Modeling multivariate health and demographic outcomes in low- and middle-income countries using sample registration and complex survey data

Austin Edward Schumacher

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

Doctor of Philosophy

University of Washington

2022

Reading Committee:

Jon Wakefield, Chair

Noah Simon

Tyler McCormick

Program Authorized to Offer Degree:
Biostatistics

University of Washington

Abstract

Modeling multivariate health and demographic outcomes in low- and middle-income countries using sample registration and complex survey data

Austin Edward Schumacher

Chair of the Supervisory Committee:
Jon Wakefield
Biostatistics

Improving the health of the most disadvantaged populations requires obtaining reliable health and demographic indicators in order to inform policy and interventions. Unfortunately, the low- and middle-income countries with the largest burden of disease and disability also tend to have the least comprehensive data. Calls for more and higher quality data collection have led (and continue to lead) to additional sources of information, namely sample registration systems and household surveys. These new data beget the development of statistical methods to produce the highest quality estimates. While the majority of current methods center on modeling univariate outcomes, improved estimates may be attained by borrowing strength across related outcomes via multivariate modeling techniques. This dissertation develops multivariate modeling frameworks using data from sample registration systems and surveys with complex sampling designs in an attempt to improve estimation of key indicators compared to modeling univariate outcomes separately.

TABLE OF CONTENTS

LIST OF FIGURES

LIST OF TABLES

ACKNOWLEDGMENTS

This has been quite a long journey for me to get this PhD, and I owe so much to those who supported me along the way.

First off, I give the deepest thanks possible to my advisor, Jon Wakefield. Saying "this wouldn't have been possible without you" is a cliche, but it's about as close to the truth as it can get. Your mentorship, advice, and guidance was invaluable. I can't wait to hit up more live shows at Seattle's finest music venues in the future.

I also want to thank the rest of my reading committee. Noah Simon, thank you for being my co-chair in spirit, if not in name. You taught me how to do research and supported me in countless ways. Tyler McCormick, thank you for your expertise while working on the research in Chapter 2 and providing immense feedback throughout the entire PhD process.

To the remainder of my committee, you have also been such an immense help. Haidong Wang, you're the longest standing member of my committee. Thank you for always being there for whatever advice I need, be it for my research, career, or life. Jim Hughes, thank you for stepping in when I needed you and for providing great insights. Kenny Sherr, my GSR, I appreciated the conversations we had and hope we can work together in a greater capacity in the future.

I also had the pleasure of collaborating with a number of others on this dissertation research. First, thank you to the Spatiotemporal Analysis with Bayes (STAB) research group. You all are awesome. I can't wait to see how all of your research projects shape up, and I'm excited to continue working with you all. Second, thank you to the group at John Hopkins: Li Liu, Jamie Perin, Pancho Villavicencio, and Yue Chu. The research in Chapter 2 was only possible with your hard work and vision.

The greater UW Biostatistics faculty and staff provided me with so much support during my time as a PhD student. I especially would like to thank Gitana Girafalo for your insight, personal support, and unwavering care through my entire graduate experience. I also want to thank the faculty for whom I was a Teaching Assistant. From you all, I became a much better instructor, and I plan to channel all the lessons I learned from you when I teach any courses in the future.

I also want to extend a massive thanks to my healthcare providers for keeping me in the necessary shape to push through. I battled through a lot over the course of my PhD, and my therapist, nurse practitioner, sports chiropractor, and physical therapists were all instrumental in preserving my helath.

To my family and friends, too numerous to list here—I cannot adequately express how much your love means to me. You all are rockstars.

Finally, I acknowledge the land I occupy today and in which this research took place is the traditional home of the Duwamish, Tulalip, Muckleshoot, and Suquamish tribal nations. I thank the original caretakers of this land who are still here. Consult https://native-land.ca/ to learn more about the occupied lands of North America. The primary goal of the following research is to improve estimation of the demographic and health landscape for the most disadvantaged populations—often, these populations have been abused through colonial impositions of power. Although the examples that will be presented do not regard tribal nations, I believe the methodology can be used for this in future research projects. Please consider this as you read any and all of this dissertation.

DEDICATION

Dedicated to the world.

Chapter 1

INTRODUCTION

Reliable estimates of health and demographic indicators in low- and middle-income countries (LMICs) are paramount to describe the health and developmental landscape of the places in most need, serving to expose inequalities that can be addressed by policy and interventions. Unfortunately, many areas suffer from poor data—either not enough is collected, or the data is not collected with the frequency, consistency, or coverage required to produce reliable estimates of population characteristics of interest. While nearly all high-income countries have near complete coverage of civil registration and vital statistics systems, the coverage of these systems in LMICs is highly inadequate. Thus, the bulk of data comes from sample registration systems (SRS) and household surveys. Many practitioners have been pushing to increase the temporal frequency, amount, and quality of data collection to address these shortcomings, both in terms of implementing new SRS and increasing frequencies of large surveys. And with more data comes the need for statistical methodology to analyze it and produce as much useful information as possible. One crucial route to leverage these data is through jointly modeling multiple outcomes. Univariate modeling of a single outcome at a time is the default for many scientific enterprises in this realm, so a shift to modeling multivariate outcomes has the potential to improve estimates of key indicators by borrowing strength across related outcomes. This dissertation will develop multivariate modeling frameworks for a variety of critical demographic and health outcomes using both SRS and survey data. This chapter will include a brief introduction to the motivating examples and describe the organization of the rest of the dissertation.

1.1 Motivating examples

1.1.1 Estimating age- and cause-specific child mortality rates from sample registration data

The Maternal and Child Health Surveillance System (MCHSS) is China's SRS devoted to maternal and child health. The MCHSS has a multistage, stratified, clustered sampling design that is nationally and subnationally representative. All children under 5 years of age living in the surveillance sites and all live births of mothers who are either permanent residents of the sites or who have lived in the sites for at least 1 year are included. Deaths are assigned a single underlying cause that is ascertained via verbal autopsy, death certificate, or last clinical diagnosis depending on the area in which the individual lives and where/how they die. This data source provides the most detailed information on child mortality in China.

Chapter 2 develops a flexible Bayesian modeling framework to estimate age- and cause-specific child mortality over time at finer age and cause resolution than previous methods. Starting at first principles with a competing risks survival process, we specify a suitable likelihood for the tabulated form of the available data. This allows for joint modeling of cause-specific mortality rates via many different types of models, and we carry out simulation studies showing the utility of models developed in this framework compared to those currently in fashion. Furthermore, this chapter provides guidance on how to choose a model depending on the available data and the goals of analysis, and uses the MCHSS data as a case study to describe how this is done.

1.1.2 Area- and unit-level small area estimation models for multivariate outcomes using complex survey data

Many LMICs do not have well-performing SRS and thus rely on data from household surveys. The DHS Program, in collaboration with many LMICs, conducts nationally-representative household surveys that provide data to estimate many demographic and health indicators.

The wide range of variables available provide an attractive setting for multivariate modeling. We have selected two motivating examples: jointly modeling height-for-age and weight-for-age in children under age 5, and modeling three categories of contraceptive use in women aged 15-49. We will be using these data from the 2014 DHS survey in Kenya.

An important feature of DHS surveys is their complex sampling design, which must be taken into account when developing statistical models for these data. The analysis of complex survey data has a rich history. Due to the limited size of many surveys along with the desire to produce estimates that are as fine -grained as possible (in terms of space, time, age group, etc), the literature on survey sampling is highly in-step with the literature on small area estimation (SAE). SAE is an umbrella term for statistical methods used to produce estimates for domains that have sparse or non-existent response data, and typically work by borrowing strength using auxiliary information. Methodologically, SAE methods can be divided into two camps: *area-level*, where covariate information and modeling parameters are specified at the desired level of aggregation for final estimates, and *unit-level*, for which data is modeled at the level of the individual or cluster. Chapter 3 will develop an area-level multivariate modeling framework, while Chapter 4 will consider a unit-level approach. Both chapters will review the relevant literature and currently used methodology, propose shared component models that account for spatial dependence with a Bayesian approach, perform simulation studies to demonstrate the performance of these models, and use them to produce estimates at the county level.

1.2 Organization of Dissertation

The following three chapters present the methodological contributions of this dissertation, which concern the development of novel multivariate modeling frameworks to estimate demographic and health data from sample registration and complex survey data in LMICs. Each of these chapters begins with an introduction, in which the relevant literature is briefly reviewed. The data from the motivating examples is described, after which the main statistical methods are presented in detail along with a simulation study. The methodology is

then used to develop models and produce estimates for the relevant motivating example(s). Each chapter ends with a discussion. Chapter 5 concludes with a review of the research and discusses the next steps for extending the methods proposed in the dissertation.

Chapter 2

A FLEXIBLE BAYESIAN FRAMEWORK TO ESTIMATE AGE- AND CAUSE-SPECIFIC CHILD MORTALITY OVER TIME FROM SAMPLE REGISTRATION DATA

The research in this chapter has been published in the *Annals of Applied Statistics*:

Schumacher A. E., McCormick T. H., Wakefield J., Chu Y., Perin J., Villavicencio F., Simon N., & Liu L. (2022). A flexible Bayesian framework to estimate age- and cause-specific child mortality over time from sample registration data. *The Annals of Applied Statistics*. 2022 Mar; 16(1):124-43.

2.1 Introduction

The United Nations Inter-agency Group for Child Mortality Estimation (2020) estimated that 5.2 million children worldwide died before five years of age in 2019. The international community is increasing investment to develop and implement age-targeted, disease-specific interventions and policy (Aponte et al., 2009; Glass, Guttmacher, & Black, 2012; Keenan et al., 2018; O'Brien et al., 2009; Penny et al., 2016) that require knowing the patterns of child deaths for multiple causes across age and time. The burden of child deaths is heaviest in low and middle-income countries (LMICs) that lack high quality vital registration (VR) systems to register all births and deaths, creating massive uncertainty. The global health community has been pushing for drastic improvements in child health, most notably with the Sustainable Development Goals (SDGs) from the United Nations (2015). SDG 3 contains age- and cause-specific targets for reducing mortality. Assessing progress toward these goals and identifying areas for improvement require accurate estimation of cause-specific child

mortality.

High quality VR data is the gold-standard for cause-specific child mortality. In most LMICs, however, VR systems are inadequate (AbouZahr, De Savigny, Mikkelsen, Setel, Lozano, & Lopez, 2015; AbouZahr, De Savigny, Mikkelsen, Setel, Lozano, Nichols, et al., 2015; Mikkelsen et al., 2015; Phillips et al., 2015). Instead, age- and cause-specific mortality data come from sample registration systems (SRS) and national/subnational surveys. Household verbal autopsy (VA) surveys comprise the bulk of national cause-specific mortality data (Soleman, Chandramohan, & Shibuya, 2006). However, these data lack continuous monitoring provided by SRS. India and China have led the way in implementing nationally representative SRS (S. Liu et al., 2016; Mahapatra, 2010; Yang et al., 2005), and calls for more and higher quality data collection (Bchir et al., 2006; Boerma & Stansfield, 2007; Jha, 2012) have encouraged establishment of SRS in countries such as Indonesia (C. Rao et al., 2010) and Mozambique (Nkengasong et al., 2020). Empirical estimates from SRS are noisy (see Figure 2.1), so as SRS data become increasingly available, developing a relevant modeling framework is crucial to produce cause-specific child mortality estimates that provide timely and useful information. Three main methods are used to estimate cause-specific child mortality that are applicable to SRS data. The first, described by L. Liu et al. (2016), models cause-specific mortality fractions (CSMFs) with a multinomial logistic regression model and then multiplies these by all-cause mortality rates estimated in a separate Bayesian framework (Alkema & New, 2014; D. You et al., 2015) to produce cause-specific mortality rates (CSMRs). The second, used in the Global Burden of Disease study and detailed in Vos et al. (2020), models either rates or probabilities of death separately for each cause with an ad hoc ensemble modeling technique and then combines these with all-cause mortality rates estimated separately using a complex regression model described in H. Wang et al. (2020). The third, described by He et al. (2017), calculates all-cause mortality rates from a single SRS in China using a 3-year moving average, proportionately scales them so they sum to the all-cause mortality rates estimated in D. You et al. (2015), and then multiplies these by estimated CSMFs at the age-region level that have been smoothed over time using a weighted

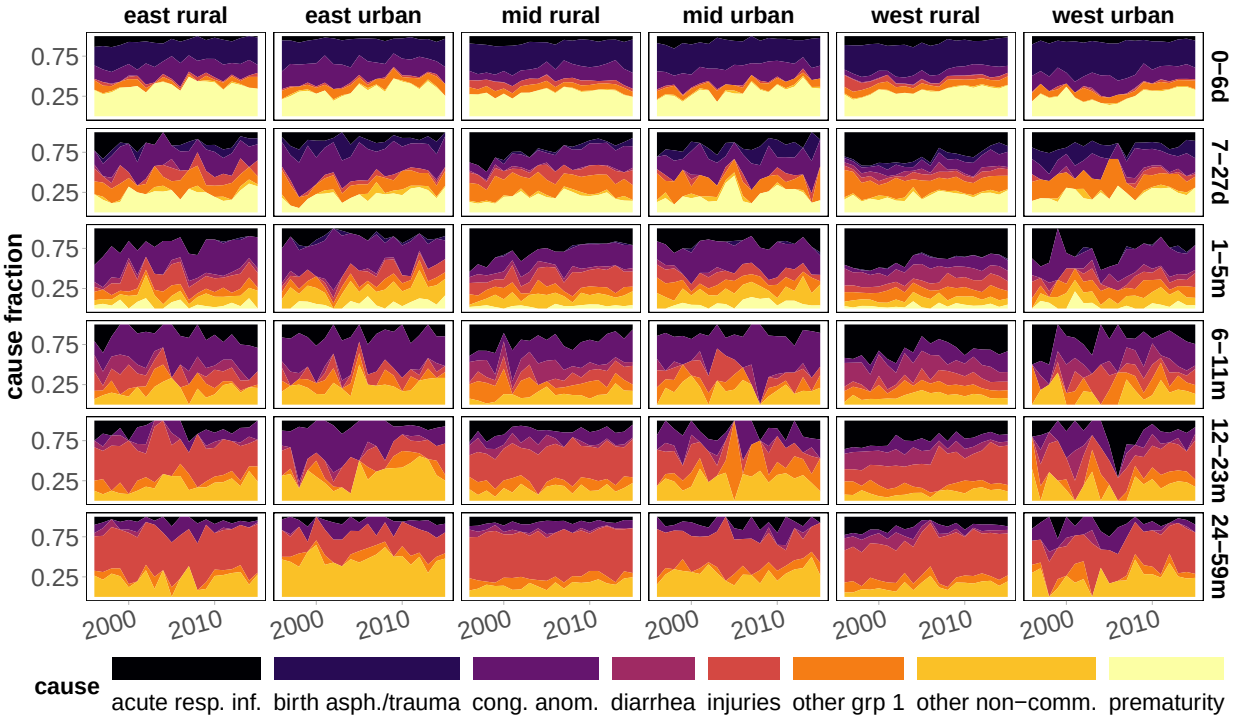


Figure 2.1: Empirical cause-specific mortality fractions over time by region and age group that were observed in the Maternal and Child Health Surveillance System in China.

seven-year moving average.

The primary issue with these methods is their use of multistage approaches that estimate all-cause and cause-specific mortality in separate, disconnected frameworks that do not "feed back" information and uncertainty between stages. A similar issue arises in other fields, for example pharmacokinetic/pharmacodynamic modeling (Bennett & Wakefield, 2001) and air pollution modeling (Keller et al., 2015), which has been addressed by using the full probability model. Joint versus non-joint modeling has been discussed in Plummer (2015). However, the multistage approaches introduced in the previous paragraph leave the models separate. Additionally, they reuse data in both stages, since data sources used for estimating cause-specific mortality are also used in the all-cause mortality models, violating assumed independence between data in the two stages and compromising uncertainty estimates.

Moreover, these multistage procedures cannot account for some features of the data that are important when modeling cause-specific mortality. For example, correlation between causes can arise when causes share a common underlying factor, such as measles and pertussis—these are both influenced by vaccine access, which may not be included in the model. Vos et al. (2020) and He et al. (2017) model each cause separately, so no correlation parameters can be included. L. Liu et al. (2016) do not model correlations, though the framework allows it. Beyond this, the multistage approach in He et al. (2017) facilitates improperly scaling to the national level using sampling probabilities without accounting for uncertainty.

Lastly, the approaches of L. Liu et al. (2016) and He et al. (2017) produce estimates for broad 0–1 month and 1–59 month age groups that are not sufficient to capture variation in cause of death by age. The cause distribution for the ages within these broad age groups have much heterogeneity (Figure 2.1), but these approaches do not disaggregate further because the all-cause mortality estimates come from large global models that only estimate for broad ages, and a lack of smoothing across available variables prevents stable estimation for subgroups with limited data.

To address these issues, this chapter brings together multiple strands of research to provide a modeling framework to estimate cause-specific child mortality rates from SRS data. We combine ideas from competing risks, loglinear modeling, and temporal smoothing to construct a framework for developing Bayesian models with an efficient implementation. Our primary contribution is estimating all-cause and cause-specific child mortality in a unified process, rather than an ad hoc multistage approach, that provides more accurate inference and predictions. The flexibility of our framework allows different functional forms to be used as needed depending on context to model mortality at fine granularity by age, cause, region, time, or other strata. We demonstrate building a model using our framework and discuss issues to consider in a motivating example.

The remainder of this chapter is structured as follows. Section 2.2 describes SRS data in general and details the data from our motivating example, the Maternal and Child Health

Surveillance System (MCHSS) in China. Section 2.3 develops our proposed framework and makes recommendations for model development. Section 2.4 demonstrates improvements over multistage modeling approaches via simulation studies. Section 2.5 describes the use of our framework to develop and fit a model to the MCHSS data and compares our results to those from He et al. (2017). Section 2.6 provides a discussion and future steps.

2.2 Data

Here, we briefly describe the data for our motivating example, the Maternal and Child Health Surveillance System (MCHSS) in China which is China's SRS devoted to maternal and child health. For further information on the history and scope of these data, see He et al. (2017). For a description of SRS data in general, see Appendix A.1.

The MCHSS has a multistage, stratified, clustered sampling design that is regarded to be representative of the six region-residency strata, henceforth referred to as regions: east urban, east rural, mid urban, mid rural, west urban, and west rural. All children under 5 years of age living in the surveillance sites and all live births of mothers who are either permanent residents of the sites or have lived in the sites for at least 1 year are included. Deaths were assigned a single underlying cause that was ascertained via verbal autopsy, death certificate, or last clinical diagnosis. Deaths were aggregated to the six regions, 20 years of surveillance (1996–2015), six age groups (0–6 days, 7–27 days, 1–5 months, 6–11 months, 12–23 months, and 24–59 months) and 16 cause groups, and then adjusted upwards using a 3 year moving average of under-reporting rates estimated from an annual quality control study. Due to small numbers of deaths, the 16 causes were aggregated into 8 mutually exclusive, collectively exhaustive groups: prematurity, birth asphyxia/trauma, congenital anomalies, other non-communicable, injuries, diarrhea, acute respiratory infections, and other communicable. For prematurity and birth asphyxia/trauma, we deleted as outliers 20 deaths which were in age groups older than 6 months due to implausibility and to prevent unstable estimation from small death counts.

Person-years at risk were not available at the granularity of these six age groups. Therefore, tabulated exposure times were estimated by standard demographic methods described in Appendix A.2. Although estimating exposure time is less exact than using recorded person-years under surveillance, exposure times are estimated in many other mortality estimation contexts using demographic methods or by making assumptions, e.g. using mid-year population estimates from the United Nations Population Division (2019). Furthermore, corrections for complex sampling and underreporting are commonly performed prior to analyses of census, survey, VR, and SRS data along with many other demographic modeling contexts (He et al., 2017; L. Liu et al., 2016; Vos et al., 2020; H. Wang et al., 2020; Wheldon, Raftery, Clark, & Gerland, 2013). These and other data issues are discussed in Section 2.6. For further details on the production of adjusted and aggregated birth and death counts from these data, we refer to the detailed description in He et al. (2017). Figure 2.1 shows empirical CSMFs over time by region and age. We provide death counts, estimated person-years, and log mortality rates by region, age, cause, and time period in Appendix A.3.

2.3 Statistical framework

2.3.1 Likelihood

SRS data arise from a competing risks failure process, as described in Prentice et al. (1978). We assume that each death occurs from a single cause, and the set of causes is mutually exclusive and collectively exhaustive. We will begin by describing the individual-level likelihood and then consider modeling data tabulated by age, time period, and strata. A full derivation is provided in Appendix A.4.

Let $i \in \{1, \ldots, n\}$ index individuals and $c \in \{1, \ldots, C\}$ index causes of death. Define T as a continuous random variable representing survival time and J as a random variable representing cause of death. We will parameterize survival time by age, i.e. time from birth for each individual. Let $\boldsymbol{z}$ be the value of a covariate vector that we assume is fixed for convenience, although this work extends to time-varying covariates in the natural way.

We define the cause-specific hazard, which in our case is the mortality rate, as $\lambda_c(t|\boldsymbol{z}) = \lim_{\Delta t \to 0} P\big(t \leq T < t + \Delta t, J = c | T \geq t, \boldsymbol{z}\big)/\Delta t$.

Suppose we have data where t_i is the time of observation, c_i is the cause of death, $\delta_i = 1$ if a death is observed and $\delta_i = 0$ otherwise (censored), and $\boldsymbol{z}_i$ is a vector of fixed covariates for subject i. Let d_{ic} indicate that individual i dies from cause c. Note that $d_{ic'} = 0$ for all $c' \neq c_i$, and for any censored observations, $d_{ic} = 0 \; \forall \; i, c$. We rewrite the likelihood from Prentice et al. (1978), up to proportionality and with independent censoring, as

$$\mathcal{L} = \prod_{i=1}^{n} \prod_{c=1}^{C} \left[\lambda_c(t_i; \boldsymbol{z}_i)^{d_{ic}} \exp\left(- \int_0^{t_i} \lambda_c(u; \boldsymbol{z}_i) du \right) \right]. \tag{2.1}$$

Next, suppose we instead have data tabulated into age groups $k = 1, \ldots, K$; here, observation time is parameterized as the age of each individual rather than calendar time. Suppose the data is also tabulated into additional strata $h = 1, \ldots, H$. These strata may include time period and region, for example. Notably, an individual may have observation time spent across multiple strata, e.g. observed across different time periods. Using arguments from Holford (1976), Holford (1980), and Laird & Olivier (1981), we define K age intervals with breakpoints $0 = \tau_0 < \tau_1 < \cdots < \tau_K$. Define $d_{ihkc} = 1$ if individual i dies of cause c in age group k and strata h, and define t_{ihk} as the total observation time that individual i is observed in age group k and strata h. Note that $d_{ic} = \sum_{k=1}^{K} \sum_{h=1}^{H} d_{ihkc}$ because a person dies while in only one age group and strata, and additionally an individual's total observation time can be expressed as $t_i = \sum_{k=1}^{K} \sum_{h=1}^{H} t_{ihk}$. Define the observed data as $y_{hkc} = \sum_{i=1}^{n} d_{ihkc}$, the number of deaths in age group k and strata h from cause c, and $t_{hk} = \sum_{i=1}^{n} t_{ihk}$, the total person-time observed in age group k and strata h.

We will assume that cause-specific hazards are constant within each age-strata tabulation group, and for simplicity, we will assume no covariates. Thus, $\lambda_c(t_i; \boldsymbol{z}_i) = \lambda_{hkc}$ for individual i in strata h with $t_i \in [\tau_{k-1}, \tau_k)$. However, $\lambda_c(t; \boldsymbol{z})$ can depend on covariates and the following derivation extends naturally. The tabulated likelihood is

$$\mathcal{L} = \prod_{h=1}^{H} \prod_{k=1}^{K} \prod_{c=1}^{C} \left[\lambda_{hkc}^{y_{hkc}} \exp\left(-\lambda_{hkc} t_{hk} \right) \right]. \tag{2.2}$$

This is identical to the kernel of the likelihood that would arise if $y_{hkc}|\lambda_{hkc}, t_{hk} \sim \text{Poisson}(\lambda_{hkc}t_{hk})$. Therefore, we can make likelihood-based inference using separate Poisson distributions for each cause and age group.

The product of Poisson likelihoods in Equation ((2.2)) is equivalent to a model in which the all-cause death counts and person-years have a Poisson distribution and the cause-specific counts conditional on the total death counts have a multinomial distribution. This equivalency of likelihoods is detailed in Lee, Green, & Ryan (2017) and is commonly exploited in modeling multinomial count data. More details are provided in Section 4.5.1 of Chapter 4. This gives rise to a specific multistage modeling specification: First estimate all-cause mortality rates using a Poisson distribution, and then estimate CSMFs conditional on the all-cause death counts using a multinomial distribution. We show the consistency of this two-stage model via simulation in Appendix A.5. However, none of the multistage modeling approaches detailed in Section 2.1 specify this consistent two-stage likelihood.

Statistical modeling based on our framework provides the flexibility to choose a model that is driven by the study-specific context. It is natural to work with loglinear hazard models due to the constraint that mortality rates must be positive. Then, given a vector of parameters $\boldsymbol{\eta}$, we can construct a model as $\log(\lambda_{hkc}) = f_{hkc}(\boldsymbol{\eta})$.

2.3.2 Model development

Using our proposed modeling framework, one can specify functional forms for $f_{hkc}(\boldsymbol{\eta})$ that contain fixed and random effects (Breslow & Clayton, 1993), use copula functions (Smith & Khaled, 2012), or many other methods. This flexibility is critical to account for the main drivers of mortality, but it also begets the need for careful model construction. We will discuss choosing a functional form for $f_{hkc}(\boldsymbol{\eta})$ by considering cause of death, time period, age, geography, and the interactions between these variables. We will also discuss accounting for overdispersion as well as model validation and comparison.

Cause of death

The relative distribution of causes of death can vary immensely depending on the context of data collection (Clark, Setel, & Li, 2019). In order for the competing risks framework to hold, we must have an exhaustive list of mutually exclusive cause groups. For modeling, we need a sufficiently large number of deaths in each group to provide stable estimation. In most situations, we expect that only a small number of cause groups would be plausible to model, and these groups will have distinct differences in mortality. Thus, we recommend using a fixed intercept for each group.

Another consideration is that causes may be correlated due to covariates that were not collected but influence multiple causes of death, such as environmental factors. Modeling correlations may improve estimates, especially for time periods/regions/ages with little or no data. Unfortunately, estimating correlation parameters is difficult and requires large amounts of data (we discuss this further in Section 2.6). If data permits stable estimation of correlation parameters, we recommend a hierarchical modeling approach using multivariate normal cause-specific intercepts, i.e. for C causes, $f_{hkc}(\boldsymbol{\eta})$ would include a vector of fixed effects $\boldsymbol{\beta} \sim \text{Normal}_C(0, \boldsymbol{\Sigma})$, where the off-diagonal entries of the covariance matrix $\boldsymbol{\Sigma}$ parameterize the correlations between causes, which can be estimated from the data. A simple example of this model for two cause groups is presented in a simulation in Section 2.4.2.

Time period

Prompt and accurate time trend estimates support policy enactment, intervention targeting, and resource allocation in an agile fashion (Friberg et al., 2010). In addition, they allow evaluation of performance toward child survival targets, such as the SDGs, and also serve as important quality indicators for global health statistics and their estimation (N. Walker, Bryce, & Black, 2007). For data with a smaller time span than the MCHSS, modeling can be done via linear trends. As the number of available time points grows, we recommend more flexible methods such as those commonly used in other child mortality models. Second-order

random walks are popular (Z. R. Li et al., 2019), in which case $f_{hkc}(\boldsymbol{\eta})$ would include a vector of random effects $(\gamma_1, \ldots, \gamma_T)$ for T time points that parameterizes the second differences in time as arising from a normal distribution, i.e. $\gamma_t - 2\gamma_{t+1} + \gamma_{t+2} \sim \text{Normal}(0, \sigma_\gamma^2)$, with σ_γ^2 parameterizing how quickly the time trend can vary from year to year (Rue & Held, 2005). Other options include cubic spline or B-spline models (Alkema & New, 2014); notably, a certain class of cubic splines are equivalent to second-order random walks (Speckman & Sun, 2003). Another option is random effects with an autoregressive distribution (Chi & Reinsel, 1989).

Age

Mortality trends vary drastically by age; for example, the neonatal period has much higher mortality than the rest of the under-five age range. Age is commonly tabulated into groups, either for data reporting standards, for convenience of modeling, or because of interval censoring. In the past, data for under-5 mortality has been disaggregated into the first year and the combined remaining four years; thankfully, recent data collection and estimation has favored further disaggregation into early- and late-neonatal, along with further breakdown of the 1–4 year period (L. Liu et al., 2016; Vos et al., 2020). Finer age groups allow for more useful estimates to direct health interventions. Treating age as a categorical variable allows flexible, non-monotonic relationships, a modeling choice that is facilitated by the commonly-available tabulated form of SRS data. We recommend using fixed intercepts to capture the main effects of age due to large differences in mortality among commonly available tabulated groups.

Geography

In the SDG era, health policy and program decision making are becoming decentralized with many decisions now happening at the district level. Subnational mortality estimates help adapt the development of health statistics to meet changing needs that vary by region

(Boerma, 2013). Often the geographic information consists of the administrative area in which the child resides. When there are few areas, spatial modeling is difficult because of the paucity of information. In this situation, fixed effects for areas can be used.

If more detailed location information is available, many methods can be used with our framework. Data availability and the scale for reporting estimates should drive this modeling choice. For areal data, popular choices include the reparameterized Besag, York and Mollié (BYM) model (Riebler, Sørbye, Simpson, & Rue, 2016) and DAGAR models (A. Datta, Banerjee, Hodges, Gao, et al., 2019), amongst many others (Heaton et al., 2019). For point-referenced data, such as individual- or cluster-level data, one can use the Gaussian Markov random field representation of the stochastic partial differential equations approach with a Matérn family for point-referenced data (Lindgren, Rue, & Lindstrom, 2011) or nearest neighbor Gaussian process models (A. Datta, Banerjee, Finley, Hamm, & Schaap, 2016).

Interactions

Interactions between cause of death, time period, age, and geography must be considered as we expect differential effects for different combinations of these variables. The amount of data available restricts the number of interactions that can be modeled, which means the context of the data analysis is crucial when choosing a model.

Due to the importance of time trends for global health policy and interventions, modeling different time trends for each age, region, and cause is paramount. For example, mortality rates from injuries are likely to change differently for infants compared to older children, and these may further be different in rural locations compared to urban locations. Modeling these interactions involves allowing for time trends to vary by age-region-cause strata. How to achieve this goal depends on the choice of how main effects are modeled. For example, if temporal trends are modeled via a random walk, one can use different random walk parameters for different combinations of age, region, and cause of death. More complex space-time models can also be used, such as those described in Wakefield et al. (2019).

Non-temporal interactions to consider include the different distributions of causes of death among age groups and regions, as well as different age effects among regions (Abdullah et al., 2007; Snow et al., 1997; C. L. F. Walker et al., 2013; WHO Collaborative Study Team on the Role of Breastfeeding on the Prevention of Infant Mortality, 2001). Accurate modeling of these allows interventions to be targeted to populations in most need. Again, the decision on how to model these interactions depends on how the main effects were chosen to be modeled. For example, if cause of death, age, and geography were all modeled with fixed effects, then one can simply include two-way (and possibly three-way) interactions of these variables. Models with more complicated main effects will necessitate more complex interaction models.

While child mortality is likely to vary by all possible interactions among cause of death, time, age, and geography, one cannot typically include all of these interactions in a model due to the limited amount of available data. Therefore, it is crucial to carefully choose which interactions are most important to model. We believe the end scientific goal should drive the interactions on which to focus. Once the main goals are decided upon, initial data exploration and simulations can be used to determine the extent to which the amount of data allows for different interaction models. We present examples of these exercises in Section 2.5 with additional details in Appendices A.9, A.10, and A.16.

Overdispersion

Overdispersion is common in child mortality data due to within-strata variability, such as from unobserved covariates, measurement error, non-systematic errors from the data pre-processing steps, and cause misattribution. To account for this, we recommend including observation-specific random effects in $f_{hkc}(\boldsymbol{\eta})$, i.e. $\epsilon_{hkc} \overset{iid}{\sim} \text{Normal}(0, \sigma_\epsilon^2)$. This popular method allows for additional variability, although it does not facilitate differentiating between different sources of overdispersion.

Model validation and comparison

Due to the flexibility of our framework and the many decisions that must be made when choosing how to model cause, time, age, geography, and their interactions, models must be checked for adequacy. We recommend plotting standardized residuals grouped by all combinations of each of the strata used in modeling (including two-way, three-way, and higher combinations if necessary) and examining these plots for any distinguishable patterns. Furthermore, one can perform hold out experiments to evaluate predictive performance for important aspects of the model. For example, one can hold out the final year of observations, fit the model, and compare predictions to the held out data in order to evaluate predictive validity of short term time trends. Finally, we recommend comparing the performance of a suite of candidate models via traditional model comparison metrics such as the deviance information criteria (DIC), Watanabe-Akaike information criterion (WAIC), and conditional predictive ordinates (CPO). For a detailed discussion of model comparison, see Gelman, Hwang, & Vehtari (2014) and Held, Schrödle, & Rue (2010).

2.4 Simulations

This section compares models fit using our framework, which we call unified models, to models fit with multistage frameworks that combine separate cause-specific models. We fit models in this section with the `INLA` package for fast estimation using integrated nested Laplace approximation (Rue, Martino, & Chopin, 2009) in the R statistical computing environment (R Core Team, 2013). Models use the default prior distributions in `INLA`. Full descriptions of each simulation are provided in Appendix A.6. Replication code is available at http://www.github.com/aeschuma/SRS-child-mortality.

2.4.1 Scenario 1: Extra-Poisson variability

Let $h \in \{1, \ldots, H = 720\}$ index strata with 6 age groups, 6 regions, and 20 years, and let $c \in \{1, \ldots, C\}$ index cause. Define N_h as the total exposure time and y_{hc} as the death counts

from cause c in strata h. Define λ_{hc} as the CSMRs, $y_{h+} = \sum_c y_{hc}$ as the all-cause death counts and $\lambda_{h+} = \sum_c \lambda_{hc}$ as the all-cause mortality rates. For $C = 8$ causes, we specify

$$\begin{aligned} y_{hc}|\lambda_{hc} &\sim \text{Poisson}\left(N_h \lambda_{hc}\right) \\ \log(\lambda_{hc}) &= \alpha + \sum_{c'=2}^{C} \beta_{c'} \mathbb{1}_{[c'=c]} + \epsilon_{hc} \\ \epsilon_{hc}|\sigma_\epsilon^2 &\overset{iid}{\sim} \text{N}(0, \sigma_\epsilon^2). \end{aligned}$$

This exemplifies a scenario with different mortality for each cause along with extra-Poisson variability. We set $\alpha = -5$ and $\beta_c = 0.5$ for all c, which yields death counts in the same range as the MCHSS data with a similar overall mean mortality rate of 0.01. We set $\sigma_\epsilon^2 = 0.2$, which is approximately equal to the level estimated from our model fit to the MCHSS data described in Section 2.5.

For the multistage model, we estimate all-cause mortality rates using a Poisson GLMM with an overall intercept and IID Normal random effects on strata such that

$$\begin{aligned} y_{h+}|N_h, \lambda_{h+} &\sim \text{Poisson}(N_h \lambda_{h+}) \\ \log(\lambda_{h+}) &= \alpha + \gamma_h \\ \gamma_h|\sigma_\gamma^2 &\overset{iid}{\sim} \text{Normal}(0, \sigma_\gamma^2). \end{aligned}$$

To estimate the CSMFs, we use separate Poisson generalized linear mixed models for each cause. These models each have an overall intercept and IID Normal random effects on strata. For each cause c, we have

$$\begin{aligned} y_{hc}|N_h, \lambda_{hc} &\sim \text{Poisson}(N_h \lambda_{hc}) \\ \log(\lambda_{hc}) &= \alpha_c + \xi_{hc} \\ \xi_{hc} &\overset{iid}{\sim} \text{Normal}(0, \sigma_{\xi c}^2). \end{aligned}$$

Taking samples $s = 1, \ldots, 1000$ from the posteriors, for each sample we calculate CSMFs as $\widehat{p}_{hc}^{(s)} = \widehat{\lambda}_{hc}^{(s)} / \sum_c \widehat{\lambda}_{hc}^{(s)}$, all-cause mortality rates as $\widehat{\lambda}_{h+}^{(s)}$, and combine these to calculate log CSMRs as $\log\left(\widehat{\lambda}_{hc}^{(s)}\right) = \log\left(\widehat{\lambda}_{h+}^{(s)} \widehat{p}_{hc}^{(s)}\right)$.

To compare with the multistage model, we fit a unified model correctly specifying the data generating mechanism and draw 1000 posterior samples for each log CSMR. For both the multistage and unified models, we perform 100 simulations each for $N_h \in \{1000, 10000, 100000\}$. These are in the range of the 5th, 50th, and 95th percentiles of exposure time in the MCHSS data, which are 292, 8749, and 110974, respectively. We compare the relative bias, coverage and width of posterior 95% intervals for log mortality rates as functions of exposure time in Figure 2.2. Neither approach is biased but the unified model has better coverage due to appropriately wider uncertainty intervals. The problem is likelihood misspecification, because the multistage model parameterizes the log of the sums of mortality rates as normally distributed, whereas in the generated data, the sum of the log mortality rates are normally distributed, which the unified model specifies correctly. This substantial undercoverage in the multistage model is derived analytically in Appendix A.7. While we are imposing a normal distribution on the log cause-specific mortality rates in this example, which may not reflect the real world, we believe this parametric form is a reasonable assumption since it corresponds to the canonical assumptions of exponential failure times in competing risks survival analysis.

2.4.2 Scenario 2: Correlated causes

To show the benefits of our flexible framework for a contextually relevant situation in which a multistage approach fails, we simulate data for two causes with correlated CSMRs, representing causes with similar underlying drivers not captured in the data.

We use the same data generating mechanism as Scenario 1 with two changes. First, we include region and age as covariates with omitted reference groups and all associated coefficients equal to 0.5, in order to more closely resemble a situation where mortality depends on region and age. These again yield death counts and an overall mean mortality rate similar to the MCHSS data. Second, we specify $\boldsymbol{\epsilon}_h = (\epsilon_{h1}, \epsilon_{h2}) \overset{iid}{\sim} \text{Normal}_2(\mathbf{0}, \boldsymbol{\Sigma})$ for the two causes. We set the diagonal terms of $\boldsymbol{\Sigma}$ to be equal to σ^2, which controls overdispersion. The off-

diagonal term is $\sigma^2\rho$, where ρ controls the correlation between causes. We set the exposure time to the median in the MCHSS.

We fit the same multistage model as the previous simulation, except both stages also include fixed effects for region and age, and compare it to a correctly specified unified model. We perform 100 simulations for each of nine scenarios: $\sigma^2 \in \{0.01, 0.1, 1\}$ crossed with $\rho \in \{-0.5, 0, 0.5\}$. The values of σ^2 span a range that are much smaller, similar to, and much larger than that estimated from the MCHSS data, and the values of ρ span a range that are consistent with residual correlations estimated from the MCHSS data in an exploratory analysis in Appendix A.8.

Relative bias, coverage and width of 95% intervals for estimated log mortality rates as functions of σ^2 for each value of ρ are shown in Figure 2.2. The bias in both models is negligible, although it increases slightly with higher overdispersion due to small-sample bias. We also see little direct impact of ρ. The coverage of the multistage model ranges between 80% and 90%, while the unified model has coverage near the nominal level with appropriately wider uncertainty intervals, although it displays overcoverage with small levels of overdispersion. This is not surprising because the elements of Σ are all nearly 0 in this case, so our model is estimating parameters for which there is very little information. This leads to overcoverage from additional uncertainty. The multistage model's undercoverage is again due to likelihood misspecification. In summary, the flexibility of our framework facilitates modeling aspects of the data, such as overdispersion and correlation, that a multistage model cannot, which leads to more accurate inference.

2.5 Estimating mortality from SRS data

This section uses our framework to develop a model to estimate child mortality from the MCHSS.

2.5.1 Model description

We index region by $r \in \{1, \ldots, R = 6\}$, age group by $a \in \{1, \ldots, A = 6\}$, year by $t \in \{1, \ldots, T = 20\}$, and cause by $c \in \{1, \ldots, C = 8\}$. Let $N_{r,a,t}$ be person-years and $y_{r,a,t,c}$ be death counts due to cause c in region r, age group a, and year t. To estimate cause-specific mortality rates by region and age over time, we specify the model as

$$\begin{aligned} y_{r,a,t,c} | N_{r,a,t}, \lambda_{r,a,t,c} &\sim \text{Poisson}(N_{r,a,t}\lambda_{r,a,t,c}) \\ \log(\lambda_{r,a,t,c}) &= \alpha + \beta_r^R + \beta_a^A + \beta_c^C + \\ &\quad \beta_{a,c}^{AC} + \beta_{r,c}^{RC} + \beta_{a,r}^{AR} + \\ &\quad \gamma_{r,a^\star[a],c^\star[c]}(t) + \epsilon_{r,a,t,c} \end{aligned} \tag{2.3}$$

In order to properly use a Poisson likelihood, we rounded deaths up to the nearest integer because the death counts in the MCHSS were fractional due to the underreporting adjustment described in Section 2.2.

In Equation ((2.3)), α is the overall intercept, and β_r^R, β_a^A, and β_c^C are fixed effects for region, age, and cause, respectively. These are specified with omitted reference groups, i.e. $\beta_1^R = \beta_1^A = \beta_1^C = 0$. We chose to use fixed effects due to the small numbers of regions, age groups, and causes along with the strong differences in mortality among the different strata. Notably, due to regions being defined not simply by geography but by urban/rural status as well, we did not include any spatially dependent effects because there were only three geographic regions.

Notably, we did not model correlations between causes. We performed a simulation study, presented in Appendix A.16 and further discussed in Section 2.6, in which we simulated two data sets with correlated CSMRs—one that was the same size as the MCHSS and one with 100 regions rather than six. The correlation parameters were well-estimated for the data with 100 regions but not for the MCHSS-sized data. Consequently, we did not model correlations in our final model.

To capture first-order interactions among cause, age, and region, we include $\beta_{a,c}^{AC}$, $\beta_{r,c}^{RC}$,

and $\beta_{a,r}^{AR}$ as fixed effects with omitted reference groups. To determine how to model these non-temporal interactions, initial data exploration using non-Bayesian GLMs showed that a generalized linear model with all three of these interactions had a substantially lower AIC than any model with only one of them included, and had only slightly higher AIC than a model that additionally included a three-way interaction. Because the model with a three-way interaction had approximately twice as many parameters, we chose to only include two-way interactions for parsimony.

We place proper but flat priors on all the above fixed effects except α for which we use an improper flat prior due to our choice to model temporal trends with a second-order random walk (discussed later). Prior choice is important in more data sparse situations, but here estimating with stronger priors makes little difference, as we establish with sensitivity analyses described in Section 2.6.

The parameter $\gamma_{r,a^\star[a],c^\star[c]}(t)$ is a random effect on time with a second-order random walk distribution, which we denote as $\gamma_{r,a^\star,c^\star}(t) \sim \text{RW2}(\sigma_\gamma^2)$. We include different random walks for various age-region-cause combinations and we share random walks among certain ages and causes, hence indexing by $a^\star$ and $c^\star$. We define $a^\star[a] = 1$ for observations in the 0–6 day age group ($a = 1$) and $a^\star[a] = 2$ otherwise ($a = 2, \ldots, 6$). We define $c^\star[c] = 1$ for *diarrhea* and *other communicable diseases* ($c = 1, 2$), $c^\star[c] = 2$ for *congenital anomalies* and *other non-communicable diseases* ($c = 3, 4$), and $c^\star[c] = 3, \ldots, 6$ for the remaining causes ($c = 5, \ldots, 8$, respectively). This results in 6 (region) $\times$ 2 (age) $\times$ 6 (cause) = 72 random walks. By using different random walks to allow age-region-cause strata to have distinct trends, we accomplish a similar goal as the model in He et al. (2017) without using ad-hoc weighted rolling averages. All random walks share a variance parameter, σ_γ^2 for parsimony and to reduce the number of estimated parameters to aid computation. For identifiability, we use sum-to-zero constraints on each random walk along with an improper prior on the intercept α, which is required for correct specification of a second-order random walk (Rue & Held, 2005). We use a penalized complexity prior (Simpson et al., 2017) on the variance parameters such that there is a 1% probability that $\sigma_\gamma > 1$.

We use a second-order random walk to encourage the estimated mortality rates to vary smoothly in time. A second-order random walk model penalizes deviations from linearity and is more smooth than a first-order random walk, which models the first differences in time as being normally distributed and allows for sharp year-to-year variation. The additional smoothness of the second-order random walk is in line with what we expect in the MCHSS data.

We share random walks among certain ages and causes to aid computation. The categories for sharing random walks were chosen via a data-driven exercise that accounted for the scientific context. We fit a suite of Poisson generalized linear models that contained interactions between time and all one-way, two-way, and three-way combinations of region, age, and cause, and then analyzed the residual plots for common patterns that were consistent with the context of child mortality in China. The 0–6 day age group had consistent patterns in the residuals that were different than the other ages, which is reasonable due to the biological uniqueness of this age group such as higher mortality and its dependence on birth-related interventions of health facilities. The causes that share random walks also had similar residual patterns which are reasonable because *diarrhea* and *other communicable diseases* are communicable, while *congenital anomalies* and *other non-communicable diseases* are non-communicable. This is fully detailed in Appendix A.9.

We tested the feasibility of the shared variance parameter by separately fitting random walk models for the data in each of the 72 age-region-cause combinations and comparing the estimated standard deviation parameters, which are presented in Appendix A.10. The estimates ranged from 0.005 to 0.1, with the majority below 0.025. Sharing the variance parameter will shrink the rates of change in some of the time trends toward the average, but not drastically.

We specify $\epsilon_{r,a,t,c} \overset{iid}{\sim} \text{Normal}(0, \sigma_\epsilon^2)$ with a penalized complexity prior on σ_ϵ^2 such that there is a 1% probability that $\sigma_\epsilon > 5$. We include this term to account for overdispersion as discussed in Section 2.3.2. We treat $\epsilon_{r,a,t,c}$ as an error term rather than true signal and do not include it in the posterior distribution of our final estimates. While this parameter likely

captures some true signal, we believe the relative strength of the signal is low because the data is a sample with quality issues, which we discuss in Section 2.6. With more covariates and higher data quality, the magnitude of the noise component would decrease. Furthermore, by omitting $\epsilon_{r,a,t,c}$, our final estimates reflect the underlying smooth time trends. We may wish to include this parameter in contexts with more data, higher data quality, or where the goal is to estimate true numbers of deaths in a population rather than underlying mortality rates.

To demonstrate model adequacy, we plotted $(y_{r,a,t,c} - N_{r,a,t}\widehat{\lambda}_{r,a,t,c})/(N_{r,a,t}\widehat{\lambda}_{r,a,t,c})^{1/2}$, the standardized residuals, by all two- and three-way combinations of age, region, cause, and time. These plots were examined for patterns that may suggest inadequate model fit to the data. As an example of evaluating predictive performance, we held out the final year of data, fit the model, predicted the log mortality rates, and then plotted these against the held out data. This particular hold-out experiment was performed because short-term predictions are relevant for policy decisions. Finally for model comparison, we fit a model with no interactions between fixed effects to assess the necessity of these interactions, and a model using first order autoregressive (AR1) processes rather than second-order random walks to assess how we modeled temporal trends. We compared these to our final model via DIC, WAIC, and negative sum of the log CPO.

We fit the model with the `INLA` package in R. Code for data processing, model fitting, and plotting results is available at http://www.github.com/aeschuma/SRS-child-mortality.

2.5.2 Results

Figure 2.3 shows estimated posterior medians and posterior 80% intervals for the log mortality rates over time in each age group for selected regions and causes. In order to show how the fixed effects component of the model contributes to these estimates, we include posterior medians and 80% intervals for the sum of the fixed effects only; to see the contribution of

the random effect error terms, $\epsilon_{r,a,t,c}$, we also show the posterior medians and 80% intervals for the estimated log mortality rates with these added. As discussed in Section 2.5.1, there is ambiguity in the source of these error terms, but their inclusion allows for a greater sense of the sampling variation in the data to be realized.

We first present estimates for prematurity in the west rural region, which represents the highest number of deaths in the data. Our model fits well, although the estimates are consistently higher than the observed data in the 7–27 day age group in later years, which is due to borrowing strength across other strata. This may indicate data errors, such as underreporting missed by the adjustment, or it may be due to shrinkage induced by the random walks. In comparison to preliminary models with random walks for each age-region combination only, the time trends for west rural prematurity were carried over to all causes and did not fit the data well. This is testament to the importance of including random walks by age-region-cause strata. Looking at a different cause and region, other non-communicable diseases in the east urban region, we see largely flatter time trends with wider uncertainty reflecting the smaller amount of data. In the random walk fitting exercise described in Section 2.5.1, this strata had a much lower estimated random walk variance parameter than the previous two strata presented, but all have acceptable fits here. Plots of all estimates from our final model are available in Appendix A.11. Additionally in Appendix A.12, we present estimated CSMFs over time for each age group and region to show differences in the disribution of causes by age and region over time. For example, the percentage of deaths due to congenital anomalies in the 1-5 and 6-11 month age groups is fairly constant in the east urban region but increasing in the mid rural region.

To show model adequacy, we present plots of standardized residuals stratified by all two- and three-way combinations of age, region, cause, and time in Appendix A.13, and plots of predicted vs. observed from a model with the final year held out in Appendix A.14. These show no gross misspecification and predictions with no systematic biases. Furthermore, in Appendix A.15 we present comparisons of the DIC, WAIC, and negative sum of the log CPO among our final model, a model with no interactions between fixed effects, and a model

using AR1 processes rather than second-order random walks. Not surprisingly, leaving out interactions caused a substantially poorer fit; the AR1 model was close but still inferior. Since the method of He et al. (2017) also estimates cause-specific child mortality from the MCHSS data, we compare their results to our estimated CSMFs in the east rural region as an example in Figure 2.4. One primary benefit of our estimates is the more detailed age granularity. The approach in He et al. (2017) produced estimates for the 0–1 month and 1–59 month age groups. Any further disaggregation would presumably lead to unstable estimation as their model performs temporal smoothing from moving averages for each region-age-cause separately. In comparison, our method borrows strength in two key ways in order to facilitate modeling more granular age groups. First, we include fixed effects for interactions among age groups, causes of death, and regions. Second, we share random walks in time among different strata. Our estimates show substantial heterogeneity across the age groups that is masked by the large age bins in He et al. (2017). Figure 2.4 also shows smoother temporal trends in our estimates, demonstrating less year-to-year variability which is what we expect for the population-level parameters in China that we are estimating from our sample.

This comparison illustrates important conclusions that are missed by He et al. (2017). In the 0–1 month age group, He et al. (2017) estimate a large decline in the cause fraction for acute respiratory infections, resulting in a near-negligible percentage in 2015. However, our model shows that this cause is still meaningful in the 7–27 day age group. For intervention funding allocated in this region, our results indicate continued investment in acute respiratory infection prevention among children aged 7-27 whereas the results from He et al. (2017) do not. Another key takeaway is the comparatively larger portion of deaths due to injuries in the 24–59 month age group, along with smaller portions due to acute respiratory infections and congenital anomalies. This pattern could influence policies and resources to be enacted in a more targeted manner.

2.6 Discussion

We have introduced a unified, flexible framework for estimating age- and cause-specific child mortality over time using tabulated death counts and exposure time from SRS data. This framework is based on an individual-level competing risks likelihood along with Bayesian smoothing priors. We have shown that it performs better than multistage modeling on simulated data with overdispersion and correlation, and we used the framework to develop a model for the MCHSS data from China.

Our framework improves upon current methodologies by simultaneously estimating all-cause and cause-specific mortality in a unified framework rather than using improper multistage models. We compared our model for the MCHSS data in China to the model in He et al. (2017). Our model improved estimation in four key ways: (1) He et al. (2017) improperly scales to national estimates as part of their multistage model, which is effectively unnecessary in our model because all-cause and cause-specific estimates are estimated simultaneously and consistently, which properly quantifies uncertainty; (2) we directly derive a likelihood for tabulated death and exposure time data commonly reported from SRS which allows us to fully specify a flexible estimation model in a statistically well-grounded fashion; (3) we model temporal trends with second-order random walks rather than using ad hoc moving averages; and (4) we employ proper smoothing and variance estimation that allows stable estimates at finer subgroup granularity. Our estimates are smoother and tease out important heterogeneity in finer age groups that is missed in the results from He et al. (2017).

While we developed our framework to estimate child mortality from SRS data, it is applicable to other scenarios. We can use cause-specific child mortality data tabulated from household surveys (as long as sampling probabilities are correctly accounted for), although modeling the individual data in this case may be preferred as discussed below. Additionally, high quality VR can be modeled with our framework, for which we recommend less smoothing due to near complete population coverage. Furthermore, by incorporating multiple types of

data from multiple countries, our framework can be adapted for large scale estimation akin to L. Liu et al. (2016) and Vos et al. (2020). This extension would be computationally expensive, however, necessitating developments to be explored in future research. Beyond child mortality, we can use our framework to develop models for estimating any rates with competing risks from data sources that provide tabulated counts and time-at-risk, or from which these values can be calculated. Some examples are cause-specific mortality for ages beyond 5 years, incidence of non-fatal diseases, and rates of traffic accidents by severity.

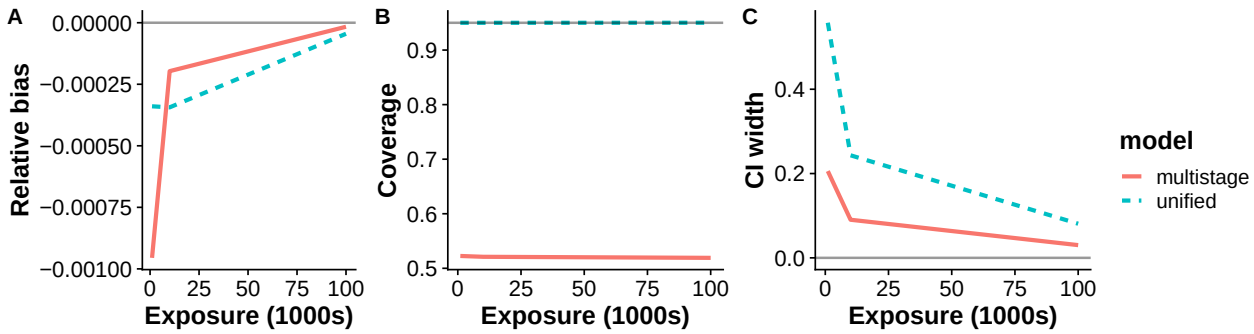
The high degree of flexibility in our framework allows complex trends to be estimated, but data availability should drive model choice. Simulation experiments with data generated similar to the true data can reveal what forms are estimable. As one example, we attempted an alternate specification for the MCHSS data that accounted for correlations between causes. We simulated data with the same size as the MCHSS that had correlated CSMRs and fit a correctly specified model. Recovery of the correlation parameters was poor and posterior distributions were wide. We then simulated data with 100 regions rather than six and fit a similar model. The correlation parameters were recovered well with narrow posterior distributions. Thus, we chose not to model correlation. A full description of this simulation exercise and its results are provided in Appendix A.16.

Additionally, choosing prior distributions requires careful thought. We used diffuse priors for the fixed effects and penalized complexity priors for the variance parameters of the random walks and IID Normal random effects. The latter parameters are more sensitive to prior choice. We recommend penalized complexity priors due to the reasons outlined in Simpson et al. (2017), in particular allowing for specifying context-relevant prior distributions. As sensitivity analyses for our model fit to MCHSS data, we fit one model with Gamma(5, 0.00005) priors on the precisions and one model with stronger Normal(0, 5) priors on the fixed effects. No appreciable differences were found; the largest absolute difference in posterior median log mortality rates was 0.05.

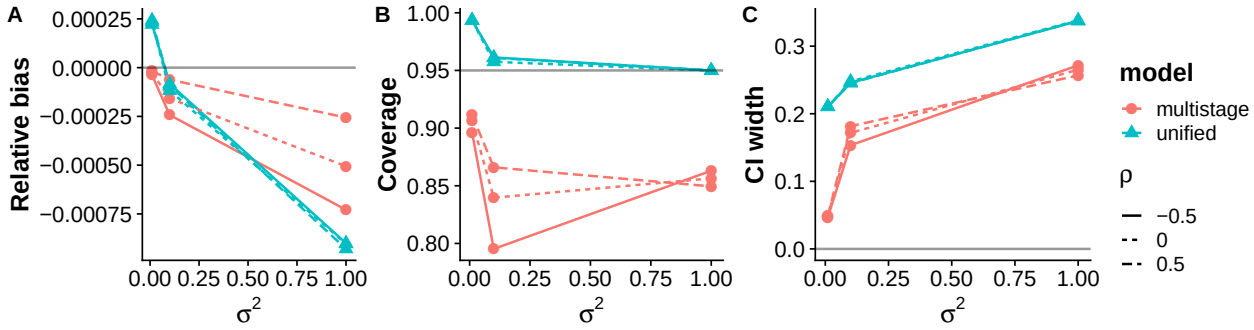
We propose that the data context suggests the most important aspects of child mortality to model for the setting at hand. From this viewpoint, one could fit multiple candidate

models and perform model assessment to choose the most suitable model for the scientific goal. The independent random effects are useful for this endeavor, for example plotting them against time to detect patterns. Cross validation is also useful and could include multiple levels depending on the context (e.g., leaving one observation out, leaving one region out, leaving one time period out). Different levels of cross validation explore aspects of the data for which different models may perform better (Roberts et al., 2017).

Finally, the MCHSS data provide an example of common problems to be addressed in SRS data. Underreporting adjustment and exposure time estimation are common data preprocessing steps. Beyond this, errors in cause attribution introduce substantial variability (Desai et al., 2014; Murray et al., 2014). These issues naturally lead to using random effects in order to induce overdispersion which may accommodate this extra variability, though they cannot of course remove bias. If available, the unadjusted, unaggregated data would be used and the aggregation and completeness adjustments would be included as components of the Bayesian model. Using individual-level data allows proper calculation of person-time, necessitating proper handling of the censoring for individual-level survival data. Individual-level data also require explicit incorporation of sampling probabilities either with design-based or model-based estimation (Pfeffermann, 2013). With individual-level VA data, probabilistic cause assignment could also be included, for example using the method in McCormick et al. (2016). Extending our model in this manner would be necessary when using smaller scale surveillance data on individuals, e.g. data from COMSA Mozambique (Nkengasong et al., 2020), HDSS sites in the INDEPTH network (Sankoh & Byass, 2012), or HDSS sites in the ALPHA network (Maher et al., 2010). Future work could also expand this framework to include survey data, notably VA data, in a single model.

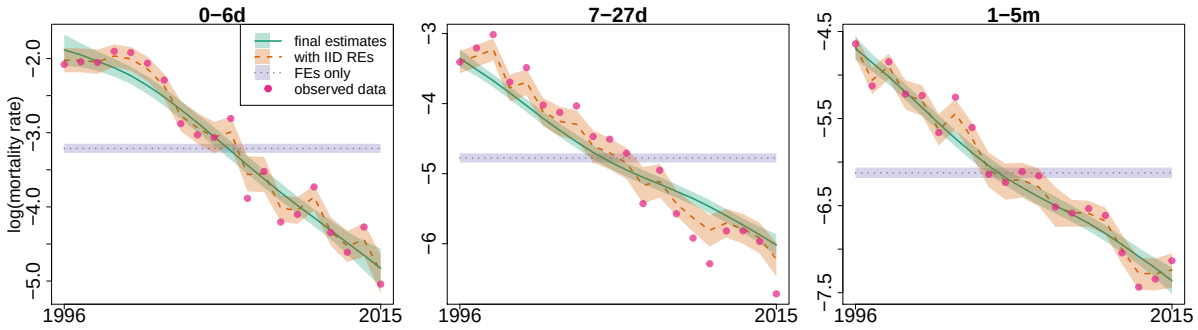


(a) Scenario 1: eight causes with extra-Poisson variability

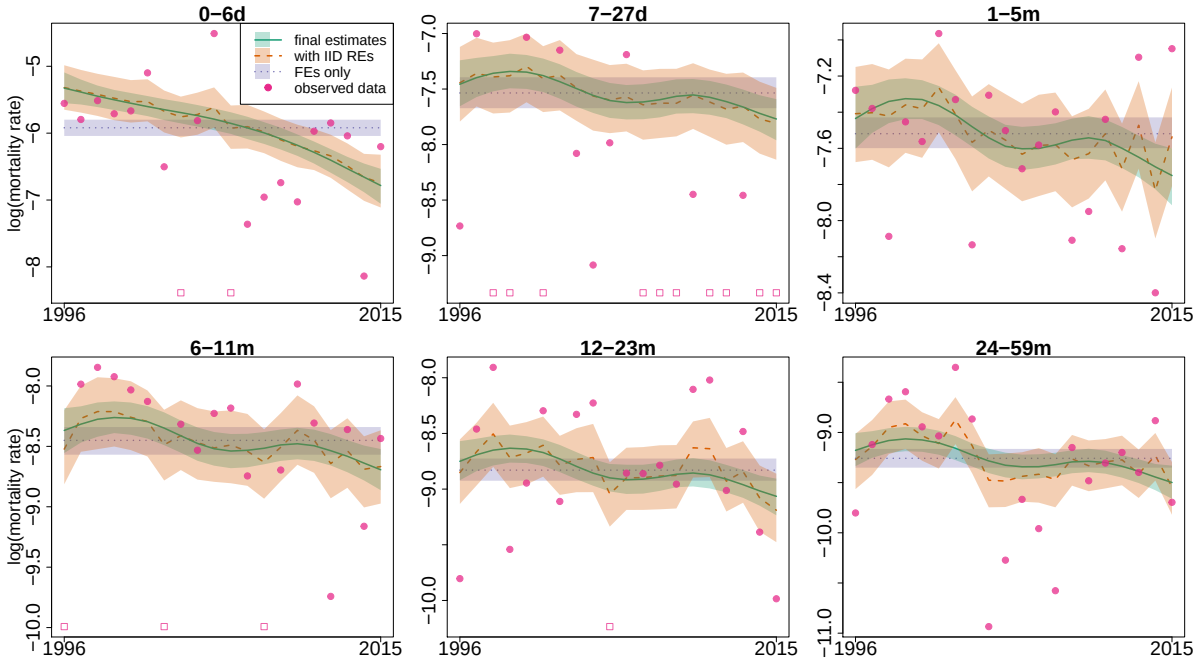


(b) Scenario 2: two correlated causes

Figure 2.2: Relative bias, coverage and width of 95% intervals for log mortality rates estimates from multistage and unified models. For (a), data were generated with IID Normal random effects for each observation and three possible exposure values. For (b), data were generated with bivariate IID Normal random effects for each region-age-year strata, defining the diagonal elements of the covariance matrix as σ^2 and the off diagonal elements as $\rho\sigma^2$. Estimates are averaged over all observations and simulations per scenario.



(a) prematurity in the west rural region



(b) other non-communicable diseases in the east urban region

Figure 2.3: Selected results from the MCHSS data showing empirical data, estimated posterior medians, and posterior 80% intervals for log mortality rates. Combinations with no deaths are represented by an open square.

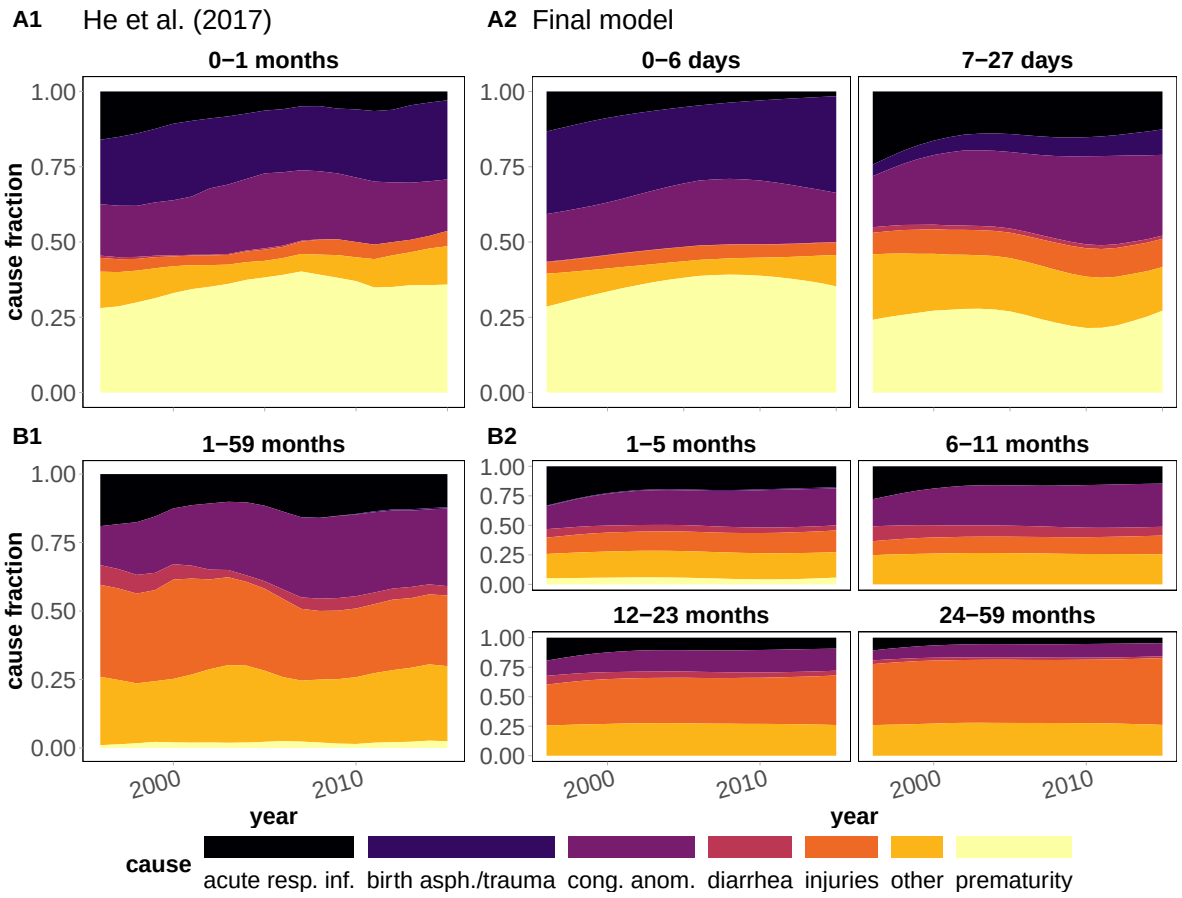


Figure 2.4: Comparisons of estimated CSMFs between our model and the model in He et al. (2017) in the east rural region. Panels A1 and B1 show cause fractions from He et al. (2017) for the 0–1 month and 1–59 month age groups, respectively, while panels A2 and B2 show the estimates from our model for the finer age groups as labeled that are within the broad age groups in A1 and B1.

Chapter 3

AREA-LEVEL MODELS FOR SMALL AREA ESTIMATION OF MULTIVARIATE OUTCOMES FROM COMPLEX SURVEY DATA

3.1 Introduction

Small area estimation (SAE) is an umbrella term for statistical methods that are used to produce estimates of various quantities, such as means and totals, for domains that have sparse or non-existent response data. These domains are typically geographic areas, but may be any subsets of a population defined by variables such as demographic groups. SAE is carried out in many fields, but this chapter centers on estimating health and demographic outcomes in low- and middle-income countries (LMICs), for which SAE can be used to highlight disparities, prioritize resource allocation, and motivate development of interventions. In order to produce reliable estimates for small areas, SAE methods aim to borrow strength using auxillary information. One common way of borrowing strength is modeling with covariates that are shared across domains; another, which is one particular focus of this chapter, is exploiting similarities across geographic areas via spatial models. J. N. Rao & Molina (2015) is the standard reference on SAE, and Pfeffermann (2013) gives an in-depth methodological review. Wakefield, Okonek, & Pedersen (2020) provides an overview specific to health outcomes.

This chapter focuses on SAE using data from surveys with complex design because, as stated in Chapter 2, many LMICs do not have adequate civil registration and vital statistics systems. Consequently, household surveys, such as the Demographic and Health Surveys (DHS) Program (Corsi, Neuman, Finlay, & Subramanian, 2012) and the Multiple Indicator Cluster Surveys (MICS) (Khan & Hancioglu, 2019), are the primary source of nationally-

representative data used to estimate health and population indicators (Boerma & Stansfield, 2007). SAE is typically based on a subset of individuals whose data often come from surveys, although the term is sometimes used when modeling other types of data such as disease counts from registries. One notable subset of SAE is prevalence mapping (Wakefield, 2020), which describes the production of maps displaying counts of disease cases or the prevalence of health and demographic indicators. Prevalence mapping of diseases, or disease mapping, is traditionally based on a complete enumeration of outcomes, in contrast to the predominant use of survey data for SAE.

Using data from complex surveys necessitates accounting for the survey sampling design, which is an important and sometimes overlooked component of SAE. Acknowledging the survey design in SAE has led to *design-based* (or *randomization*) approaches, which are distinct from *model-based* approaches that are in line with much of modern spatial statistics. Skinner & Wakefield (2017) provides a discussion of both approaches. Design-based methods assess frequentist properties of estimators, averaging over all possible samples that could be drawn under a specific sampling design using the randomization distribution. This viewpoint holds responses as fixed rather than random. The distinction is made between *direct* estimators that only use data within a target area for inference, and *indirect* estimators that borrow information from "similar areas" to improve estimation. Additionally, some methods use the design-based framework via model-assisted approaches (Särndal, Swensson, & Wretman, 2003) which allow a model to be specified while retaining desirable design-based properties even with a misspecified model.

Model-based approaches, which can be frequentist or Bayesian, usually condition on the selected sample and inference is with respect to a specified model. These methods can be split into two types: *area-level*, where covariate information and modeling parameters are specified at the desired level of aggregation for final estimates, and *unit-level*, for which data is modeled at the level of the individual or cluster. The seminal Fay-Herriot model (Fay & Herriot, 1979) uses a possibly transformed weighted estimate as the response and uses random effects for each area. Many extensions of the Fay-Herriot model, including the

use of discrete spatial models, have been proposed (Marhuenda, Molina, & Morales, 2013; Mercer et al., 2015; Watjou et al., 2017; Y. You & Zhou, 2011). Crucially, the survey design is accounted for via weighted estimators and their associated variances. This chapter is concerned with area-level modeling, and multivariate Fay-Herriot models in particular. For an approach based on unit-level modeling, see Chapter 4.

Most work on area-level SAE involves single outcomes, which has seen the bulk of recent developments in Fay-Herriot models (Bell, Datta, & Ghosh, 2013; G. S. Datta, Ghosh, Steorts, & Maples, 2011; G. S. Datta, Kubokawa, Molina, & Rao, 2011; Esteban, Herrador, Hobza, & Morales, 2011; Ghosh & Steorts, 2013; González-Manteiga, Lombarda, Molina, Morales, & Santamaría, 2010; Herrador, Esteban, Hobza, & Morales, 2011; Jiang, Nguyen, & Rao, 2011; Kubokawa, 2012; Pfeffermann, Sikov, & Tiller, 2014; Slud & Maiti, 2011; Ybarra & Lohr, 2008). However, the scientific question of interest may involve multiple related outcomes, or we may wish to use information from multiple outcomes to improve the accuracy of predictions. In these settings, jointly modeling multiple outcomes allows for borrowing strength across these outcomes. Some examples include modeling both height and weight for age, cause-specific mortality, road traffic crashes at different severities, and poverty proportions and gaps. Many methods exist for multivariate SAE (Arima, Bell, Datta, Franco, & Liseo, 2017; G. S. Datta, Day, & Maiti, 1998; G. S. Datta, Fay, & Ghosh, 1991; Fay, 1987; Huang & Bell, 2006). Benavent & Morales (2016) surveys this literature while also proposing a multivariate Fay-Herriot model for various linear and mixed model formulations. Some recent developments in multivariate Fay-Herriot models can be found in Saegusa, Sugasawa, & Lahiri (2020) and Franco & Bell (2021).

Much research has been done on multivariate spatial modeling in contexts other than SAE (Carlin & Banerjee, 2003; Carlin & Ma, 2007; Gelfand & Vounatsou, 2003; Jin, Banerjee, & Carlin, 2007; Jin, Carlin, & Banerjee, 2005; Neelon, Gelfand, & Miranda, 2014). One method in particular for discrete spatial analysis is the shared component model, popularized in the bivariate case by Knorr-Held & Best (2001), whereby some parameters of the model are shared among the multiple outcomes. Earnest et al. (2010) proposed a shared compo-

nent conditional autoregressive (CAR) model for a zero-inflated Poisson likelihood. MacNab (2010) surveys the multivariate CAR literature while proposing a general formulation of shared component models that includes allowing for errors in covariates. Additionally, a bivariate shared component model can be reframed as a particular ecological regression model with errors in covariates, which allows the coefficient for the latent shared component to be interpreted as the regression effect of unobserved covariates (Knorr-Held & Best, 2001). In the context of jointly modeling multiple outcomes using data from surveys in LMICs, we often believe that unobserved covariates likely exist that are shared among outcomes, which makes shared component models an attractive option. While some research has been done on multivariate spatial area-level SAE such as Porter, Wikle, & Holan (2015) and Guha & Chandra (2021), these did not investigate shared component models. In particular, multivariate shared component Fay-Herriot models have not been developed, and our research aims to fill this gap.

This chapter proposes a class of multivariate shared component area-level SAE models, focusing on discrete spatial modeling, using complex survey data. We use a Bayesian modeling approach, developing latent Gaussian models that are fit using integrated nested Lapace approximation (INLA). We present two motivating examples that use data from the 2014 Kenya Demographic and Health Survey (2014KDHS): jointly modeling height for age and weight for age in children under age 5, and modeling three categories of contraceptive use in women aged 15-49. The organization of this chapter is as follows. Section 3.2 describes the data for our working example. Section 3.3 provides a brief overview of survey sampling and traditional SAE methods. Section 3.4 details our proposed multivariate shared component model and presents a simulation study for validation. Sections 3.5 and 3.6 describe using this model to produce estimates for our motivating examples, comparing results among univariate models and bivariate non-shared models. Section 3.7 provides a discussion and presents avenues for future research.

3.2 Data

The DHS Program, in collaboration with many LMICs, conducts nationally-representative household surveys that provide data for a wide range of monitoring and impact evaluation indicators in the areas of population, health, and nutrition. The DHS Program uses a set of consistent sampling approaches from country to country, with methods described in the 2012 DHS Sampling and Household Listing Manual (ICF International, 2012, sec. 5.2, p. 80–85). The standard design is a stratified two-stage cluster sampling scheme with stratification by county crossed with urban/rural. The first sampling stage involves selecting enumeration areas (EAs) using probability proportional to size (PPS) sampling, where the probability of sampling each EA is proportional to the listed number of households in that EA, and the second stage involves simple random sampling of (typically) 25 households within each EA. Mothers within the household are then asked a number of questions about their children.

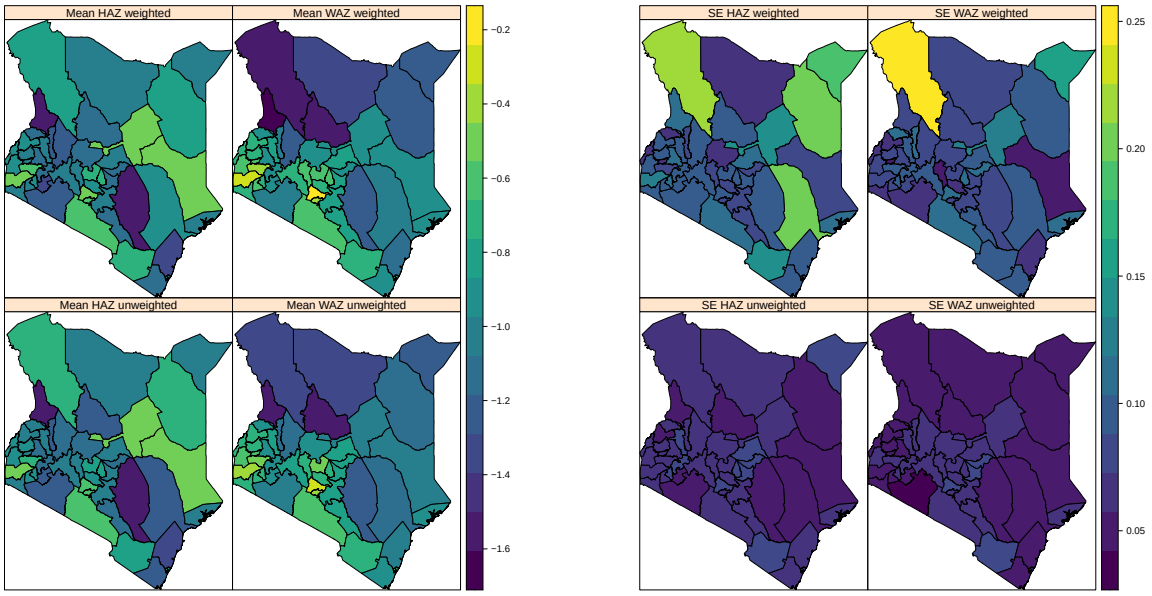
This chapter's motivating examples will use data from the 2014 Kenya DHS (KDHS) (Kenya National Bureau of Statistics, 2015). The 2014 KDHS was designed to produce representative estimates for most of the survey indicators at the national level, for urban and rural areas separately, at the regional (former provincial) level, and for selected indicators at the county level. It follows the typical DHS scheme with 1,612 clusters being sampled out of the 96,251 total EAs that were in the 2009 Kenya Population and Housing Census (Kenya National Bureau of Statistics, 2010). Of these clusters, 995 are urban while 617 are rural, with urban areas oversampled in the majority of the 47 counties. Mombasa and Nairobi are entirely urban and the remaining 45 counties have both urban and rural areas, so that there are 92 strata in total. The interviewers visited only the preselected households, and no replacement of the preselected households was allowed during data collection. Because of the non-proportional allocation to the sampling strata and the fixed sample size per cluster, the survey was not self-weighting. The resulting data include sampling weights in order to be representative at the national, regional, and county levels.

The 2014 KDHS captures many outcomes relating to child health. As a motivating

example for jointly modeling two continuous outcomes, we will use two key child health indicators: height for age z-scores (HAZ) and weight for age z-scores (WAZ), which are measures of child growth. We expect these variables to be correlated (at the individual level, cluster level, and area level) due to shared unobserved risk factors at each of these levels. Previously, HAZ and WAZ were jointly modeled using data from a survey in Papua New Guinea using multivariate CAR models (Gelfand & Vounatsou, 2003). In this analysis, the authors investigated how HAZ and WAZ were associated with dietary factors such as main source of protein, and included joint spatial random effects; however, this analysis did not account for the complex survey design.

HAZ and WAZ scores in the 2014 KDHS were collected for all children born in the 5 years prior to the survey to mothers who were sampled. The final data set contained 18,498 children. Z-scores were calculated using the CDC Standard Deviation-derived Growth Reference Curves derived from the NCHS/FELS/CDC Reference Population (Centers for Disease Control and Prevention, 1996). This reference population, however, is based on data from the USA, which is a limitation for interpreting results of analyses using these data.

We present the county-level univariate naive unweighted estimates (not accounting for the complex survey design) and weighted estimates (accounting for the complex survey design) in Figure 3.1. We see differing spatial gradients between HAZ and WAZ, with a much more consistent north (lower) to south (higher) gradient for WAZ. Additionally, the urban areas in south-central Kenya tend to have both the highest HAZ and WAZ scores. However, the standard errors of the univariate direct estimates (described in Section 3.3) are fairly wide, especially in some of the more rural counties, and the levels are different between HAZ and WAZ. Ignoring the survey design gives national mean estimates (standard error) of -1.00 (0.0099) for HAZ and -0.89 (0.0089) for WAZ, while the weighted estimates are -0.96 (0.016) and -0.78 (0.016), respectively. The weighted estimates are higher because the weights account for the urban oversampling, and urban areas tend to have better health outcomes. The increased standard error is due to the clustering. We present in Figure 3.2 a scatterplot of the county-level design-weighted direct estimates of HAZ versus WAZ with uncertainty.



(a) Mean estimates

(b) Standard errors

Figure 3.1: Univariate direct estimates of means (left two columns) and standard errors (right two columns) for HAZ and WAZ from the 2014 KDHS. We present both naive unweighted estimates that do not account for the complex survey design, and weighted estimates that do.

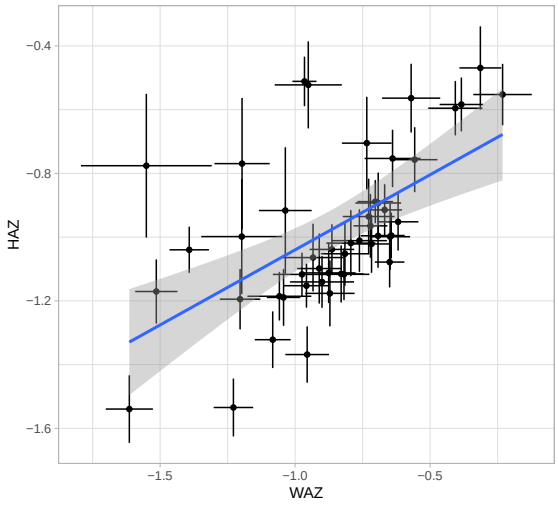


Figure 3.2: Weighted univariate direct estimates of mean HAZ verses WAZ at the county level. Error bars represent plus/minus one standard error. Simple linear regression line with 95% CI is included.

This shows a positive association between the two measures. Additionally, we see a fairly consistent similarity in the amount of uncertainty between HAZ and WAZ in the same county. As a motivating example for multinomial data, we will model current contraceptive use in women aged 15-49 again from the 2014 KDHS. Jointly modeling multinomial data, rather than modeling one category at a time, is important due to the implicit dependencies among multinomial probabilities—they must sum to 1. For contraceptive use in particular, different methods may be used by similar subgroups that have shared, unobserved characteristics in different areas, e.g. due to cultural practices or education, which is further motivation for joint modeling. This survey asked each female respondent their current contraceptive use and categorized it into one of 14 categories. We group these into three broad groups in order to have sufficient observations in each category: *none* (no contraceptive use), *modern* (pill, IUD, injections, condom, female sterilization, male sterilization, implants/norplant, Lactational amenorrhea, female condom, and other modern method), and *other* (periodic abstinence, withdrawal, and other). These are categories as defined by the DHS surveys.

3.3 Background methodology

We begin by reviewing the relevant literature on survey sampling and SAE.

3.3.1 Notation

Let Y_{ric} be an observation of outcome $c = 1, \ldots, C$ for individual $i = 1, \ldots, N_r$ in region $r = 1, \ldots, R$, with N_r the number of individuals in region r (population size). Often, the population mean in area r is of particular interest, $\bar{Y}_{rc} = \sum_{i=1}^{N_r} Y_{ric}/N_r$. In the univariate case, $C = 1$, outcomes Y_{ri1} are typically binary or continuous. In the multivariate case, Y_{ric} typically refer to multiple continuous variables or observations of a categorical variable with C categories.

A survey is conducted to obtain a sample of n_r individuals in each region, a subset of the population. Let I_{ri} be an indicator for membership into the sample. Define π_{ri} to be the first-order inclusion probability, which is the probability that individual i in area r is selected (i.e., $P(I_{ri} = 1) = \pi_{ri}$), and y_{ri} is the observed outcome value. Define the survey design weights as $w_{ri} = 1/\pi_{ri}$, which are the inverses of the inclusion probabilities. Further define π_{rij} to be the second-order inclusion probability, which is the probability that both individuals i and j are selected in area r. Throughout this chapter, we will use bold font and drop subscripts to refer to vectors. Thus, the vector of all observed outcomes for individual i in region r is denoted $\boldsymbol{y}_{ri} = [y_{ri1}, \ldots, y_{riC}]$.

3.3.2 Design-based approaches

We start with an overview of the univariate case and thus we drop the c subscripts for readability.

Design-based methods for estimating the population mean work by averaging over all possible samples that could be drawn under a specific sampling design by using the randomization distribution (the distribution of all possible samples). This viewpoint holds responses as fixed rather than random. *Direct* estimation is a common approach to calculate the mean

of each region whereby only the response data available for the target area is used for estimation.

One popular direct estimator is the Horvitz-Thompson estimator (Horvitz & Thompson, 1952) for the population mean, which is defined as

$$\hat{Y}_r^{HT} = \frac{1}{N_r}\sum_{i=1}^{N_r}\frac{I_{ri}Y_{ri}}{\pi_{ri}} = \frac{1}{N_r}\sum_{i=1}^{n_r}\frac{y_{ri}}{\pi_{ri}} = \frac{1}{N_r}\sum_{i=1}^{n_r} w_{ri}y_{ri}.$$

When population totals are unknown, we can use the Hàjek estimator (Hájek, 1971), which is

$$\hat{Y}_r^{HAJ} = \frac{\sum_{i=1}^{n_r} w_{ri}y_{ri}}{\sum_{i=1}^{n_r} w_{ri}}.$$

Note, for the weights in DHS surveys, we have normalized weights $w_{ri}^* = w_{ri} \times g$, where $g = 1/\sum_{ri} w_{ri}$. In this case, the Hàjek estimator is still a valid estimate of the population mean while the Horvitz-Thompson estimator is not.

Variance estimation for direct estimators can be challenging, as they depend on the joint inclusion probabilities π_{rij} which are often unknown. For many cluster designs, approximations are made in order to estimate the variance. In the context of LMICs, often a jackknife estimator is used (Pedersen & Liu, 2012). In this chapter, we use a generalization of model-robust "sandwich" standard error estimators that are implemented in common software (Lumley, 2004), with derivations provided in Appendix B.1.

Moving beyond direct estimators, another class of design-based approaches is *indirect* estimation whereby information from "similar areas" is borrowed to estimate for a target area. Some examples include synthetic estimators, survey regression, and generalized regression. See Pfeffermann (2013) for details.

3.3.3 Model-based approaches

However, another way of borrowing information is to take a *model-based* approach rather than design based. Model-based approaches are another class of indirect estimators that assume a model for a theoretical infinite superpopulation from which the sample data is drawn, and then use standard statistical theory to derive the optimal (or approximately optimal) predictor of the population mean (or other characteristics). These methods are divided into *area-level* and *unit-level* approaches. Area-level approaches specify covariate information and modeling parameters at the desired level of aggregation for final estimates, while *unit-level* approaches model data at the level of the individual or cluster. Unit-level models are discussed in the following chapter and we focus here on area-level models.

The most popular type of area-level model was first proposed by Fay & Herriot (1979), whose model was defined as

$$\begin{aligned}\hat{y}_r &= \theta_r + e_r \\ \theta_r &= \boldsymbol{X}_r^T \boldsymbol{\beta}_r + u_r,\end{aligned}$$

with $\hat{y}_r$ the direct sample estimator of the population parameter of interest θ_r. For example, this chapter will be using $\hat{y}_r$ as the Hàjek estimator for θ_r the population mean. The e_r terms denote sampling error with assumed zero mean and known area-specific design (randomization) variances v_r^{des}, while the random effects u_i are assumed to be independent with zero mean and variance σ_u^2. Using this model, we can use the standard statistical machinery to produce estimates with uncertainty. Note: a linking model is often defined for a transformation of θ_r; Fay & Herriot (1979) actually assume $\log(\hat{y}_r) = \theta_r + e_r$ and proceed with the above random effects model.

Many extensions of this model have been proposed with more complicated functional forms of the mean model and more complex variance components (Bell et al., 2013; G. S. Datta et al., 1991; G. S. Datta, Ghosh, et al., 2011; G. S. Datta, Kubokawa, et al., 2011;

Esteban et al., 2011; Fay, 1987; Ghosh & Steorts, 2013; González-Manteiga et al., 2010; Herrador et al., 2011; Huang & Bell, 2006; Jiang et al., 2011; Kubokawa, 2012; Marhuenda et al., 2013; Mercer et al., 2015; Pfeffermann et al., 2014; Slud & Maiti, 2011; Watjou et al., 2017; Ybarra & Lohr, 2008; Y. You & Zhou, 2011). Models of this form are commonly referred to as Fay-Herriot models.

Much of the research on Fay-Herriot models has centeres on univariate estimation, but some work has been done on multivariate Fay-Herriot models (Arima et al., 2017; G. S. Datta et al., 1998; Franco & Bell, 2021; Porter et al., 2015; Saegusa et al., 2020). Benavent & Morales (2016) provides a thorough review. A simple multivariate Fay-Herriot model can be defined in two stages. With $\hat{\boldsymbol{y}}_r$ a C length vector of direct estimators for all outcomes in region r, the sampling model is

$$\hat{\boldsymbol{y}}_r = \boldsymbol{\mu}_r + \boldsymbol{e}_r$$

where the vectors $\boldsymbol{e}_r|\boldsymbol{V}_r^{des} \sim N_C(\boldsymbol{0}, \boldsymbol{V}_r^{des})$ are independent with known $C \times C$ design covariance matrices $\boldsymbol{V}_r^{des}$. Then, a model is specified for the $\boldsymbol{\mu}_r$ parameters. For example, González-Manteiga, Lombardía, Molina, Morales, & Santamaría (2008) considered the model

$$\begin{aligned}\boldsymbol{\mu}_r &= \boldsymbol{X}_r\boldsymbol{\beta} + \boldsymbol{1}_C u_r \\ u_r &\overset{IID}{\sim} N(0, \sigma_z^2)\end{aligned}$$

with p_c covariates for each outcome, $p = \sum_{c=1}^{C} p_c$, letting $\boldsymbol{x}_{rc} = [x_{rc1}, \ldots, x_{rcp_c}]$ be a row vector of covariates with $\boldsymbol{X}_r = \text{diag}(\boldsymbol{x}_{r1}, \ldots, \boldsymbol{x}_{rC})_{C \times P}$, and $\boldsymbol{\beta} = [\boldsymbol{\beta}_1^T, \ldots, \boldsymbol{\beta}_C^T]^T$ a matrix of regression coefficients where $\boldsymbol{\beta}_c$ is a column vector of size p_c. Also, $\boldsymbol{1}_C$ is a column vector with all elements equal to 1.

This linking model does not allow correlations between the outcome-specific random effects. Thus, Benavent & Morales (2016) propose an extension that allows for correlations as

$$\boldsymbol{\mu}_r = \boldsymbol{X}_r\boldsymbol{\beta} + \tilde{\boldsymbol{u}}_r$$
$$\tilde{\boldsymbol{u}}_r \sim N_C(\boldsymbol{0}, \boldsymbol{V}_r)$$

and has the $C \times C$ covariance matrices depend on unknown parameters. The models fit with this framework do not include models with spatial random effects, and they are also not developed for multinomial outcomes.

3.3.4 Recommendations for using design- and model-based estimators

When to use design-based or model-based approaches is an active discussion in the SAE literature. Cautiously, Lehtonen & Veijanen (2009) recommend design-based methods for situations with large or medium samples in areas, while sparse data may necessitate model-based approaches. Alternatively, G. S. Datta (2009) argues more favorably toward model-based approaches as a whole. Paige, Fuglstad, Riebler, & Wakefield (2022) conclude that direct estimators are the gold standard when sufficient data allow for an acceptable level of uncertainty, while model-based smoothed direct estimators can aid in reducing this variance, although with the introduction of small bias due to the smoothing; but when direct estimates are unreliable, cluster-level models are required which entails using a model that is consistent with the design and requires careful and thoughtful effort in specifying a model.

Furthermore, some model-based approaches in current fashion are unit-level models that do not fully acknowledge the complex survey design, and do not enjoy the theoretical consistency of design-based approaches (Diggle & Giorgi, 2019; Utazi et al., 2020). Fuglstad, Li, & Wakefield (2021) provides a recent review of SAE that compares direct estimators, area-level, and unit-level models. They conclude similarly to above that sufficient power for direct estimation requires sufficiently large sample sizes, while indirect estimation heavily benefits from reliable auxiliary population information at the target resolution. Therefore, unit-level continuous spatial models have difficulty producing areal estimates without detailed auxiliary population information. In addition, accounting for the sampling design is not

straightforward; this is problematic if the sampling is informative. Furthermore, prevalence mapping often uses model-based geostatistics (MBG) methods that specify a continuous spatial model, which contrasts with this chapter's focus on area-level SAE methods. However, some unit-level SAE approaches do use continuous spatial models (Burstein et al., 2019; Diggle & Giorgi, 2019). Additionally, aggregation of continuous models remains a major challenge, and we recommend spatial modeling at the target level for producing estimates, unless stratification necessitates finer scale modeling. Due to these recommendations on the importance of area-level models that acknowledge the survey design, this chapter will focus on Fay-Herriot models.

The general conclusion is that if you have enough data at the desired area level, then use direct (area-level) estimates. If not, but you still have a sufficiently large sample size to have reliable variance estimates from area-level model-based methods, then use those. Only if neither of the above are true should unit-level models be used, preferably with spatial models specified at the desired area level. Unit-level continuous spatial models require great care if used when the desired level of estimation is at the cluster level as they may have bias and be poorly calibrated.

3.3.5 Spatial modeling

This chapter will primarily center on spatial smoothing models which are commonly used in modeling demographic and health variables using survey data (Manda, Haushona, & Bergquist, 2020). As we are analyzing data at the area-level, only discrete spatial models will be investigated. Banerjee, Carlin, & Gelfand (2014) provides a review of choices for modeling the spatial distribution. Random effect models are commonly used, typically with either conditional autoregressive (CAR) and intrinsic conditional autoregressive (ICAR) distributions. Both of these use information from neighboring areas for local geographical smoothing in order to capture the idea that outcomes are likely to be similar in locations that are close-by. In this chapter, we model area-level spatial effects with a Besag, York and Mol-

lié (BYM) model which includes both an ICAR component for spatial auto-correlation and an independent and identically distributed (IID) normal component for non-spatial heterogeneity (Besag, York, & Mollié, 1991). In the univariate case, the BYM model can be written as $v_r + u_r$, with $v_r|\sigma_v^2 \overset{iid}{\sim} N(0, \sigma_v^2)$ the nonspatial IID normal random effects on region, and $\boldsymbol{u}|\sigma_u^2 \sim ICAR(\sigma_u^2)$ the ICAR spatial random effects, i.e. $u_r|u_{r'}, r' \neq r \sim N(\sum_j w_{rr'}u_r, \sigma_u^2)$.

We choose to use a reparameterized version of the BYM model, which we will call the BYM2 model. The BYM2 model reformulates the IID and ICAR random effects as

$$v_r + u_r = \sigma(\sqrt{1-\rho}v_r^* + \sqrt{\rho}u_r^*).$$

In this formulation, $\sigma = \sqrt{\sigma_v^2 + \sigma_u^2}$ is the standard deviation of the random effects. The unstructured component is $v_r^* \overset{iid}{\sim} N(0, 1)$, which has fixed standard deviation 1. The spatial component is u_r^*, which are ICAR random effects scaled so $\text{Var}(u_r^*) \approx 1$. One way of doing this is by scaling the model so the geometric mean of these variances is 1, using the adjacency matrix to calculate the inverse precision of the ICAR model (Riebler et al., 2016). Lastly, ρ is interpreted as the percent of the total variation that is due to the spatial random effect. This version is an improvement because the parameters are more interpretable, and it allows more intuitive specifications of prior distributions in Bayesian models. This BYM2 model will be the only spatial model considered in this chapter.

Extending discrete spatial models to the multivariate setting is crucial to leverage shared spatial information between outcomes. The simplest case would be to specify separate BYM2 effects for each outcome, i.e. $v_{rc}|\sigma_{vc}^2 \overset{iid}{\sim} N(0, \sigma_{vc}^2)$ and $\boldsymbol{u}_c \sim ICAR(\sigma_{uc}^2)$ for each outcome c. However, many demographic and health outcomes share common but unobserved covariates. This naturally leads to models in the literature called shared component models, for which latent parameters are shared among outcomes (Knorr-Held & Best, 2001). In the bivariate case, a shared component model can be parameterized as in Held, Natário, Fenton, Rue, & Becker (2005) via

$$\mu_{r1} = \beta_1 + z_{r1} + \delta z_{r2}$$
$$\mu_{r2} = \beta_2 + z_{r2}/\delta$$

with δ a scaling parameter and z_{rc} random effects (e.g., BYM2). To aid interpretation of this model, and the δ parameter in particular, Held et al. (2005) links the shared component model to an ecological regression model with errors in covariates in the bivariate case. Suppose g_r is a true but unobserved covariate. We formulate an ecological regression model for outcome 1 as

$$\mu_{r1} = \beta_1 + z_{r1} + g_r\gamma.$$

We then use the second outcome as a surrogate for exposure to the unobserved covariate and assume a measurement error model $\mu_{r2}|g_i \sim N(g_i, \omega)$ with ω the known measurement error precision. Finally, we assume $Y_{r2}|\mu_{r2} \sim N(\mu_{r2}, v)$. Assuming $\gamma > 0$, this ecological regression model is equivalent to the shared component model with

$$g_r = z_{r2}/\delta$$
$$\gamma = \delta^2$$

Thus, the coefficient δ^2 obtained from a joint modelling approach can be interpreted as the regression coefficient that would have been obtained under a measurement error ecological regression model, while the shared component divided by δ can be interpreted as the true, but unknown, ecological covariate. This is an attractive feature of the shared component model because it reflects the nature of demographic and health outcomes having shared, unobserved covariates.

In this chapter, we use a different formulation of the shared component model for simplicity of implementation:

$$\mu_{r1} = \beta_1 + z^*_{r1} + \lambda z^*_{r2}$$
$$\mu_{r2} = \beta_2 + z^*_{r2}.$$

This is equivalent to the model in Held et al. (2005) if we set $\lambda = \delta^2$ and $z^*_{r2} = z_{r2}/\delta$. Theoretically, a model with the first component shared and the final component shared are equivalent as we demonstrate in the Appendix B.2. However, in practice, especially when using Bayesian models, some differences may arise particularly in uncertainty estimation. We explore this further in the contraceptive use example in Section 3.6.

Furthermore, both of these models are different from the model proposed in Knorr-Held & Best (2001), which includes an additional outcome-specific random effect in the second outcome. This formulation is appealing due to it being symmetric, but issues with identifiability arise without strong priors on the precision parameters and/or additional constraints on the random effects.

3.4 A multivariate shared component Fay-Herriot model

Often the outcome, time period, and geographic level of interest for producing estimates of demographic and health indicators using data from the DHS and similar surveys tend to fall into the category of not having sufficient data for direct estimates to be adequately precise, yet having enough data for reliable model-based variance estimates. Thus, model-based area-level methods are of particular importance. In this section, we propose a Fay-Herriot type two-stage modeling approach for multivariate outcomes. In the first-stage, we develop a working likelihood and aggregate estimates to the region level. In the second stage, we fit a mixed effects smoothing model for the region-level estimates produced at the first stage. We describe the model for continuous outcomes—in Section 3.6 we will adapt it a multinomial outcome.

3.4.1 Stage 1 modeling: design-weighted aggregation to the district-level

For the first stage, we will take the individual-level outcomes and calculate survey weighted averages in each region accounting for the sampling probabilities and the complex survey design, along with a design-based estimated covariance matrix for these means. This procedure results in, for each region, a vector of the survey-weighted mean outcomes, as well as a design-based estimate for the covariance of this random vector.

First, for each area we will calculate a vector of means of each outcome and its associated survey design-based covariance matrix using weighted estimates and the appropriate variance. To produce mean estimates in each region, we will calculate the Hàjek mean estimate as

$$\hat{\bar{y}}_{rc} = \frac{\sum_{i=1}^{n_r} w_{ri}^* y_{ric}}{\sum_{i=1}^{n_r} w_{ri}^*}$$

We then calculate the estimated design-based covariance matrix of $\hat{\bar{\boldsymbol{y}}}_r = (\hat{\bar{y}}_{r1}, \ldots, \hat{\bar{y}}_{rC})$ in each region, which we will denote as

$$\hat{\boldsymbol{V}}_r^{des} = \begin{bmatrix} \hat{V}_{r11}^{des} & \hat{V}_{r12}^{des} & \ldots & \hat{V}_{r1C}^{des} \\ \hat{V}_{r21}^{des} & \hat{V}_{r22}^{des} & \ldots & \hat{V}_{r2C}^{des} \\ \vdots & \vdots & \ddots & \vdots \\ \hat{V}_{rC1}^{des} & \hat{V}_{rC2}^{des} & \ldots & \hat{V}_{rCC}^{des} \end{bmatrix}.$$

This estimate must account for the complex stratified cluster design of the survey. The form of $\hat{\boldsymbol{V}}_r^{des}$ and its derivation are provided in Appendix B.1. We note that these direct estimates and design-based covariance matrices are estimated separately for each region.

3.4.2 Stage 2 modeling: Bayesian smoothing model

For the second stage, we will fit a smoothing model on these region-specific estimates that uses the first stage mean vectors, $\hat{\bar{\boldsymbol{y}}}_r$, as the outcome, and that takes as fixed the asymptotic

normal design-based covariance matrices from the first stage, $\hat{\boldsymbol{V}}_r^{des}$. The smoothing model can be flexibly specified depending on the data available and the scientific question of interest; we envision that it will typically be a mixed effects model including area-level covariates along with spatial random effects.

Due to the amount of data, we will assume that the mean outcomes in a region have a multivariate normal distribution with variance equal to $\hat{\boldsymbol{V}}_r^{des}$. We propose models with the working likelihood

$$\hat{\tilde{\boldsymbol{y}}}_r | \boldsymbol{\mu}_r, \hat{\boldsymbol{V}}_r^{des} \sim N_2(\boldsymbol{\mu}_r, \hat{\boldsymbol{V}}_r^{des}) \tag{3.1}$$

with $\hat{\tilde{\boldsymbol{y}}}_r$ playing the role of "observed" data and $\hat{\boldsymbol{V}}_r^{des}$ assumed fixed and known. Then, we specify a mean smoothing model on the latent parameters $\boldsymbol{\mu}_r$. The smoothing models that we investigate in this chapter will include BYM2 random effects to leverage spatial information, and will also be formulated as shared component models in order to borrow strength among outcomes.

3.4.3 Model fitting

A Bayesian approach was used in this chapter. Models were fit within the R statistical computing environment (R Core Team, 2013). First stage models were fit using the `survey` package (Lumley, 2004). Appropriately survey-weighted direct estimates of the area-level means were calculated using the `svymean()` function, and design-based covariance matrices were extracted via the `vcov()` function applied to the output of `svymean()`. The second-stage Bayesian smoothing models were fit using Integrated Nested Laplace Approximation (INLA) (Rue et al., 2009) as implemented in the `INLA` package. INLA provides a fast alternative to MCMC for approximating the marginal posterior distributions of latent Gaussian Markov random field models. However, since the `INLA` package does not provide a bivariate likelihood, we must instead specify the a univariate Gaussian likelihood and add in bivariate Gaussian IID random effects for each region that have fixed

covariances equal to $\hat{\boldsymbol{V}}_r^{des}$. The models in this chapter are easy to specify and fast to compute using personal computing machines. All code used in this chapter is available at https://github.com/aeschuma/multivariate-sae.

3.4.4 Model selection

Model selection among candidate models is an important task in any modeling task. For Fay-Herriot models, the first stage can be considered as a "data-processing" step in some sense, as the performance of the first stage direct estimates and covariance matrix have been validated in previous research (Mercer et al., 2015). Thus in this chapter, we will take the first stage direct estimates as if they were observed data and perform model selection for a suite of candidate models in the second stage. Models will be compared via leave-one-out cross validation using a multivariate scoring strategy that accounts for the correlation between HAZ and WAZ. For region r, we define $\ell_r = \pi(\boldsymbol{y}_r|\boldsymbol{y}_{-r})$ as the posterior marginal likelihood from a model fit with region r removed evaluated at the true value of the held out data—in this case, the true value of the held out data are the direct estimates of the held out region. We perform the following steps for each region:

1. Hold out the direct estimates for region r.
2. Fit all candidate models to the direct estimates with region r removed.
3. Draw $S = 1000$ samples from the joint marginal distributions $\boldsymbol{\mu}_r^{(s)} = \pi(\boldsymbol{\mu}_r|\boldsymbol{y}_{-r})$.
4. Calculate $\ell_r = \frac{1}{S}\sum_{s=1}^{S} p(\hat{\boldsymbol{y}}_r|\boldsymbol{\mu}_r^{(s)}, \hat{\boldsymbol{V}}_r^{des})$, where $p(\boldsymbol{y}_r^{(s)}|\boldsymbol{\mu}_r^{(s)}, \hat{\boldsymbol{V}}_r^{des}) = N_2(\boldsymbol{\mu}_r^{(s)}, \hat{\boldsymbol{V}}_r^{des})$.
5. Calculate $-\sum_{r=1}^{R}\log(\ell_r)$, which we will call the *LogScore*.

We will compare the LogScore of each model. Since the LogScore is based on the negative log-likelihood, lower values indicate better performing models.

3.4.5 Simulation study to evaluate shared component models

We perform a simulation study to show the validity of our modeling procedure and explore how model misspecification of the spatial model and shared component affects results. We

will generate bivariate data similar to our working example of HAZ and WAZ scores in the 2014 KDHS, but rather than simulate individual-level data and fit a first stage model, we will instead simulate first-stage direct estimates and their associated variance estimates from their sampling distributions. Then, we will fit second-stage smoothing models using these as inputs. We will do this for nine different scenarios each with their own data generating mechanism. In each scenario, we will have seven different second stage smoothing models and compare the bias and coverage of the estimated latent area-level means. All scenarios will simulate 1000 datasets, fit the seven comparison models, and calculate the bias, absolute bias, and relative absolute bias, along with the width and coverage of both 80% and 95% credible intervals. Each scenario will generate latent area-level means, $\boldsymbol{\mu_r} = [\mu_{r1}, \mu_{r2}]$, using different mechanisms. Once these means are generated, we will simulate the area-level sample means as

$$\bar{\boldsymbol{y}}_r | \boldsymbol{\mu}_r, \boldsymbol{V}_r \sim N_2(\boldsymbol{\mu}_r, \boldsymbol{V}_r)$$

where the $\boldsymbol{V}_r$ are set to be equal to the estimated asymptotic design-based covariance matrix of the area-level mean HAZ and WAZ from the 2014 KDHS data.

Now, for an SRS, the sampling distribution of the area-level sample covariance matrices is

$$(n_r - 1)\hat{\boldsymbol{V}}_r^{srs} | \boldsymbol{V}_r, n_r \sim \text{Wishart}\left(\boldsymbol{V}_r, (n_r - 1)\right)$$

where n_r is the sample size in area r. The 2014 KDHS uses a stratified cluster design rather than an SRS, which has a different sampling variance than an SRS. In the univariate case, the ratio of the variance for a statistic calculated using a specific survey design to the variance of that statistic calculated using an SRS of the same sample size is called the design effect,

$$d^2 = \frac{\hat{V}}{\hat{V}^{srs}}.$$

Table 3.1: Candidate models to estimate HAZ and WAZ from the 2014 KDHS

Model #	Model name	Stage 1 variance	Stage 2 linear predictor
O	Bivariate Direct	$\begin{bmatrix} \hat{V}_{r11}^{des} & \hat{V}_{r12}^{des} \\ \hat{V}_{r21}^{des} & \hat{V}_{r22}^{des} \end{bmatrix}$	—
I	Univariate IID	$\begin{bmatrix} \hat{V}_{r11}^{des} & 0 \\ 0 & \hat{V}_{r22}^{des} \end{bmatrix}$	$\beta_1 + v_{r1}$ $\beta_2 + v_{r2}$
II	Univariate BYM	$\begin{bmatrix} \hat{V}_{r11}^{des} & 0 \\ 0 & \hat{V}_{r22}^{des} \end{bmatrix}$	$\beta_1 + v_{r1} + u_{r1}$ $\beta_2 + v_{r2} + u_{r2}$
III	Bivariate nonshared IID	$\begin{bmatrix} \hat{V}_{r11}^{des} & \hat{V}_{r12}^{des} \\ \hat{V}_{r21}^{des} & \hat{V}_{r22}^{des} \end{bmatrix}$	$\beta_1 + v_{r1}$ $\beta_2 + v_{r2}$
IV	Bivariate nonshared BYM	$\begin{bmatrix} \hat{V}_{r11}^{des} & \hat{V}_{r12}^{des} \\ \hat{V}_{r21}^{des} & \hat{V}_{r22}^{des} \end{bmatrix}$	$\beta_1 + v_{r1} + u_{r1}$ $\beta_2 + v_{r2} + u_{r2}$
V	Bivariate shared IID	$\begin{bmatrix} \hat{V}_{r11}^{des} & \hat{V}_{r12}^{des} \\ \hat{V}_{r21}^{des} & \hat{V}_{r22}^{des} \end{bmatrix}$	$\beta_1 + v_{r1} + \lambda(v_{2r})$ $\beta_2 + v_{r2}$
VI	Bivariate shared BYM	$\begin{bmatrix} \hat{V}_{r11}^{des} & \hat{V}_{r12}^{des} \\ \hat{V}_{r21}^{des} & \hat{V}_{r22}^{des} \end{bmatrix}$	$\beta_1 + v_{r1} + u_{r1} + \lambda(v_{r2} + u_{r2})$ $\beta_2 + v_{r2} + u_{r2}$

where $\hat{V}$ and $\hat{V}^{srs}$ are the univariate variances.

In a typical DHS, the average design effect across all indicators is $d^2 = 1.5^2 = 2.25$ (see https://userforum.dhsprogram.com/index.php?t=msg&goto=3448&S=Google). Then, the effective sample size in the stratified design is approximately equal to $n_r^* = \frac{n_r}{d^2}$, since the sample variance is linear with respect to the sample size. Thus, we calculate n_r^* as the observed sample size in region r divided by 2.25 and use this to adjust for the survey design effect. This scaling of variances extends to the bivariate case, giving us

$$\hat{\boldsymbol{V}}_r | \boldsymbol{V}_r, n_r^* \sim \frac{1}{(n_r^* - 1)} \text{Wishart}\left(\boldsymbol{V}_r, (n_r^* - 1)\right).$$

In each simulation within a scenario, we will have the same values of μ_r and V_r, and we use these to generate $\bar{\boldsymbol{y}}_r$ and $\hat{\boldsymbol{V}}_r$. We will treat the $\bar{\boldsymbol{y}}_r$ as pseudo-direct estimates and calculate asymptotic 95% confidence intervals as $\hat{y}_{rc} \pm z_{0.975}\sqrt{\hat{V}_{r,cc}}$ with $\hat{V}_{r,cc}$ the diagonal entries of the sampled covariance matrix.

These pseudo-direct estimates and their simulated covariance matrix will be the first of the seven comparison models. The other six will match the data generating mechanisms for Scenarios 1 through 6 (described below), and they are also listed in Table 3.1. All comparison models will be fit to the same data for each simulation using Bayesian models with vague priors fit in INLA.

We will now describe the data generating mechanisms for the latent means μ_{rc} in each of the nine different scenarios. Let β_c be fixed intercepts for each outcome $c = 1, 2$. With $r = 1, \ldots, R$ indexing regions, let v_{rc} and u_{rc} be IID normal and ICAR random effects, respectively. When data is generated and models are fit with both spatial and nonspatial random effects, we will use the reparameterized BYM2 model as previously described.

In all scenarios, the parameter values used to generate the latent means will be set equal to parameters estimated from the described model fit to the univariate direct estimates of HAZ and WAZ from the 2014 KDHS.

The nine scenarios are as follows:

Scenario 1, Univariate IID, 2014 KDHS parameters: This scenario will set the diagonal entries of the $\boldsymbol{V}_r$ to be the diagonal entries of the covariance matrices from the first stage direct estimates of the 2014 KDHS HAZ and WAZ data, and the off-diagonal elements will be set to 0 in all regions, which is equivalent to separate univariate generating mechanisms for HAZ and WAZ. The latent area means will be generated as

$$\mu_{r1} = \beta_1 + v_{r1}$$
$$\mu_{r2} = \beta_2 + v_{r2}.$$

Scenario 2, Univariate BYM, 2014 KDHS parameters: This scenario will use the same $\boldsymbol{V}_r$ as Scenario 1. The latent area means will be generated as

$$\mu_{r1} = \beta_1 + v_{r1} + u_{r1}$$
$$\mu_{r2} = \beta_2 + v_{r2} + u_{r2}.$$

Scenario 3, Bivariate nonshared IID, 2014 KDHS parameters: This scenario will set $\boldsymbol{V}_r$ to be the covariance matrices from the first stage direct estimates of the 2014 KDHS HAZ and WAZ data. The latent area means will be generated as

$$\mu_{r1} = \beta_1 + v_{r1}$$
$$\mu_{r2} = \beta_2 + v_{r2}.$$

Scenario 4, Bivariate nonshared BYM, 2014 KDHS parameters: This scenario will use the same $\boldsymbol{V}_r$ as Scenario 3. The latent area means will be generated as

$$\mu_{r1} = \beta_1 + v_{r1} + u_{r1}$$
$$\mu_{r2} = \beta_2 + v_{r2} + u_{r2}.$$

Scenario 5, Bivariate shared IID, 2014 KDHS parameters: This scenario will use the same $\boldsymbol{V}_r$ as Scenario 3. The latent area means will be generated as

$$\mu_{r1} = \beta_1 + v_{r1} + \lambda v_{r2}$$
$$\mu_{r2} = \beta_2 + v_{r2}.$$

Scenario 6, Bivariate shared BYM, 2014 KDHS parameters: This scenario will use the same $\boldsymbol{V}_r$ as Scenario 3. The latent area means will be generated as

$$\mu_{r1} = \beta_1 + v_{r1} + u_{r1} + \lambda(v_{r2} + u_{r2})$$
$$\mu_{r2} = \beta_2 + v_{r2} + u_{r2}.$$

Scenario 7, Bivariate shared BYM, larger first stage variances: This scenario will set the correlations of the $\boldsymbol{V}_r$ to be the correlations of the covariance matrices from the first stage direct estimates of the 2014 KDHS HAZ and WAZ data, and the diagonal elements will be set to range between 0.1 and 0.2, which is a range that is approximately ten-fold greater than estimates from the 2014 KDHS. The latent area means will be generated the same as Scenario 6.

Scenario 8, Bivariate shared BYM, smaller first stage variances: This scenario will set the correlations of the $\boldsymbol{V}_r$ to be the correlations of the covariance matrices from the first stage

direct estimates of the 2014 KDHS HAZ and WAZ data, and the diagonal elements will be set to range between 0.001 and 0.002, which is a range that is approximately ten-fold smaller than estimates from the 2014 KDHS. The latent area means will be generated the same as Scenario 6.

Scenario 9, Bivariate shared BYM, one precise and one imprecise outcome that are correlated: This scenario will set the correlations of the $\boldsymbol{V}_r$ to be 0.8, which is in the higher end of the range of correlations estimated from the 2014 KDHS. The variances in each region for the first outcome will be set to 0.001 (which represents a well-measured outcome with little intrinsic variability), while the variances for the second outcome will be set to 0.2 (representing a poorly measured outcome). This reflects a useful scenario for which joint models are useful—the idea is that modeling the two outcomes together will allow precise estimates of the first outcome to aid estimation of the second.

Simulation results: Tables of simulation results are presented in Appendix B.3. We graph the bias, relative bias, variance, and mean squared error (MSE) of the estimated latent means for each outcome, along with the coverage and width of 95% credible intervals, averaged over all regions, in Figure 3.3 for scenarios 1 through 6 and Figure 3.4 for scenarios 7 through 9. Starting with scenarios 1–6, we see little bias in general. The direct estimates are the least biased, but have the highest variance and MSE, as well far and away the widest intervals. In scenarios 1–3, all smoothing models perform comparably in terms of variance, MSE, coverage, and width. In scenario 4, in which data is generated via a bivariate nonshared BYM mechanism, the bivariate nonshared BYM and bivariate shared BYM models have the lowest MSE and similar close to nominal coverage, but the shared model actually has the narrowest intervals which indicates it performing better than the correctly specified model. In scenarios 5 and 6, the bivariate shared IID and bivariate shared BYM data generating mechanisms, we see that the bivariate nonshared models have undercoverage while the other

models all perform well. Additionally, the bivariate shared BYM model performs best when it is the correct model, and the bivariate shared IID model performs best when it is the correct model, but these models are quite similar in both scenarios. This stresses the importance of accounting for a shared component when there indeed is one and you are fitting a bivariate model. From these results, we can conclude that with data the size of the 2014 KDHS with outcomes similar to HAZ and WAZ, the choice of second stage model is fairly unimportant as long as a shared component is specified when one truly exists—but some sort of second-stage smoothing is beneficial compared to the direct estimates in order to narrow the uncertainty intervals without adding bias and still preserving close to nominal coverage.

In scenarios 8 and 9, we draw similar conclusions as scenarios 1–6. This means that for scenario 9, we do not see that a clearly better performance of the shared component model, because the bulk of the work accounting for the correlation between the outcomes is done in the first stage model which is the same in all bivariate models. However, the univariate models do not perform too poorly, which means that the utility of these models isn't too important in this particular scenario with one large and one small variance for two correlated outcomes. Further simulation scenarios may be able to tease out specific scenarios where a bivariate shared modeling approach has better performance compared to incorrectly specified models.

Scenario 7, with much higher first stage variability, is also interesting. This scenario mimics having much less data than the 2014 KDHS and/or having outcomes with much higher variability (e.g., measured with high imprecision). The direct estimates again have the least overall bias but the other models are still relatively unbiased. The direct estimates still have the largest variance, MSE, and interval width, and the bivariate shared BYM model has the least variance, MSE, and interval width, although the other models are close. We again see that the bivariate nonshared models have the narrowest uncertainty intervals, and these are too narrow as they also demonstrate substantial undercoverage. The univariate BYM and bivariate shared BYM models are the best performing, with slightly better coverage for the univariate BYM but narrower intervals for the bivariate shared BYM model. So we

see that the most important modeling choice is the spatial component, and to be sure that if the data actually has a shared component, it is better to not use a nonshared model.

3.5 Area-level modeling of HAZ and WAZ from the 2014 Kenya DHS

We model HAZ and WAZ scores from the 2014 KDHS as our first working example. We will present the modeling specifications for seven candidate models, compare them using a bivariate scoring approach, and then present detailed results for the best performing model.

3.5.1 Modeling specifications

This section presents a set of seven candidate models which will all use the same first stage model but will have different second stage Bayesian smoothing models. The names, first stage variance structure, and form of the linear predictors are listed in Table 3.1.

The first stage model will be the bivariate case of the model in Section 3.4.1. Specifically, with y_{ir1} and y_{ir2} the HAZ and WAZ scores, respectively, for individual i in region r, our first stage model estimates are

$$\hat{\bar{y}}_{r1} = \frac{\sum_{i=1}^{n_r} w_{ri} y_{ri1}}{\sum_{i=1}^{n_r} w_{ri}}$$
$$\hat{\bar{y}}_{r2} = \frac{\sum_{i=1}^{n_r} w_{ri} y_{ri2}}{\sum_{i=1}^{n_r} w_{ri}}$$

with asymptotic normal likelihood

$$\hat{\bar{\boldsymbol{y}}}_r | \boldsymbol{\mu}_r, \hat{\boldsymbol{V}}_r^{des} \sim N_2\left(\boldsymbol{\mu}_r, \hat{\boldsymbol{V}}_r^{des}\right) \tag{3.2}$$

where the estimated covariance matrix is denoted

$$\hat{\boldsymbol{V}}_r^{des} = \begin{bmatrix} \hat{V}_{r11}^{des} & \hat{V}_{r12}^{des} \\ \hat{V}_{r21}^{des} & \hat{V}_{r22}^{des} \end{bmatrix}.$$

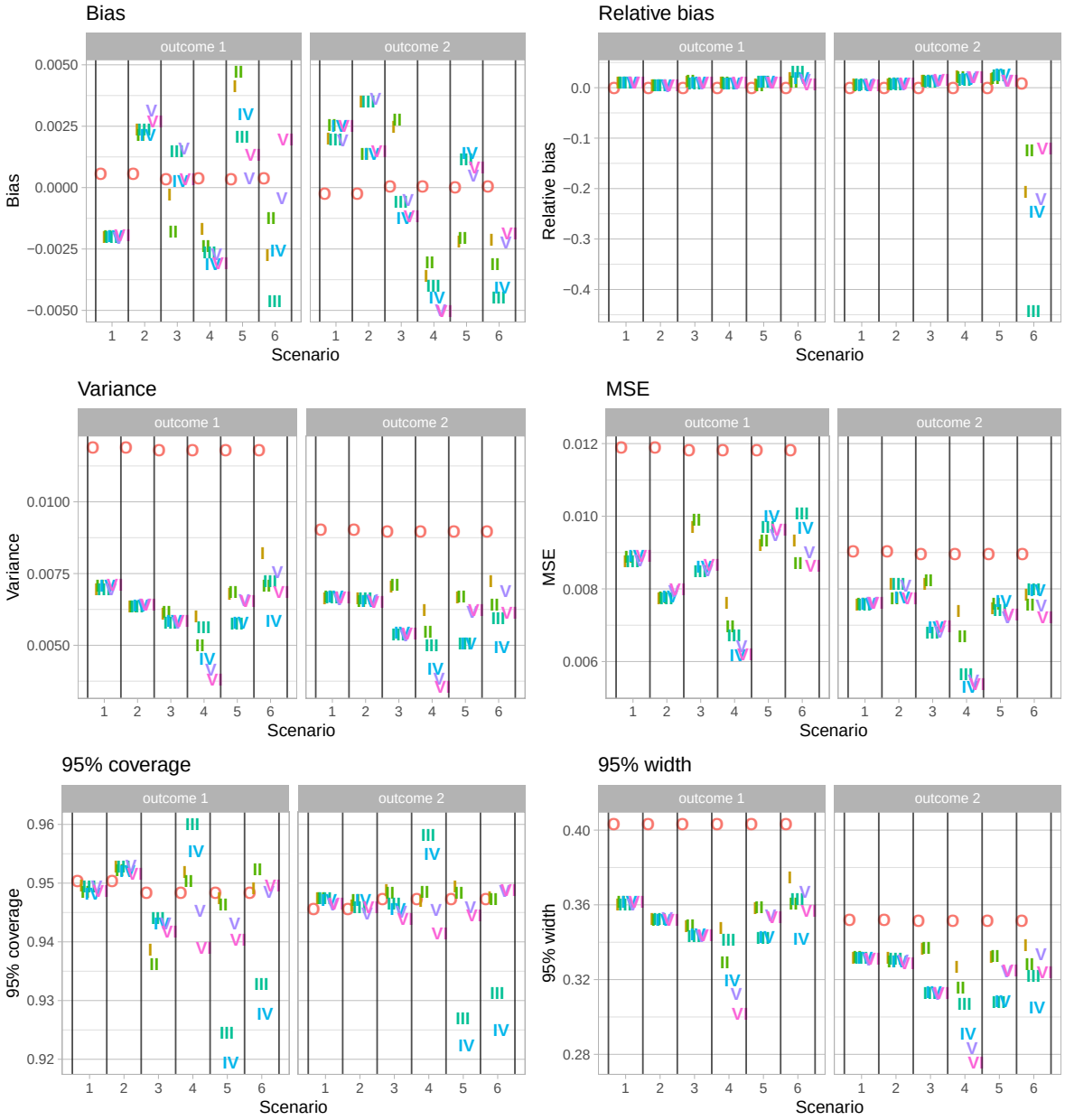


Figure 3.3: Bias, absoulte bias, variance, MSE, 95% interval coverage, and 95% interval width for all models across simulation scenarios 1–6. Scenario descriptions: 1 - univariate IID; 2 - univariate BYM; 3 - bivariate nonshared IID; 4 - bivariate nonshared BYM; 5 - bivariate shared IID; 6 - bivariate shared BYM. Models: O - direct estimates; I - univariate IID; II - univariate BYM; III - bivariate nonshared IID; IV - bivariate nonshared BYM; V - bivariate shared IID; VI - bivariate shared BYM.

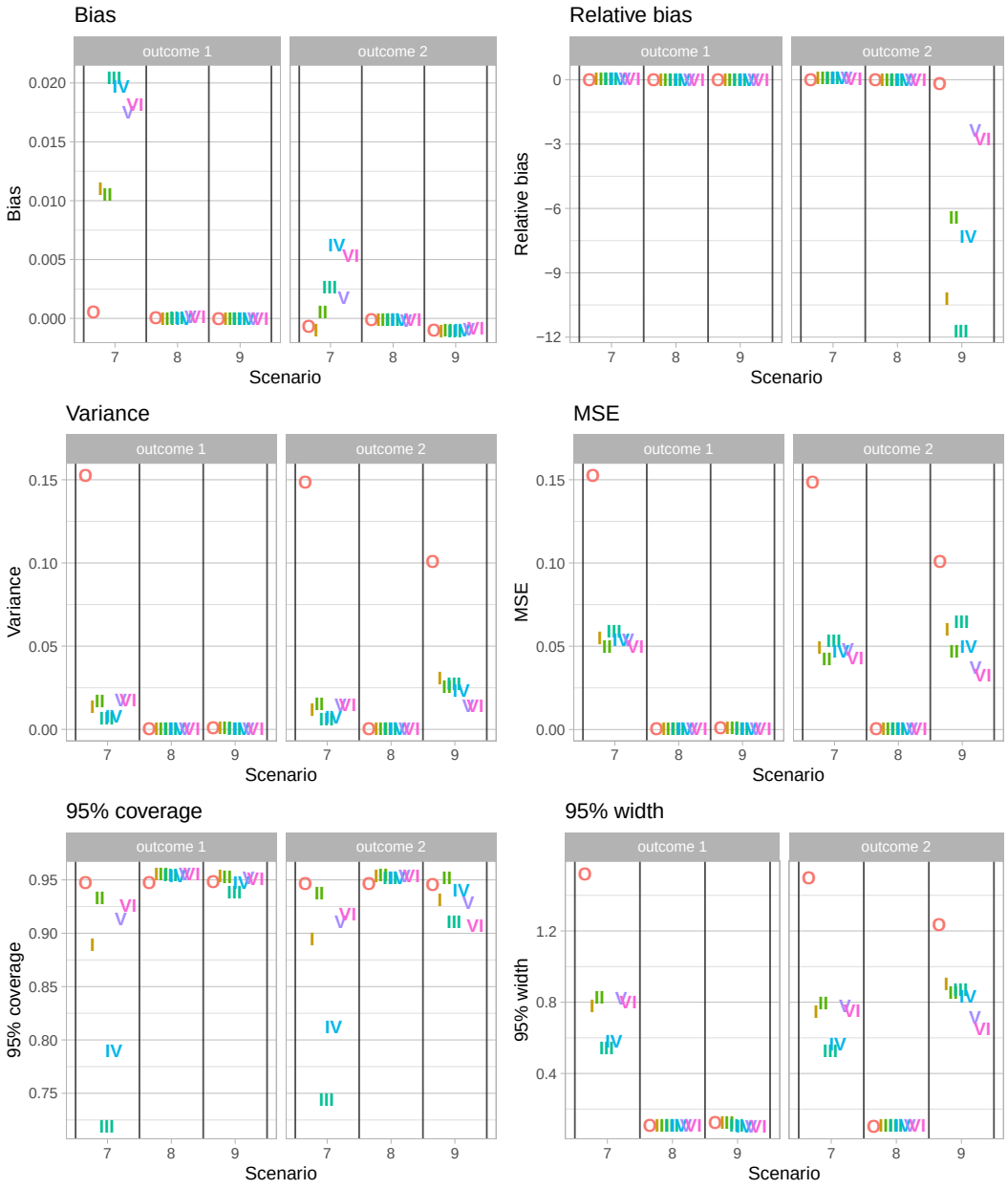


Figure 3.4: Bias, absoulte bias, variance, MSE, 95% interval coverage, and 95% interval width for all models across simulation scenarios 7–9. Scenario descriptions: 7 - bivariate shared BYM, large first stage variances; 8 - bivariate shared BYM, small first stage variances; 9 - bivariate shared BYM, correlated one large and one small first stage variance. Models: O - direct estimates; I - univariate IID; II - univariate BYM; III - bivariate nonshared IID; IV - bivariate nonshared BYM; V - bivariate shared IID; VI - bivariate shared BYM.

Now, we will describe the second stage models. The functional forms are written in more detail in Table 3.1. Candidate model 0 is just the first stage direct estimates and thus has no second stage smoothing. Candidate models I and II are univariate second stage smoothing models, which means we model HAZ and WAZ scores separately. In order to facilitate model comparisons, we will instead jointly fit the univariate models in a single bivariate model but set $\hat{V}_{r12}^{des} = \hat{V}_{r21}^{des} = 0$, which is equivalent to separately fitting two univariate models. Model I will have IID random effects on region while model II will perform spatial smoothing via BYM random effects. Candidate models III and IV are bivariate second stage smoothing models that do not have shared components. Model III will use IID random effects only and model IV will use BYM random effects. Candidate models V and VI are bivariate second stage smoothing models with a shared component. Model V will parameterize the latent means of WAZ with IID random effects only, and it will parameterize the latent means of HAZ as IID random effects and a coefficient multiplied by the same IID random effects from the WAZ latent means. Model VI will parameterize the latent means of WAZ with BYM random effects only, and it will parameterize the latent means of HAZ as BYM random effects and a coefficient multiplied by the ICAR component of the WAZ BYM random effects. All BYM random effects are parameterized using the formulation in Riebler et al. (2016) as previously described.

As we are using Bayesian inference, we must specify prior distributions. Since this example has a large amount of data, we use relatively noninformative priors. For the fixed effects we use improper flat priors. For the IID random effects in IID-only models (I, III, and V), we use penalized complexity (PC) priors (Simpson et al., 2017) on the standard deviations such that there is a 1% chance to be larger than 1. For BYM random effects, we use PC priors on the total variance of the IID and ICAR random effects such that there is a 1% probability that the total variance is greater than 1. We use Beta(1, 1) prior on ρ_c, the percent of the variation that is spatial for each of the c outcomes, as recommended in Riebler et al. (2016).

3.5.2 Results

First, we present the bivariate correlations estimated from the first stage model in Figure. The correlations range from 0.34 to 0.95. This indicates fairly strong correlations between HAZ and WAZ at the individual and cluster levels, which is what the first stage correlations are capturing (the shared component in the second stage models will help capture the correlation between HAZ and WAZ among regions, i.e. due to unobserved covariates). This range of correlations demonstrates the potential for improved estimates due to multivariate modeling. We also note that the correlations are higher in the rural regions and lower in the urban regions, which lends some credence to there being unobserved covariates shared between HAZ and WAZ.

We compare the posterior medians of the latent mean HAZ and WAZ scores among all models, and these are presented in Figure 3.6. The comparisons are separated into univariate comparisons, bivariate IID comparisons, and bivariate BYM comparisons. Here, we see that posterior median estimates of all of the IID models generally agree, and the same is true for all BYM models. Comparing IID models to BYM models, we see the impact of the spatial smoothing in shrinking the estimates toward the overall mean—however, since there is an abundance of data, this shrinkage is less pronounced than would be seen when analyzing data from a smaller survey.

We present a comparison of the posterior standard deviations of the latent means in Figure 3.7. Here, we see that the nonshared BYM and nonshared IID models have the smallest posterior standard deviations. However, as discussed in Section 3.4.5, the smaller variance leads to narrower uncertainty intervals in nonshared models which are incorrectly anticonservative in scenarios for which there is a true shared component (for example, if there are unobserved covariates shared between the outcomes).

A comparison of the LogScore (Section 3.4.4) for each of the seven candidate models is provided in Table 3.2. Recall the LogScore is akin to the average of the posterior predictive negative log likelihood, and that a lower score means a better performing model. We see

that Model VI, the bivariate shared BYM model, performs best. This makes sense since there are likely shared unobserved factors that influence both HAZ and WAZ that we are not modeling, so a shared component is an intuitively reasonable choice. We also note that we fit a bivariate shared BYM model with the HAZ component shared, rather than the WAZ component shared, and estimates were nearly identical.

Due to these results, we present our final estimates of mean HAZ and WAZ scores from the 2014 KDHS from the results of model VI. Summaries of the posterior distributions of parameter estimates are presented in Table 3.3. Figure 3.8 shows a map of the posterior median HAZ and WAZ scores along with the associated widths of the 95% credible intervals. We see a clearer spatial gradient for median WAZ compared to HAZ. This is also reflected in the ρ_1 and ρ_2 estimates, for which WAZ had a slightly higher proportion of the random effect variance that was spatial (91.0% vs. 86.5%), although the uncertainty intervals are wide. We also see that while areas of high uncertainty in HAZ generally correspond to high uncertainty in WAZ, there are still numerous regions with substantially higher HAZ uncertainty than WAZ uncertainty and vice versa.

We plot maps of the IID and ICAR random effects for both HAZ and WAZ in Figure 3.9. The differences in the spatial gradients of the HAZ ICAR and the shared ICAR components is particularly striking. This is because the shared ICAR component is modeling the spatial variation in WAZ, while the HAZ ICAR component is modeling the residual spatial variation beyond that which is shared with WAZ. Because WAZ has a much stronger spatial gradient, we see near opposite gradients. We provide a more in-depth discussion of this phenomenon when we examine the unit-level model for these data in Chapter 4.

3.6 Area-level modeling of contraceptive use from the 2014 Kenya DHS

As our next motivating example, we model at contraceptive use in women aged 15-49 also from the 2014 KDHS data. Contraceptive use was categorized into three categories: none, modern, and other. First, we will describe how to adapt our multivariate Fay-Herriot model

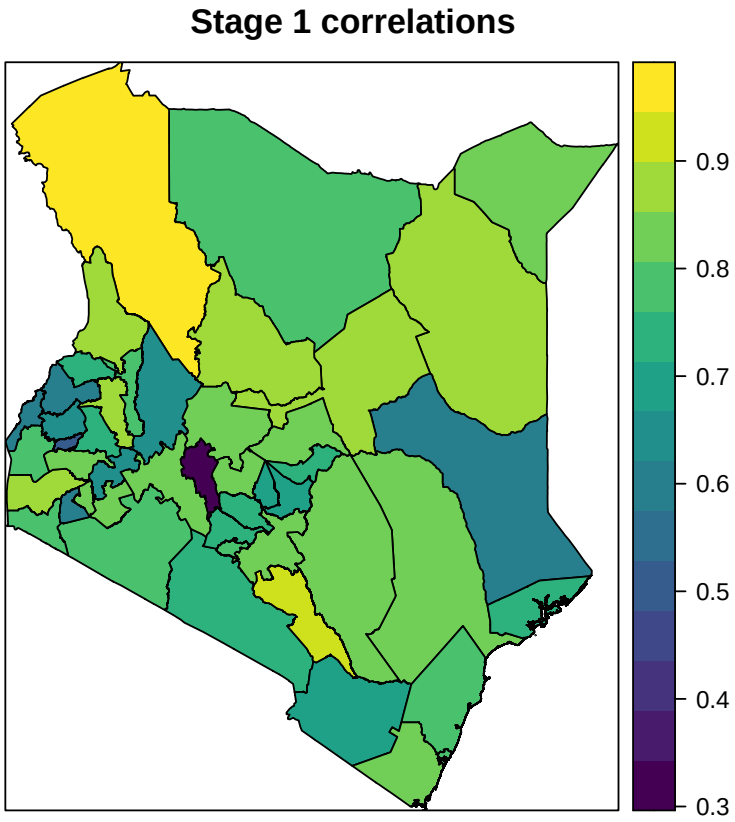
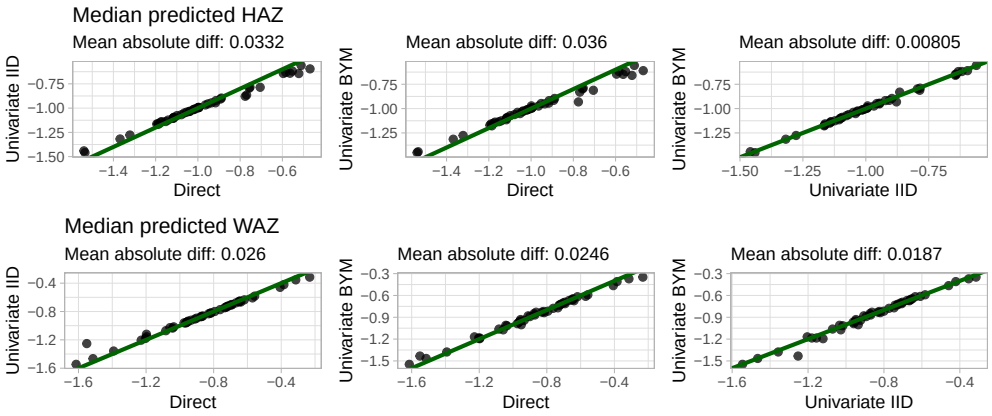
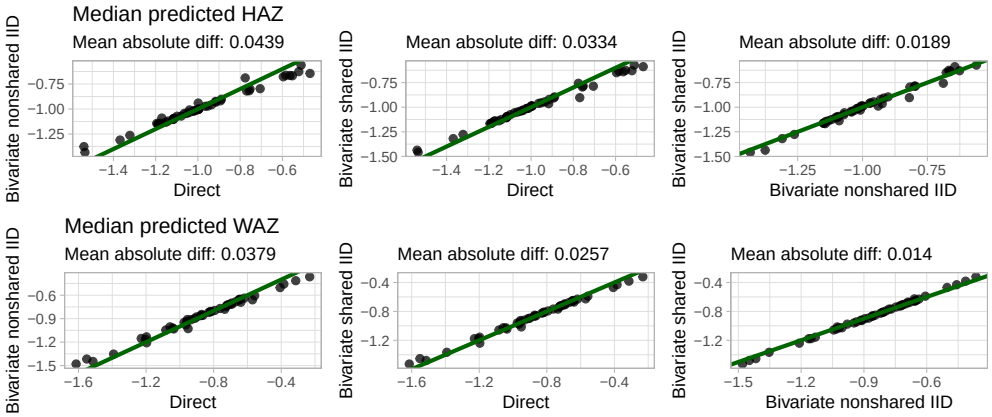


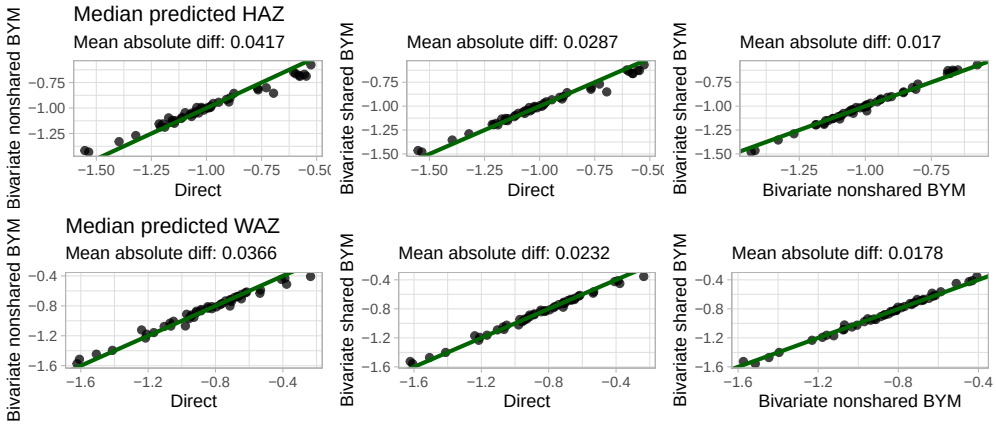
Figure 3.5: Bivariate correlations from the first stage model for estimating HAZ and WAZ from the 2014 KDHS.



(a) Univariate comparisons



(b) Bivariate IID comparisons



(c) Bivariate BYM comparisons

Figure 3.6: Comparison of posterior median latent mean estimates of HAZ and WAZ from the 2014 KDHS.

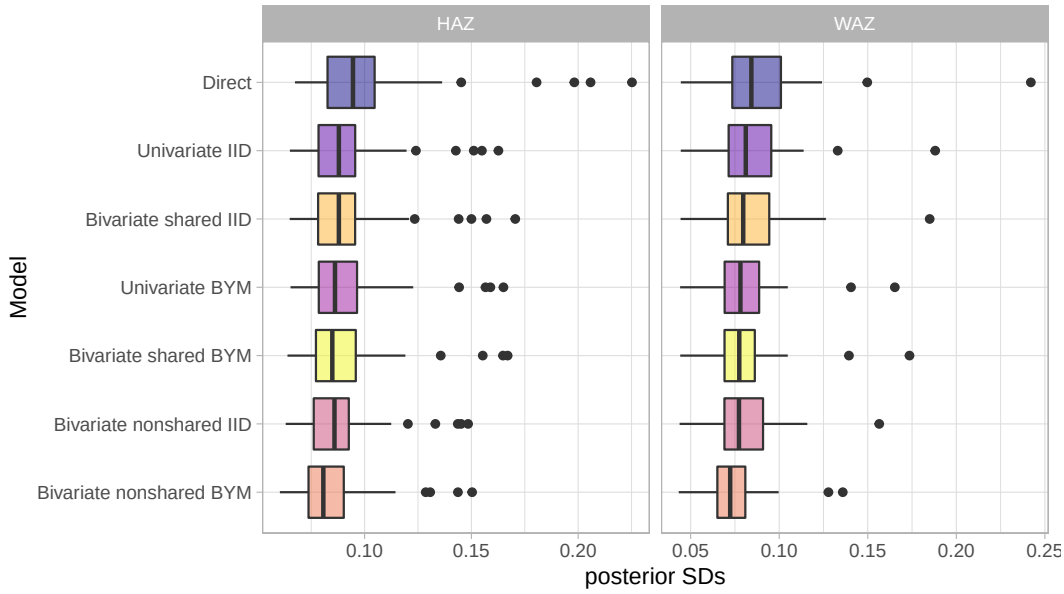


Figure 3.7: Comparison of posterior standard deviations of latent means for each of six models used to estimate HAZ and WAZ from the 2014 KDHS.

Table 3.2: Bivariate LogScore for each model for HAZ/WAZ. Bold indicates the best performing model

	Model	LogScore
(III)	Bivariate nonshared IID	0.379
(I)	Univariate IID	0.274
(V)	Bivariate shared IID	0.139
(II)	Univariate BYM	-0.16
(IV)	Bivariate nonshared BYM	-0.232
(VI)	**Bivariate shared BYM**	**-0.47**

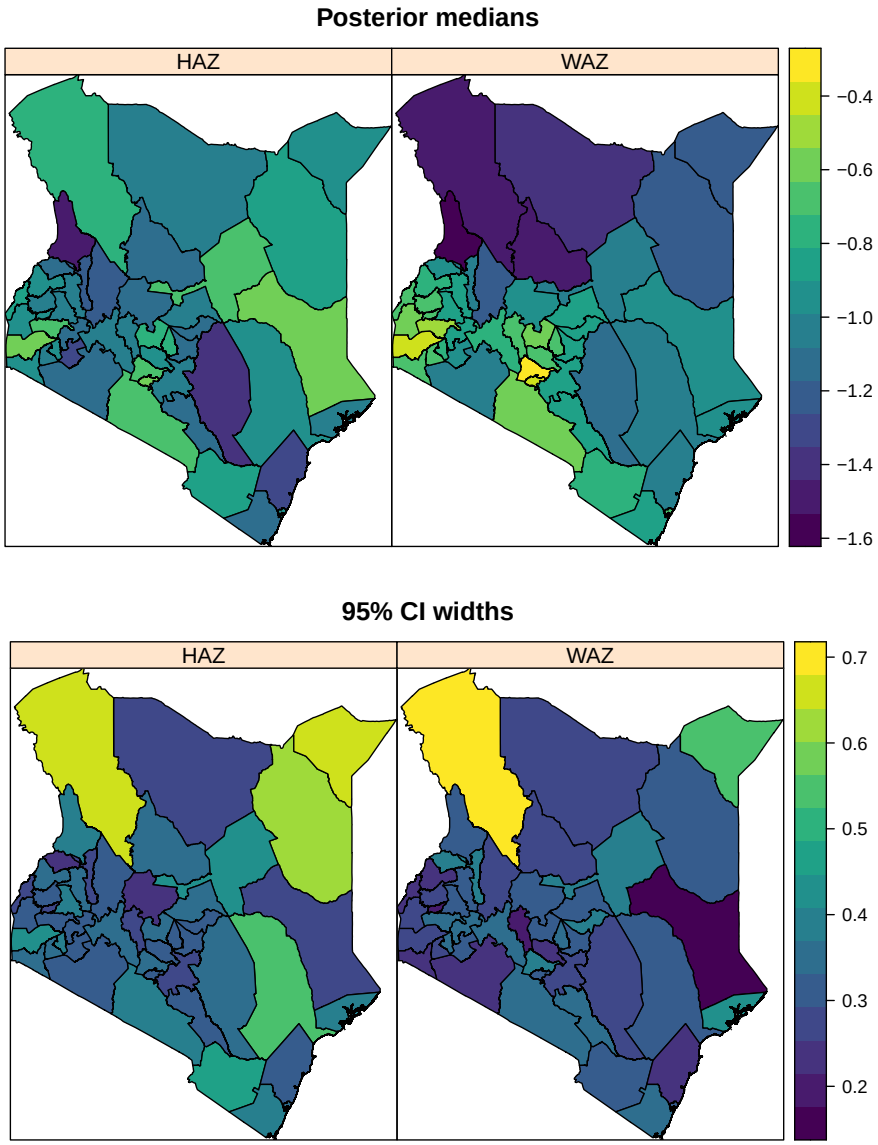


Figure 3.8: Estimated posterior medians and widths of 95% credible intervals for HAZ and WAZ estimated from our bivariate shared BYM model from the 2014 KDHS.

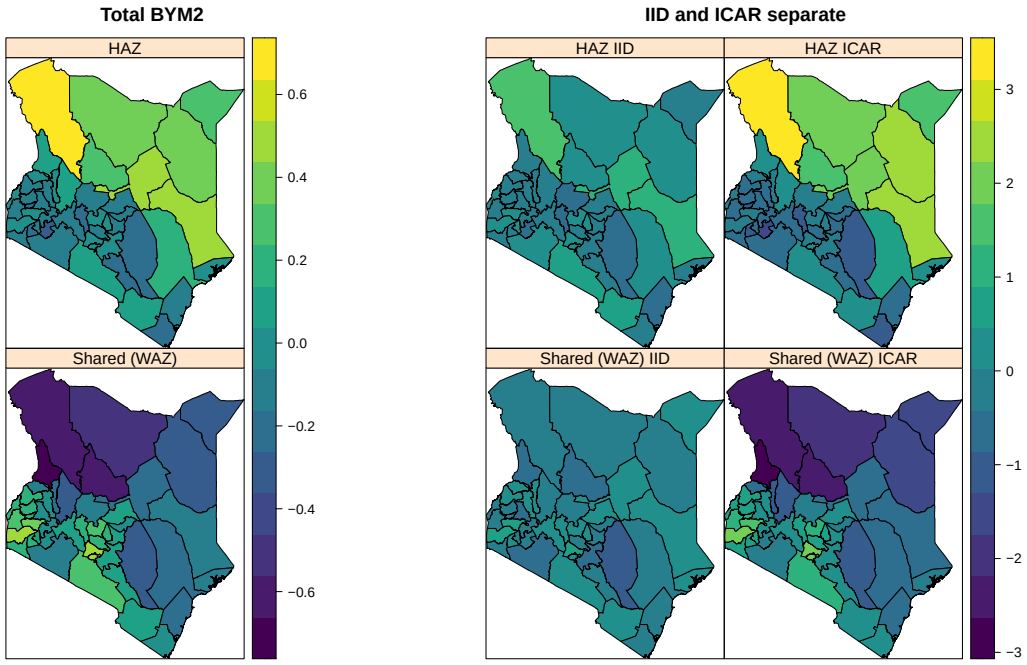


Figure 3.9: Estimated total BYM random effects, along with the IID and ICAR random effect components (normalized to have variance 1), for HAZ and WAZ estimated from our bivariate shared BYM model.

Table 3.3: Posterior medians (Est) and credible intervals (CI) for fixed effects and hyperparameters from the bivariate shared BYM model estimating HAZ and WAZ from the 2014 KDHS data.

Parameter	Est	80% CI	95% CI
β_{haz}	-0.98	(-1.01, -0.96)	(-1.03, -0.94)
β_{waz}	-0.87	(-0.89, -0.85)	(-0.91, -0.83)
σ_1	0.19	(0.16, 0.22)	(0.15, 0.24)
ρ_1	0.79	(0.52, 0.92)	(0.35, 0.96)
σ_2	0.25	(0.21, 0.29)	(0.19, 0.31)
ρ_2	0.92	(0.73, 0.98)	(0.57, 0.99)
λ	0.75	(0.57, 0.92)	(0.47, 1.01)

to the multinomial case. Then, we will present the modeling specifications for seven candidate models, compare them using a bivariate scoring method, and present detailed results for the best performing model.

3.6.1 Adaptation for multinomial data

This area-level multivariate modeling framework can also be adapted for use with multinomial outcomes, with some minor changes. We will use the same notation, except now we will let $c = 1, \ldots, C$ denote levels of a categorical variable. For cluster region r, define N_r as the total number of individuals, Y_{rc} as the number of individuals that are in category c (in vector form $\boldsymbol{Y}_r = [Y_{r1}, \ldots, Y_{rC}]'$), with p_{rc} the associated probability of being in category c (in vector form $\boldsymbol{p}_r = [p_{r1}, \ldots, p_{rC}]'$). Hence, $N_r = \sum_{c=1}^{C} Y_{rc}$.

For the first stage, we will use the individual-level data to calculate survey weighted probabilities in each region along with the estimated design-based covariance matrix, accounting for the complex survey design. We will use a baseline category multinomial logistic general-

ized linear model with only intercepts fit in each region separately. Without loss of generality, we let $c = C$ be the reference group. Then, we define our model for $c = 1, \ldots, C-1$ as

$$\log(p_{rc}/p_{rC}) = \beta_{rc}.$$

Note that there are only $C-1$ estimated parameters due to the constraint that probabilities sum to 1. This is the standard generalized linear modeling approach for multinomial data (Hartzel, Agresti, & Caffo, 2001). However, we must take into account the sampling design. We can calculate sampling weighted point estimates through standard tools for frequency weights in vectorized generalized linear models, since sampling weights are equivalent to frequency weights in these models with independent observations (Lumley, 2020). For each region, this results in a vector of the estimated survey-weighted log relative probabilities (which are called the multinomial log odds or logit probabilities), $\hat{\beta}_{rc}$.

Next, we calculate the design-consistent covariance matrices, $\hat{\boldsymbol{V}}_r^{des}$, via linearization. Details are provided in Lumley (2021). Briefly, linearization requires computing the influence functions of the parameters $h_i(\boldsymbol{\beta}) = -\boldsymbol{\mathcal{I}}_w^{-1} U_i(\boldsymbol{\beta})$, where $-\boldsymbol{\mathcal{I}}_w^{-1}$ is the weighted estimate of the population Fisher information and $U_i(\boldsymbol{\beta}) = \partial_{\boldsymbol{\beta}} \ell_i(\boldsymbol{\beta})$ is the loglikelihood contribution of observation i with weight w_i. The sum of the weighted influence functions (i.e. the population total of the influence functions) are asymptotically equal to the variance of $\hat{\boldsymbol{\beta}}$. Care must be taken for the constraints of the multinomial model. Detail on this, as well as further information on this linearization procedure, can be found in the vignette for the `svyVGAM` package in R (Lumley, 2021). We use the `svy_vglm()` function from this package to calculate $\hat{\boldsymbol{\beta}}_r$ and $\hat{\boldsymbol{V}}_r^{des}$. In the multinomial case, $\hat{\boldsymbol{\beta}}_r$ is a vector of length $C-1$ and $\hat{\boldsymbol{V}}_r^{des}$ is a $(C-1) \times (C-1)$ covariance matrix.

An alternative perspective here is that we are using a pseudo-likelihood approach in which the likelihood is weighted by the appropriate sampling weights. In this sense, the estimated statistics and their corresponding variances are computed from the variance of weighted estimating functions, as described by Binder (1983). This gives asymptotically

correct results for both point estimates and standard errors.

One further issue is for cases on the boundary—this arises if a region has no observations for one of the categories. In this case, the multinomial model cannot estimate a logit probability. Since in most cases (including our example of contraceptive use) we do not believe this to be a true zero but instead we were simply unlucky and did not see any observations in this category. As a practical method of overcoming this, we add a single "phantom" individual who has an observation in this category in each of these regions, but we set the sampling weight to be half of the smallest weight observed in the region in order to minimize the effect on the results yet allow us to estimate $\hat{\beta}_{rc}$ for ever category and every region. Further discussion of this is given in Section 3.7.

Once we have these first stage logit probability estimates, the second stage smoothing model follows exactly like the continuous outcome case, with working likelihood

$$\hat{\boldsymbol{\beta}}_r | \boldsymbol{\mu}_r, \hat{\boldsymbol{V}}_r^{des} \sim N_{C-1}(\boldsymbol{\mu}_r, \hat{\boldsymbol{V}}_r^{des}).$$

We will estimate posterior distributions for the area-specific smoothed logit probabilities (since we are using Bayesian methods), which we can then transform back to probability space via

$$p_{rc} = \frac{\exp(\mu_{rc})}{1 + \sum_{c'=1}^{C-1} \exp \mu_{rc'}}$$

for $c = 1, \ldots, C-1$, and for the reference category C,

$$p_{rC} = \left[1 + \sum_{c'=1}^{C-1} \exp \mu_{rc'} \right]^{-1}.$$

For our motivating example of contraceptive use in three categories, we will use the same candidate models as we did for the HAZ/WAZ example which can be found in Table 3.1.

3.6.2 Results

In our analysis of the contraceptive use data, we set the reference category to be no contraceptive use, such that in region r, β_{r1} is the log ratio of the probability of using a modern method to no method, and β_{r2} is the log ratio of the probability of using an "other" method to no method. This is a natural reference category (typically the "null" category is chosen), and it also allows the effects for the "other" group to be modeled directly rather than used as the basis of comparison for the other two groups. This is because the "other" group has phantom observations added, and it would not be ideal to include these inaccuracies in the reference group because they would propagate to all the logit probability estimates.

First, we present the bivariate correlations estimated from the first stage model in Figure 3.10. The correlations range from -0.48 to 0.64. This indicates fairly strong correlations at the individual and cluster levels, both positive and negative, as well as some areas without much correlation. This range of correlations also demonstrates the potential for improved estimates due to multivariate modeling. Note that these correlations are on the logit probability scale, so they are not simply due to the dependence between multinomial probabilities.

We compare the posterior medians of the latent logit probabilities among all models, and these are presented in Figure 3.11. The comparisons are separated into univariate comparisons, bivariate IID comparisons, and bivariate BYM comparisons. Here, we see similar results as in the HAZ/WAZ example—posterior median estimates of all of the IID models generally agree, which is similar for BYM models, and spatial smoothing leads to shrinking the estimates toward the overall mean. The shrinkage is more apparent here compared to what was seen in the HAZ/WAZ example, notably for the other vs. none logit probabilities. This is mostly due to the very small values of the logit probabilities and especially the two values in the regions where we had to add "phantom" observations, since these values had very low survey weights and thus very high design-consisted variance estimates. We present a comparison of the posterior standard deviations of the latent means in Figure 3.7. Here, we see that the shared BYM has the smallest posterior standard deviations. A comparison

of the LogScore (Section 3.4.4) for each of the seven candidate models is provided in Table 3.2. Recall that a lower score means a better performing model. We see that Model VI, the bivariate shared BYM model, performs best.

Due to these results, we present our final estimates of contraceptive use from the 2014 KDHS as the results of model VI. Summaries of the posterior distributions of parameter estimates are presented in Table 3.5. Figure 3.13 shows a map of the posterior logit probabilities along with the associated widths of the 95% credible intervals. We again see a north-south spatial gradient. In the northern areas, the logit probabilities of modern vs. none and modern vs. other are lower than the south, which means that relative to southern areas, the northern areas have less contraceptive use of any kind, but the uncertainty in these regions is also larger.

We plot maps of the total BYM2 random effects, as well as the IID and ICAR random effects separately, in Figure 3.14. These are on the logit probability scale. The shared "modern vs. none" random effect is the main driver of the spatial dependence, as we see that the scale of its ICAR component is much greater than the scale of the "other vs. none" ICAR component which has a very subtle spatial gradient. Thus, most of the spatial variation is shared between the two logit probability estimates.

We plot a map of the posterior median estimated probabilities of each outcome and their 95% credible intervals in Figure 3.15. These reflect the observations made from the maps of the logit probabilities, and these estimates serve as what might be a typical "deliverable" from this data analysis. Finally, we also fit a bivariate shared BYM model with the "modern vs. none" BYM random effects being the shared component, rather than the "other vs. none" random effects in the above models. In other words, we used the parameterization $\mu_{r1} = \beta_1+v_{r1}+u_{r1}$ and $\mu_{r2} = \beta_2+v_{r2}+u_{r2}+\lambda(v_{r1}+u_{r1})$, rather than $\mu_{r1} = \beta_1+v_{r1}+u_{r1}+\lambda(v_{r2}+u_{r1})$ and $\mu_{r2} = \beta_2 + v_{r2} + u_{r2}$. We compare the posterior median estimates of the proportion in each contraceptive use category as well as widths of 95% credible intervals in Figure 3.16. The posterior medians are nearly identical, while there are more differences in the widths of posterior intervals—these are due to slight numerical instability in the models along

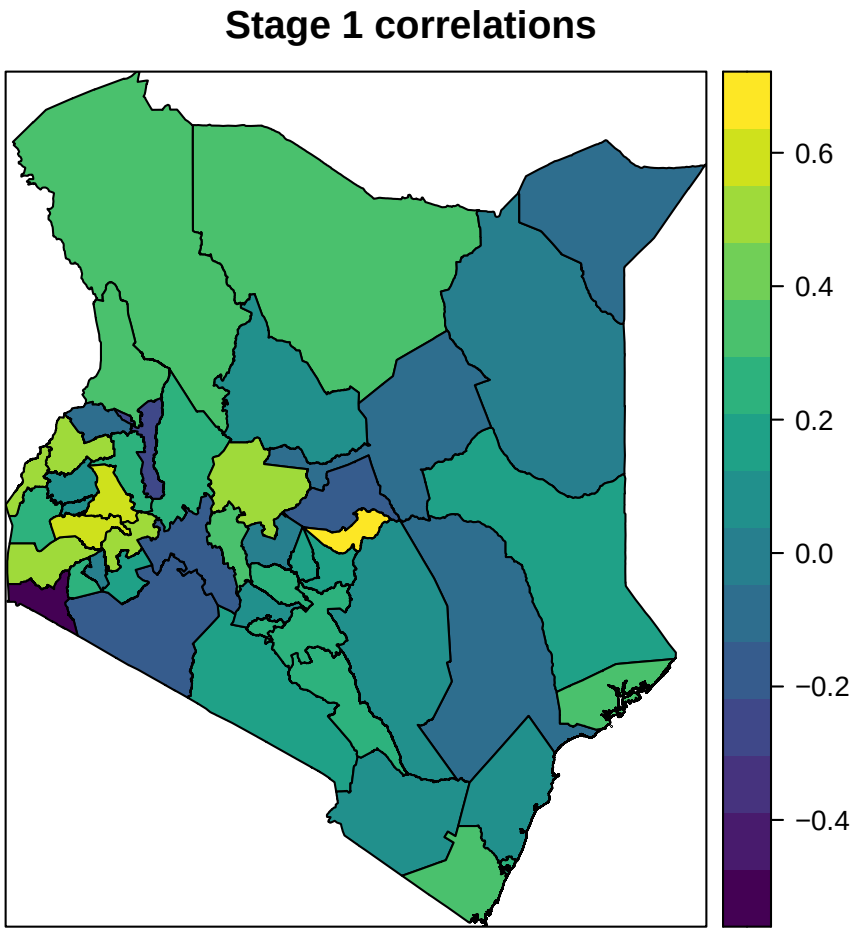
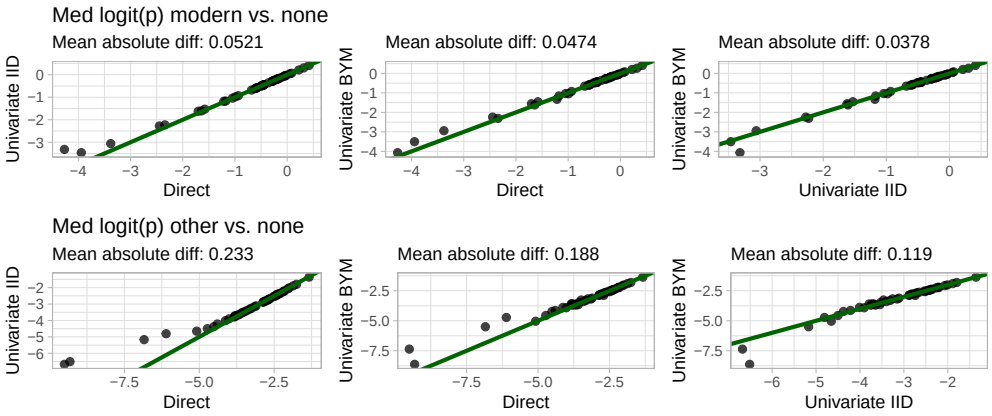
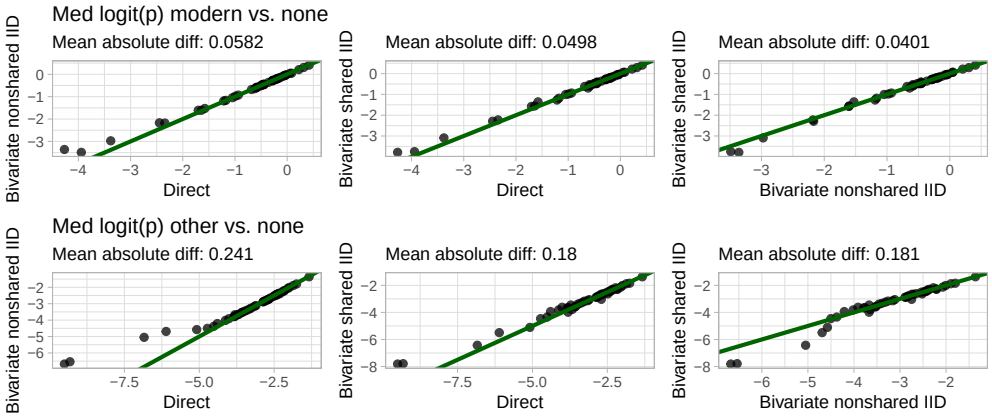


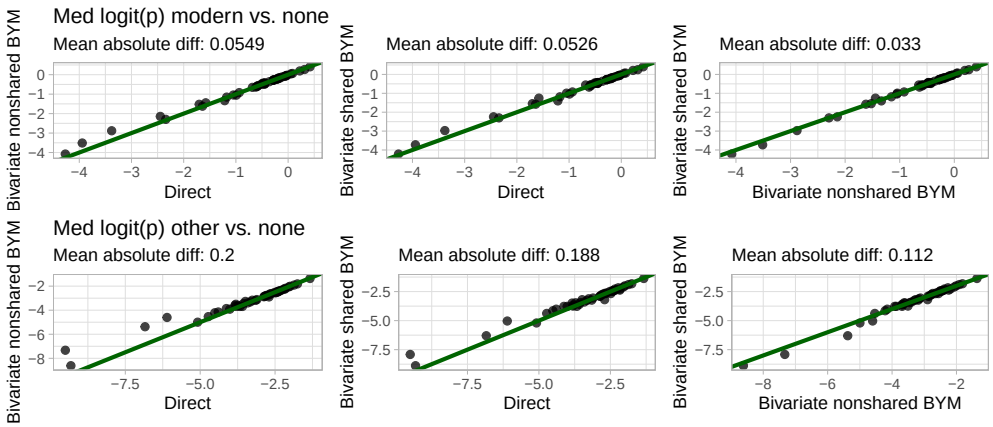
Figure 3.10: Bivariate correlations between the logit probabilities of modern and other contraceptive use from the first stage model for estimating contraceptive use from the 2014 KDHS.



(a) Univariate comparisons



(b) Bivariate IID comparisons



(c) Bivariate BYM comparisons

Figure 3.11: Comparison of posterior median latent mean estimates of contraceptive use from the 2014 KDHS.

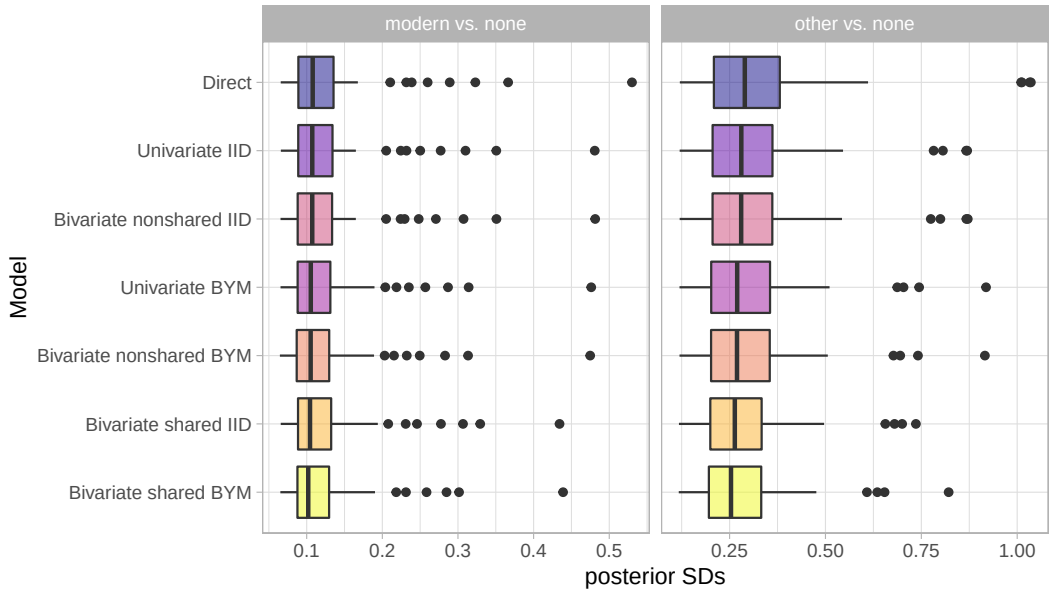


Figure 3.12: Comparison of posterior standard deviations of latent means for each of six models used to estimate contraceptive use from the 2014 KDHS.

Table 3.4: Bivariate LogScore for each model for contraceptive use. Bold indicates the best performing model

	Model	LogScore
(III)	Bivariate nonshared IID	3.51
(I)	Univariate IID	3.33
(V)	Bivariate shared IID	2.87
(II)	Univariate BYM	2.05
(IV)	Bivariate nonshared BYM	2.03
(VI)	**Bivariate shared BYM**	**1.79**

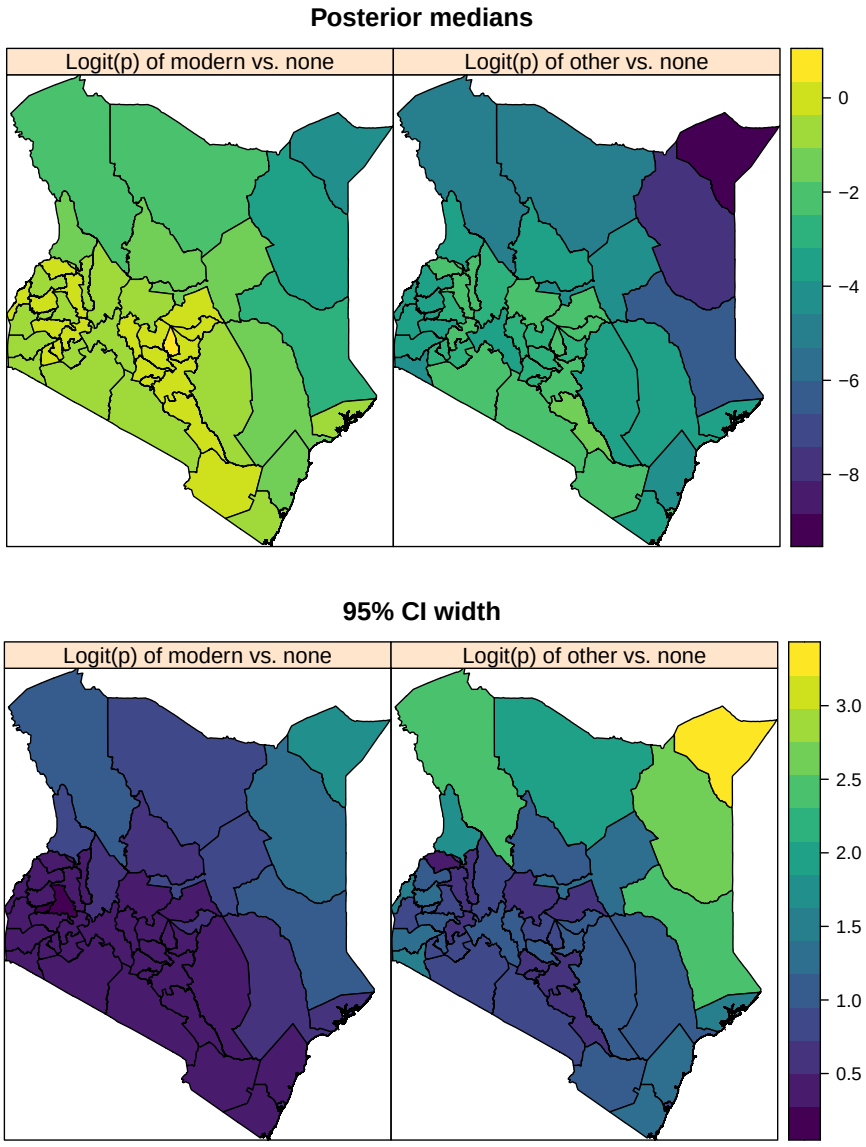


Figure 3.13: Estimated posterior medians and widths of 95% credible intervals for logit probabilities of contraceptive use estimated from our bivariate shared BYM model from the 2014 KDHS.

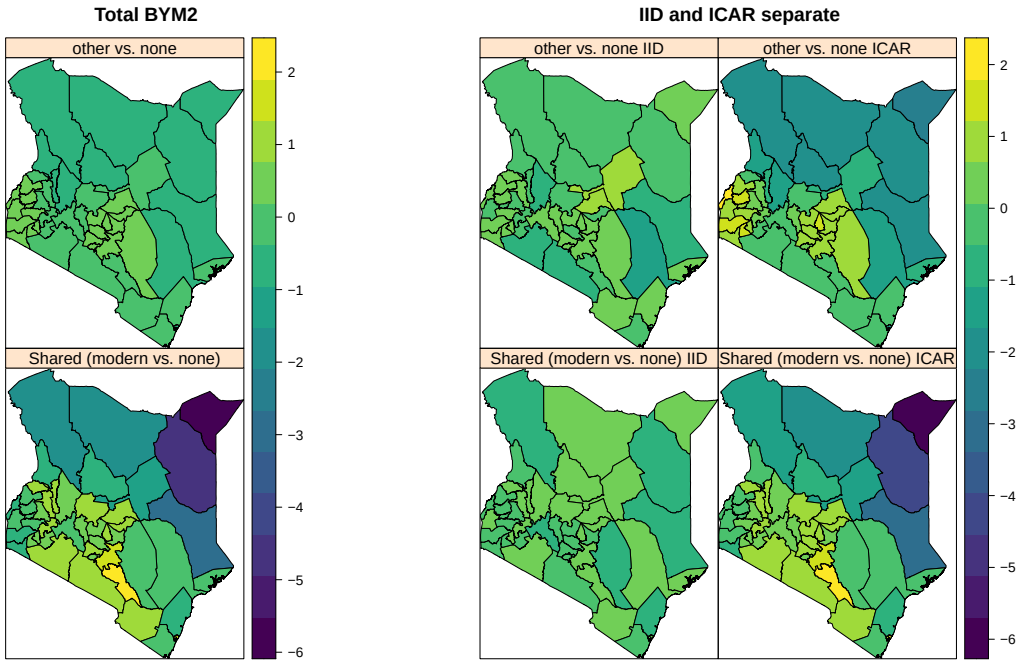


Figure 3.14: Estimated total BYM2 random effect, as well as IID and ICAR random effects (normalized to have variance 1), for contraceptive use estimated from our bivariate shared BYM model. Estimates on the logit(p) scale.

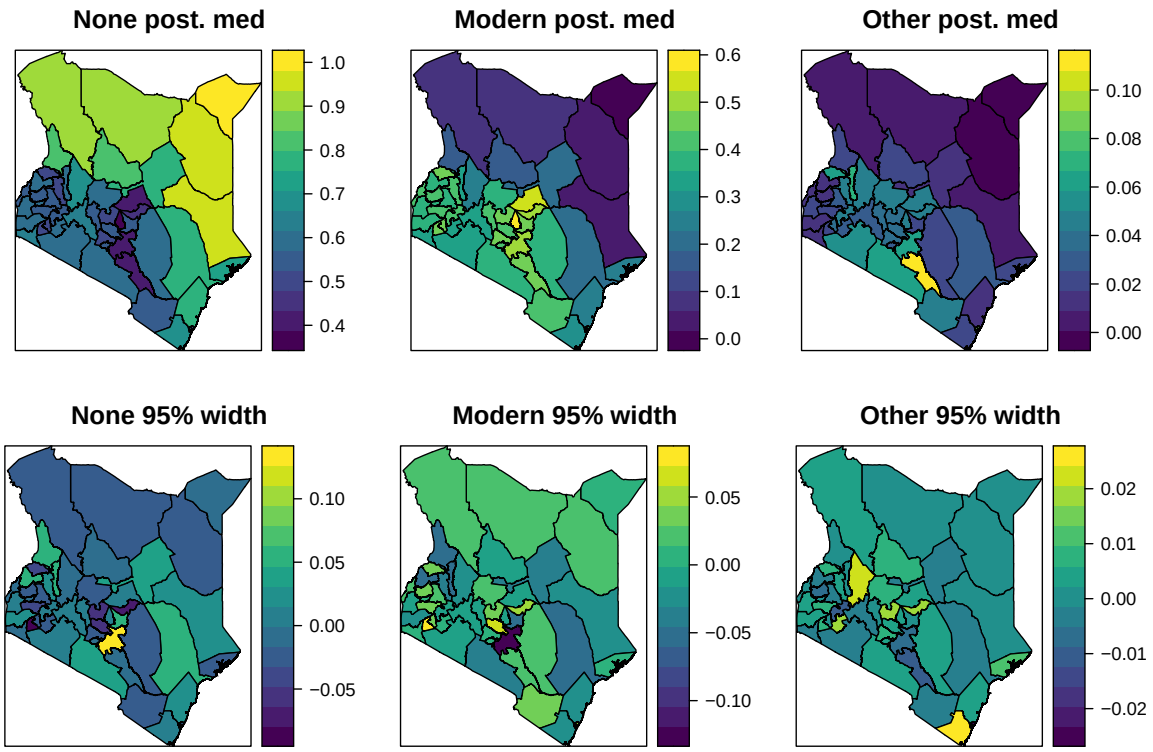


Figure 3.15: Estimated posterior medians and widths of 95% credible intervals for probabilities of contraceptive use estimated from our bivariate shared BYM model from the 2014 KDHS.

Table 3.5: Posterior medians (Est) and credible intervals (CI) for fixed effects and hyperparameters from the bivariate shared BYM model estimating contraceptive use from the 2014 KDHS data.

Parameter	Est	80% CI	95% CI
β_1	-0.72	(-0.78, -0.67)	(-0.81, -0.64)
β_2	-3.32	(-3.42, -3.22)	(-3.48, -3.16)
σ_1	0.37	(0.31, 0.44)	(0.28, 0.49)
ρ_1	0.86	(0.62, 0.97)	(0.48, 0.99)
σ_2	1.03	(0.88, 1.21)	(0.82, 1.32)
ρ_2	0.92	(0.77, 0.98)	(0.66, 0.99)
λ	0.48	(0.39, 0.57)	(0.34, 0.62)

with slightly inconsistently specified prior distributions. The differences are small, which is comforting, but it brings to attention that care should be taken when parameterizing shared component models, and this type of model checking can be important.

3.7 Discussion

We have developed a class of multivariate shared component Fay-Herriot area-level SAE models using complex survey data. We presented simulation results showing how various models in this class compare across multiple scenarios, including comparison to direct estimates. These showed that in many scenarios, the choice of second stage smoothing model is relatively unimportant—especially in cases with large amounts of data—while in others a correctly specified model, or at the very least a model that correctly specifies the spatial component and either correctly specifies a shared component or fits univariate models separately, performs best. In motivating examples, we jointly modeled height for age and weight for age in children under age 5 as well as modeled contraceptive use in women age 15-49

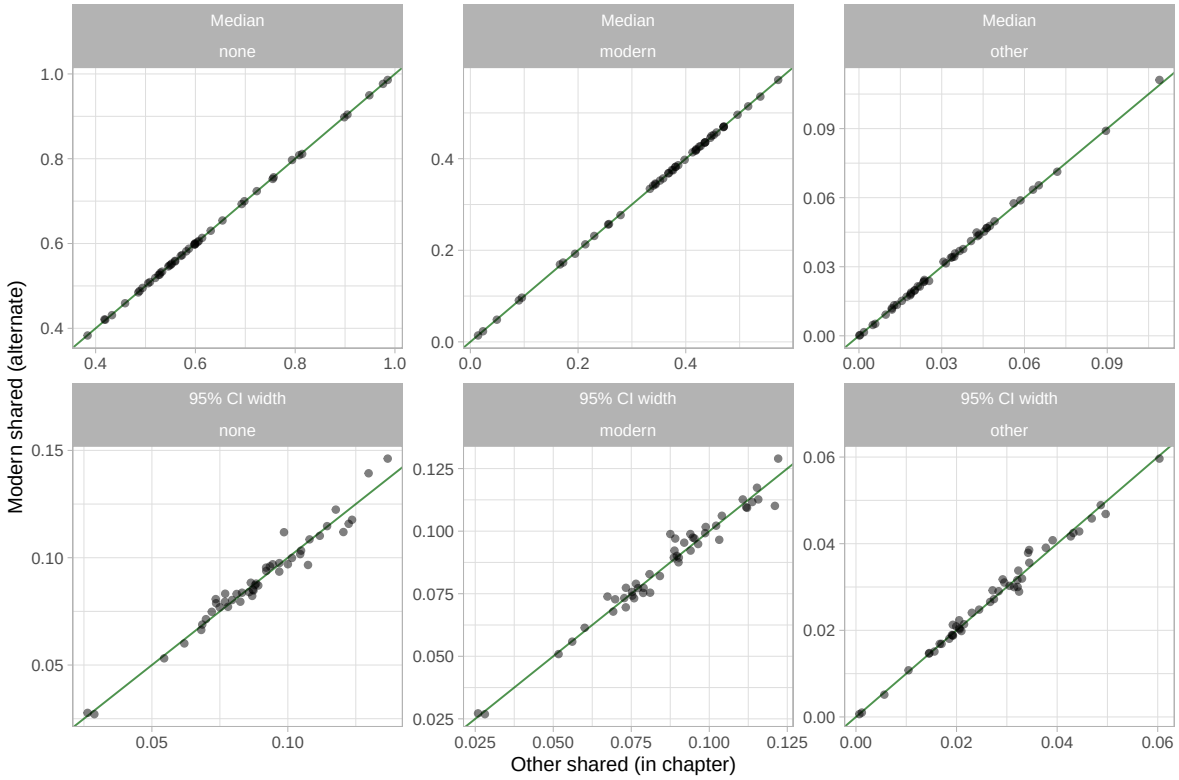


Figure 3.16: Comparison of estimated posterior medians and widths of 95% credible intervals for probabilities of contraceptive use estimated from bivariate shared BYM models from the 2014 KDHS. The other shared (in chapter) model is the one used for final estimates, while the modern shared (alternative) model is the comparison that switches which logit probability has the shared component.

using data from the 2014 Kenya DHS, In these modeling exercises, we compared a suite of models within the class we have developed and presented results from the best performing model.

This class of models provides a novel method for modeling multivariate outcomes from complex survey data. While the utility of the shared component part of our class of models was limited in our simulation study, our analysis of HAZ and WAZ from the 2014 KDHS showed that the bivariated shared component BYM model was the best performing which lends credence to the importance of this type of model. Further research into its performance on other outcomes is a priority.

One aspect of these data from the motivating examples that was not considered was measurement error. For HAZ and WAZ, measurement is notoriously difficult (Ulijaszek & Kerr, 1999). If the measurement error is not systematically biased, then the uncertainty of our estimates will not reflect the true uncertainty in the data, thus being anticonservative. Any bias, however, will propagate through and affect the mean estimates. For example, if a specific DHS interviewer that only worked in one geographic region tended to underweight children, then WAZ in those regions would be lower artificially, and our spatial estimates will be less accurate. Future work is critical to assess the impact of measurement error on estimates from these models, and any data analysis should take this into consideration when interpreting results. Furthermore, adapting these models to account for measurement error may be a fruitful pursuit.

Our modeling exercise from the motivating examples use relatively uninformative PC priors. Due to the amount of data available in the 2014 KDHS, this is an adequate choice. Analyses with more informative prior distributions did not affect estimates greatly (results not shown). However, in other situations with less data and/or more variable outcomes, the choice of prior distribution may be important, and further research is needed into how different priors on each of the parameters affects resulting estimates. In particular, priors on the shared component are an important avenue for exploration.

This chapter also restricted the class of models it used to BYM spatial models which use an ICAR model for spatial dependence. Other spatial models could also be used, such the aforementioned CAR models or models with Leroux spatial random effects (Leroux, Lei, & Breslow, 2000), and these should be explored in future research. In practice, ICAR models are often used (Mercer et al., 2015; Wakefield et al., 2020), which is why they were the focus of this chapter. Additionally, our work can be extended to include spatiotemporal models. Spatiotemporal two stage Fay-Herriot modeling approaches are used in many applications (Esteban, Morales, & Pérez, 2016; Z. R. Li et al., 2019; Marhuenda et al., 2013; Rumiati, Kuswanto, et al., 2019). A spatiotemporal extension of our model that adapts space-time interaction models currently used in the literature can be used in these contexts, such as jointly modeling under-5 mortality from different causes or jointly modeling HIV prevalence by contraceptive use. Recent developments on spatiotemporal shared component models (Ahmadipanahmehrabadi, Hassanzadeh, & Mahaki, 2019; Blangiardo et al., 2020; Mahaki, Mehrabi, Kavousi, & Schmid, 2018; Paradinas, Conesa, López-Quílez, & Bellido, 2017) can be explored to examine whether these approaches translate to the context of estimating demographic and health outcomes from complex survey data in LMICs.

This leads to a discussion of how to model more complex outcomes. One particularly interesting case is modeling cause-specific mortality. Here, the first stage model becomes much more complicated. Rather than simply aggregating means in an appropriate way with regard to the survey sampling design, we must account for the time-to-event nature of the data, which is commonly done with discrete time survival analysis techniques (Mercer et al., 2015). This can be done in our proposed modeling framework, but requires a large amount of data. In preliminary work on modeling cause-specific under-5 mortality using data from the verbal autopsy module in the 2017 Bangladesh DHS, there was not enough data to produce stable first stage estimates before we could even contemplate using second stage models. This warrants further research into incorporating more data sources and exploring new modeling techniques.

Another interesting route to explore is comparing shared component models with models

that include covariates. Since the primary mechanism through which we motivated the shared component model is via shared but unobserved risk factors and other covariates, it would be interesting to see how models with covariates compare to shared component models without covariates. This exercise may provide further support for our class of shared component models as a whole, and it is straightforward to include area-level covariates.

Lastly, this chapter solely considered area-level models. Another possible class of models that would be applicable to this type of data would be unit-level models. One reason to explore unit-level models is the boundary issue that arose in this chapter when modeling the contraceptive use data—some regions has zero sampled individuals who used "other" contraceptive methods. To remedy this problem for our area-level models, we added a "phantom" observation with a very small sampling weight in each region with this issue, which is a somewhat ad hoc solution. Unit-level models will not require these phantom observations. In the next chapter, we develop a multivariate shared component unit-level modeling approach and explore its utility to produce area-level estimates for the same two motivating examples used in this chapter. A comparison of area- and unit-level approaches is presented in Chapter 5.

Chapter 4

UNIT-LEVEL MODELS FOR SMALL AREA ESTIMATION OF MULTIVARIATE OUTCOMES FROM COMPLEX SURVEY DATA

4.1 Introduction

This chapter continues discussion of small area estimation (SAE) using data from surveys with complex design for estimating health and demographic outcomes in low- and middle-income countries (LMICs). See the previous chapter for the background and motivation of this setting. The last chapter covered area-level model-based approaches, where covariate information and sampling models are specified at the desired level of aggregation for final estimates. Unfortunately, area-level models have two key difficulties that arise somewhat frequently. First, direct estimates may be on the boundary for a summary parameter that is not on the whole real line, in which case the estimate will be undefined once transformed to the whole real line. Unmatched linking models do not do this. In the previous chapter, this occurred in the example of modeling contraceptive use from the 2014 KDHS—two regions had no sampled individuals in the "other" category. A second difficulty is that reliable area-level variance estimates may be unavailable, particularly for areas with few/no samples. One way to address these shortcomings is the use of unit-level (also called cluster-level) models, for which data are modeled at the level of the individual or cluster. Unit-level models are the focus of this chapter.

Beyond overcoming some of the limitations of area-level models, unit-level approaches can provide more precise estimates along with easy spatial aggregation as they constitute a bottom-up approach (i.e., they utilize the finest scale of resolution of the data). Since model inputs are at the unit-level (person-level, household-level, or establishment-level), predictions

and estimates can be made at the same level as the inputs, or aggregated up to any desired level. This aggregation, however, is nontrivial, and its feasibility depends on the form of the model as well as the availability of population-size data or estimates. Furthermore, perhaps the biggest challenge for unit-level models is how to account for the survey design in the model. With area-level models, the survey design is incorporated into the model through the specification of a Gaussian sampling distribution for the weighted estimator, which includes the design-consistent variance estimates. With unit-level models, accounting for the survey design is not as simple. One challenge that arises is with informative sampling schemes, in which the response variables are correlated with the sample selection variables. In these scenarios, in order to avoid bias, it is critical to capture the sample design in the model. This can be done by including the design variables used to construct the survey weights in the statistical model. In addition, one needs to consider the clustering if cluster-based sampling is used.

Much methodological development has centered on modeling univariate outcomes. Battese, Harter, & Fuller (1988) proposed a nested error regression model at the unit level, which was one of the first instances of unit-level model-based SAE. There have been various extensions for binary, count, and spatial data (Ghosh, Natarajan, Stroud, & Carlin, 1998; MacGibbon & Tomberlin, 1989). The general reference for unit-level models is J. N. Rao & Molina (2015), Section 4.5, to which readers can refer for a review of the basics. A comprehensive recent review of unit-level models for complex survey data can be found in Parker, Janicki, & Holan (2019). Since unit-level models use traditional statistical modeling techniques, much development of models is application-specific. For applications using complex survey data from low- and middle- income countries (LMICs) for estimating health indicators, Wakefield et al. (2020) provides a useful overview. Some examples include Hobza, Morales, & Santamaría (2018), Diggle & Giorgi (2019), and Godwin & Wakefield (2021).

As discussed in the previous chapter, modeling multiple outcomes together can lead to improved estimates compared to univariate models. Much work on multivariate unit-level models have focused on modeling multinomial data. Ghosh et al. (1998) devised models for

multi-category data, while a number of studies have used multinomial logistic mixed models to estimate small area counts for a categorical variable (Molina, Saei, & Jose Lombardia, 2007; Saei & Taylor, 2012; X. Wang, Berg, Zhu, Sun, & Demuth, 2018). Beyond applications-specific ventures, general modeling approaches have also been developed (Berg & Fuller, 2014; Zhang & Chambers, 2004). For the task of jointly estimating two or more outcomes (e.g., multiple continuous variables) in the context of SAE, various multivariate models have been developed (Ito & Kubokawa, 2021; Lohr & Prasad, 2003; Ngaruye, Nzabanita, Rosen, & Singull, 2017). Models encompassing all of the above types of multivariate outcomes typically use latent variables in conditionally specified models (F. Li & Zaslavsky, 2010; Sun, Berg, & Zhu, 2021). More recently, Esteban, Lombardía, López-Vizcaíno, Morales, & Pérez (2022) describes a bivariate linear mixed model for ratio and mean estimation.

The citations in the previous paragraph do not include any examples of spatial models. Outside the realm of SAE, multivariate spatial modeling at the unit-level has been extensively developed, especially regarding pollutant data. An overview of these methods was provided in the previous chapter; notably, "coregionalization" approaches, in which latent parameters are shared among outcomes, are popular and typically use continuous spatial models (Palmí-Perales, Gómez-Rubio, & Martinez-Beneito, 2019; Schmidt & Gelfand, 2003). MacNab (2016) contains a thorough review. Coregionalization is essentially the continuous spatial analog of discrete shared component models (Held et al., 2005; Knorr-Held & Best, 2001) which were described in the last chapter.

Continuous spatial models naturally arise for unit-level analyses because they allow specification of spatial effects at the same resolution as the observed data. However, if area-level estimates are desired, aggregation is challenging because auxiliary population information is needed which can be problematic when the models use non-linear link functions. Further, the fine-grained population densities required are typically modeled quantities with uncertainty, and it is not clear how to incorporate this uncertainty into the aggregate estimates. This may be the reason that spatial models for SAE have been chiefly developed as part of area-based modeling. Thus, discrete spatial models are attractive—this motivates the use of discrete-

space shared component models in unit-level analyses. Discrete time models for unit-level data have been used in the univariate case (Godwin & Wakefield, 2021), but a gap in the SAE literature persists for spatial modeling at the unit-level using complex survey data for multivariate outcomes. The frontier of this research has considered multinomial outcomes (Parker, Holan, & Janicki, 2022), and multiple binary outcomes (Lawson, Schritz, Villarroel, & Aguayo, 2020), but the general multivariate case has not been thoroughly explored. Furthermore, the useful shared component model has not been applied.

To help fill this gap, this chapter proposes a class of multivariate shared component unit-level SAE models with discrete spatial effects to analyze complex survey data. The models developed are similar in nature to those in the previous chapter, but they are built within a unit-level framework. Again, we use a Bayesian modeling approach and develop latent Gaussian models that are fit using integrated nested Lapace approximation (INLA). We present two motivating examples that use data from the 2014 Kenya Demographic and Health Survey (2014KDHS): jointly modeling height-for-age and weight-for-age in children under age 5, and modeling three categories of contraceptive use in women aged 15-49. The organization of this chapter is as follows. Section 4.2 provides a brief overview of unit-level SAE methodology. Section 4.3 details our proposed multivariate shared component model and presents a simulation study for validation. Sections 4.4 and 4.5 describe using this methodology to produce estimates for our motivating examples, comparing results among univariate models and bivariate non-shared models. Section 4.6 provides a discussion and presents avenues for future research.

4.2 Background methodology

This chapter is concerned with *model-based* approaches to SAE, which are a class of indirect estimators that assume a model for a theoretical infinite superpopulation from which the sample data are drawn. In particular, we focus on *unit-level* rather than *area-level* approaches. Area-level approaches specify covariate information and modeling parameters at the desired level of aggregation for final estimates, while *unit-level* approaches model data

at the level of the individual or cluster. The data used will be from surveys with a complex design, which necessitates methodology from the survey sampling literature. See last chapter for a review of area-level SAE as well as survey sampling.

We will first reintroduce notation and define additional terms specific for unit-level models. We assume the units of analysis are clusters within a multistage cluster design. Let $i = 1, \ldots, N_k$ denote individuals which are from $k = 1, \ldots, K_r$ clusters sampled from $r = 1, \ldots, R$ regions, and let $c = 1, \ldots, C$ denote outcomes (i.e. multiple continuous variables like height- and weight-for-age). Let $\boldsymbol{s}_{rk}$ represent the geographical location of cluster k in region r. Let Y_{rkic} be the random variable for outcome c for individual i in cluster k of area r, with observed value y_{rkic}. Throughout this chapter, we will use bold font and drop relevant subscripts to refer to vectors. Thus, the vector of observed outcomes for individual i in cluster k of region r is denoted $\boldsymbol{y}_{rki} = [y_{rki1}, \ldots, y_{rkiC}]$.

First, we will review unit-level SAE models from a simplified univariate perspective with noninformative sampling. For simplicity, we assume the observed data is at the cluster-level (no individuals), and that regions are the desired small areas for analysis. Thus, we drop the outcome subscripts c and the individual subscripts i.

The nested error unit-level model, first proposed by Battese et al. (1988), has the form

$$y_{rk} = x_{rk}^T \boldsymbol{\beta} + u_r + \epsilon_{rk}$$

with individual-level covariates x_{rk}, random effects u_r, and residual terms ϵ_{rk} which are independent and identically distributed (IID). Parameter estimation and predictions can be carried out via calculation of the best linear unbiased predictors or via specification of a hierarchical Bayesian model. This form of model was expanded to consider a generalized linear mixed model (GLMM) approach (Ghosh et al., 1998; MacGibbon & Tomberlin, 1989) that has been broadly used for binary and count data.

However, these methods do not explicitly account for informative sampling from the complex design of most surveys. The DHS, for example, carries out a stratified cluster

sampling scheme for selecting individuals. A crucial assumption for unit-level models is that the selection probability, given covariates, does not depend on the values of the response (J. N. Rao & Molina, 2015). This implies that for modeling data from a stratified random sample, stratification variables must be included in the model. Fixed effects for each strata are recommended (Paige et al., 2022). Furthermore, cluster sampling leads to correlated responses within clusters. In this case, cluster-level random effects are typically used (Scott & Smith, 1969). Models that abide by the above recommendations have been used for unit-level SAE in the context of estimating health and demographic outcomes (Z. R. Li et al., 2019; Wakefield et al., 2019).

For the particular stratified cluster sampling design of the DHS, we have strata that consist of a binary urban/rural classification crossed with geographical administrative regions. A number of possible models can be specified for these data that account for this design. If we use fixed effects on each region and on urban/rural classification, then we would need interactions between region and urban/rural. However, many models include spatial random effects to leverage spatial dependence, and interactions would produce an unidentifiable model. Since spatial modeling with random effects is used in this chapter, we consequently use a single urban/rural fixed effect and no fixed effects on region. This imposes a constant association between the response and the urban/rural variable across areas. Another possibility, would be to include a spatially varying urban/rural coefficient, or model the associations separately in larger regions, which may be attractive if the size of the observed data allows for it or if borrowing strength across regions is a lower priority.

We now present a popular candidate model on which the remainder of the chapter is based—a similar model is presented in Wakefield et al. (2020). We do not use covariates in the models in this chapter, but it would be straightforward to do so. We start with the univariate case. Assuming a continuous Gaussian outcome $Y_{rk}|\mu_{rk}, \sigma^2 \sim N(\mu_{rk}, \sigma^2)$, one popular candidate model parameterizes the means as

$$\mu_{rk} = \beta_0 + z(\boldsymbol{s}_{rk})\gamma + S(\boldsymbol{s}_{rk}) + \epsilon_{rk}$$

where $z(\boldsymbol{s}_{rk})$ is the a binary indicator that the cluster at location $\boldsymbol{s}_{rk}$ lies in a rural strata, with γ is the associated effect of urban/rural status. The spatial random effects $S(\boldsymbol{s}_{rk})$ are associated with cluster location $\boldsymbol{s}_{rk}$, and may be continuous or discrete. The cluster-level error $\epsilon_{rk} \overset{iid}{\sim} N(0, \sigma^2_\epsilon)$ is the so-called nugget, which is traditionally taken to represent short scale variability and/or unmodeled sampling variability (also labeled a little inaccurately as "measurement error"), and is frequently not included when making model-based predictions (Dong & Wakefield, 2021).

The specification of the spatial model for $S(\boldsymbol{s}_{rk})$ generally comes down to two schools of thought. First, the model-based geostatistical (MBG) approach centers on continuous spatial models and takes $S(\boldsymbol{s}_{rk})$ as a realization of a zero-mean Gaussian process (GP), which implies that any collection of spatial random effects have a multivariate normal distribution. A popular choice for the GP variance-covariance is the Matérn covariance function (Stein, 1999), for which the covariance between is,

$$\text{cov}(S(s_1), S(s_2)) = \sigma^2_S \frac{2^{1-\nu_S}}{\Gamma(\nu_S)} \left(\sqrt{8\nu_S} \frac{||s_2 - s_1||}{\rho_S} \right) \kappa_{\nu_S} \left(\sqrt{8\nu_S} \frac{||s_2 - s_1||}{\rho_S} \right)$$

where ρ_S is the spatial range corresponding to the distance at which the correlation is approximately 0.1, σ_S is the marginal standard deviation, ν_S is the smoothness (which is usually fixed, since it is difficult to estimate), and κ_{ν_S} is a modified Bessel function of the second kind, of order ν_S. Large numbers of clusters lead to computational challenges because we need to manipulate $K \times K$ matrices which involves $O(K^3)$ operations (Rue & Held, 2005). To overcome this problem, one popular approach is using stochastic partial differential equations (SPDE) (Lindgren et al., 2011). Other approaches are described by Heaton et al. (2019).

If the desired level for reporting results is not at the unit-level but instead at the area-level, which is commonplace in analyses of demographic and health data due to policy relevance

and interpretability, then estimates must be aggregated. This is a challenging endeavor for models with continuous spatial effects. To illustrate, let μ_r be the area-level mean and let $\mu(s) = \beta_0 + z(\boldsymbol{s}_{rk})\gamma + S(\boldsymbol{s}_{rk})$ be the estimated outcome at location $\boldsymbol{s}_{rk}$. Note that ϵ_{rk} is omitted since it is viewed as measurement error. Aggregation to the area-level is carried out through

$$\mu_r = \int_{A_r} \mu(s) \times q(s) ds \approx \sum_{l=1}^{M_r} \mu(s_l) \times q(s_l)$$

with A_r denoting the area boundaries of region r, $q(s)$ is the population density at s which is needed at all locations on the approximating mesh s_l, $l = 1, \ldots, M_r$.

The challenge arises because covariates, design variables, and population densities are needed for all locations on the approximating mesh. Obtaining these values is highly non-trivial, with these values often being modeled estimates themselves—especially for population densities and covariate surfaces—for which no standard method exists for propagating uncertainty through to the final aggregated estimates. In principle, we could do it if we had (say) samples for the population and covariate surfaces. Area-level methods avoid this problem since the weights implicitly include population information from the sampling frame, but unit-level models remain appealing due to being closer to the mechanism of action for covariates. If associations are of interest, then this has the possibility of reducing ecological bias (Wakefield, 2008). However for SAE, we are generally interested in predictions at the area-level rather than associations, so interpretations is less of a concern. Nevertheless, the signal will be stronger at the cluster level.

Beyond the MBG paradigm, a second approach to choosing a spatial model uses random effects at the area level. Using discrete spatial models, rather than continuous, allows easy aggregation as well as computational advantages. An example is the use of the Besag, York and Mollié (BYM) model which includes both an intrinsic conditional autoregressive component for spatial auto-correlation and an IID random effects component for non-spatial heterogeneity (Besag et al., 1991). Recently, a reformulation of the BYM model, called

the BYM2 model, has been developed which has beneficial properties for modeling and interpretability. For more detail on BYM and BYM2 spatial models, see Section 3.3.5 in the previous chapter. Letting v_r and u_r be the IID nonspatial and ICAR spatial components of a BYM2 model, respectively, we can write our previous candidate model as

$$\mu_{rk} = \beta_0 + z_{rk}\gamma + \sigma(\sqrt{1-\rho}v_r + \sqrt{\rho}u_r) + \epsilon_{rk}$$

where z_{rk} is an indicator of cluster k in region r being rural, σ^2 is the total variance of the BYM2 random effects, and ρ is the proportion of the variation that is spatial. In this model, aggregating up to the region-level is much more straightforward—it only requires knowing the proportion of the relevant population in each area that is rural, q_r. This may come from modeled estimates, census data, or known sampling frames. Aggregation for each region is done via

$$\begin{aligned}\mu_r = (1-q_r) \times \big(\beta_0 + \sigma(\sqrt{1-\rho}v_r + \sqrt{\rho}u_r)\big) \\ + q_r \times \big(\beta_0 + \gamma + \sigma(\sqrt{1-\rho}v_r + \sqrt{\rho}u_r)\big).\end{aligned}$$

Due to their positive attributes, discrete spatial models are the focus of this chapter, rather than continuous models.

Moving on to the multivariate case, the vast majority of spatial SAE model development has been done with area-level models, which was surveyed in the previous chapter. Multinomial outcomes (Parker et al., 2022), as well as multiple binary outcomes (Lawson et al., 2020), have been modeled spatially. Extending our univariate candidate model to the multivariate case, we can begin indexing by outcome $c = 1, \ldots, C$. Naturally, we can include outcome-specific intercepts, urban/rural effects, BYM2 random effects, and nuggets, which yields

$$\mu_{rkic} = \beta_c + z_{rk}\gamma_c + \sigma_c(\sqrt{1-\rho_c}v_{rc} + \sqrt{\rho_c}u_{rc}) + \epsilon_{rkc}.$$

However, as discussed in the previous chapter, many demographic and health outcomes share common but unobserved covariates, which motivates the use of shared component models. Therefore, we can adapt the above to yield a new candidate model. For simplicity, we only consider the bivariate case, for which the pair of means are

$$\mu_{rk1} = \beta_1 + z_{rk}\gamma_1 + \sigma_1(\sqrt{1-\rho_1}v_{r1} + \sqrt{\rho_1}u_{r1}) + \lambda(\sigma_2(\sqrt{1-\rho_2}v_{r2} + \sqrt{\rho_2}u_{r2})) + \epsilon_{rk1}$$
$$\mu_{rk2} = \beta_2 + z_{rk}\gamma_2 + \sigma_2(\sqrt{1-\rho_2}v_{r2} + \sqrt{\rho_2}u_{r2}) + \epsilon_{rk2}$$

This model serves as the cornerstone to this chapter and will be discussed further in the next section.

4.3 A multivariate unit-level shared component model

To summarize our previous discussion, unit-level modeling approaches can produce usable estimates at the desired level of geographic aggregation in settings without sufficient data to produce reliable model-based variance estimates at the area-level. While continuous spatial models are commonly used, discrete spatial models are more computationally tractable and allow for far less methodological challenging aggregation to produce area-level estimates. This section proposes a unit-level modeling approach for multivariate outcomes that uses discrete spatial effects and shared latent parameters among outcomes. We first describe the model for continuous outcomes, and then adapt it for a multinomial outcome.

4.3.1 Modeling description

In this chapter, we take a Bayesian approach. We will use a Gaussian likelihood for the continuous outcomes. Recall that we have multiple outcomes observed for each individual, who are nested within clusters, which are nested within regions. Thus, for outcome c for individual i in cluster k of region r, we let specify the *individual-level* likelihood

$$Y_{rkic}|\mu_{rkic}, \omega_c^2 \sim N(\mu_{rkic}, \omega_c^2).$$

which has outcome-specific, individual-level variances ω_c^2. Of note, we do not explicitly use a multivariate likelihood across the outcomes, but instead we let the shared components induce dependency among outcomes, which is similar to continuous coregionalization models (Schmidt & Gelfand, 2003).

The discrete spatial shared component model for μ_{rkic} must be developed with regard to the complex survey design used to sample the observed data. We will consider the sampling design of the 2014 KDHS (along with most other DHS surveys), which is a stratified cluster survey. The strata consist of the geographies for the first administrative level (admin 1) crossed with the binary classification of urban or rural. Within these strata, clusters were sampled with probability proportional to size. Within a cluster, individuals are selected via simple random sampling. To account for this sampling design, as discussed in Section 4.2, we must include strata effects in our model. Due to our desire to include spatial random effects, we do not use fixed intercepts for strata because that would cause identifiability problems—rather, we use random effects on admin 1 regions along with fixed effects on urban/rural strata for each outcome, which imposes a constant urban/rural effect across areas. Following Section 4.2 and assuming two outcomes for simplicity, $c = 1, 2$, the latent parameters are modeled as

$$\begin{aligned}
\mu_{rki1} &= \beta_1 + z_{rk}\gamma_1 \\
&\quad + \sigma_1(\sqrt{1-\rho_1}v_{r1} + \sqrt{\rho_1}u_{r1}) \\
&\quad + \lambda(\sigma_2(\sqrt{1-\rho_2}v_{r2} + \sqrt{\rho_2}u_{r2})) \\
&\quad + \epsilon_{rk1} \\
\mu_{rki2} &= \beta_2 + z_{rk}\gamma_2 \\
&\quad + \sigma_2(\sqrt{1-\rho_2}v_{r2} + \sqrt{\rho_2}u_{r2}) \\
&\quad + \epsilon_{rk2}
\end{aligned}$$

As fixed effects, β_c are outcome-specific fixed intercepts, z_{rk} is the binary indicator of cluster k in region r being rural (rather than urban), and γ_c is the outcome-specific effect associated with rural areas.

For the discrete spatial component, we use the BYM2 model with $v_{rc} \overset{IID}{\sim} N(0,1)$ and $\boldsymbol{u}_c \sim ICAR(1)$ with total variance σ_c^2 and percent of the variance spatial ρ_c. For the shared component, we include the BYM2 random effects from outcome $c = 2$ in the parameterization for the mean of outcome $c = 1$. which are scaled by a coefficient λ to be estimated from the model. This is a proxy for unobserved shared covariates between our outcomes. We could alternatively share the random effects from outcome $c = 1$; these two models are equivalent (see Appendix B.2).

The cluster-level error $\epsilon_{rkc}|\sigma^2_{\epsilon,c} \overset{iid}{\sim} N(0, \sigma^2_{\epsilon,c})$ is the so-called nugget, which we consider as measurement error in this context and thus will not include it when making model-based predictions. We will assume no covariates for simplicity, though this model can include them with (potentially outcome-specific) fixed effects. Many surveys, the DHS included, collect a suite of relevant variables that we may wish to adjust for, which can be explored in future analyses.

In order to calculate area-level means, we aggregate the estimates from our model via

$$\mu_{rc} = (1 - q_r) \times \left(\beta_c + g_c\right) + q_r \times \left(\beta_c + \gamma_c + g_c\right)$$

with g_c the (potentially shared) BYM2 random effects corresponding to outcome c and q_r the proportion of the relevant population in region r that is rural. For example, when modeling height- and weight-for age in children under 5, q_r is the proportion of the under-5 population in region r that live in rural areas.

While this model has many appealing features, it may not be the most suitable for many scenarios. In general, it is recommended to fit multiple candidate models and compare them via a relevant model selection technique. In our motivating examples, we will compare the above model to three other simpler models in order to see if a model with this level of

complexity is justified. Letting $v^*_{rc}|\sigma^2_c \overset{iid}{\sim} N(0, \sigma^2_c)$ for each c (which will be used in models without spatial random effects), the four candidate models are:

I: IID nonshared

$$\mu_{rki1} = \beta_1 + z_{rk}\gamma_1 + v^*_{r1} + \epsilon_{rk1}$$
$$\mu_{rki2} = \beta_2 + z_{rk}\gamma_2 + v^*_{r2} + \epsilon_{rk2}$$

II: BYM nonshared

$$\mu_{rki1} = \beta_1 + z_{rk}\gamma_1 + \sigma_1(\sqrt{1-\rho_1}v_{r1} + \sqrt{\rho_1}u_{r1}) + \epsilon_{rk1}$$
$$\mu_{rki2} = \beta_2 + z_{rk}\gamma_2 + \sigma_2(\sqrt{1-\rho_2}v_{r2} + \sqrt{\rho_2}u_{r2}) + \epsilon_{rk2}$$

III: IID shared

$$\mu_{rki1} = \beta_1 + z_{rk}\gamma_1 + v^*_{r1} + \lambda v_{r2} + \epsilon_{rk1}$$
$$\mu_{rki2} = \beta_2 + z_{rk}\gamma_2 + v^*_{r2} + \epsilon_{rk2}$$

IV: BYM shared

$$\begin{aligned}\mu_{rki1} = \beta_1 + z_{rk}\gamma_1 &+ \sigma_1(\sqrt{1-\rho_1}v_{r1} + \sqrt{\rho_1}u_{r1}) \\ &+ \lambda(\sigma_2(\sqrt{1-\rho_2}v_{r2} + \sqrt{\rho_2}u_{r2})) + \epsilon_{rk1}\end{aligned}$$
$$\mu_{rki2} = \beta_2 + z_{rk}\gamma_2 + \sigma_2(\sqrt{1-\rho_2}v_{r2} + \sqrt{\rho_2}u_{r2}) + \epsilon_{rk2}$$

4.3.2 Model fitting

We take a Bayesian approach to modeling in this chapter. All models were fit within the R statistical computing environment (R Core Team, 2013) using Integrated Nested Laplace Approximation (INLA) (Rue et al., 2009) as implemented in the `INLA` package. INLA provides a fast alternative to MCMC for approximating the marginal posterior distributions of

latent Gaussian Markov random field models. The models in this chapter are easy to specify and fast to compute using personal computing machines. All code used in this chapter is available at https://github.com/aeschuma/multivariate-sae.

4.3.3 Model selection

Model selection among candidate models will be done via cross validation using a multivariate scoring strategy. Fuglstad et al. (2021) provides the general strategy for this task for univariate models, which we extent to the multivariate case. As the area-level means are the target of estimation, we design a procedure that validates at the area-level. Furthermore, Fuglstad et al. (2021) demonstrate that cluster-level cross validation may have insufficient discriminatory power in models similar to the ones we fit here. Each Bayesian model gives rise to a predictive distribution of the true area-level mean vectors, $\boldsymbol{\mu}_r$. If we hold out all data in region r when fitting the model, the direct estimate of the held out data (i.e. the Hàjek estimator $\hat{\tilde{\boldsymbol{y}}}_r \sim N_C(\boldsymbol{\mu}_r, \hat{\boldsymbol{V}}_r^{des})$ with $\hat{\boldsymbol{V}}_r^{des}$ the design-based covariance matrix, as described in Chapter 3) is independent of this predictive distribution from the model fit with region r removed because it is based on independent data. Thus, combining the independent sampling distribution of the direct estimator for the held out region with the posterior for the true value from the model fit with the held out data removed, we find a predictive distribution for the held out data that we can use to validate estimates from the model at the area-level.

For each region r, the procedure is as follows:

1. Hold out all observations for region r.
2. Fit all candidate models to the data with region r removed.
3. Sample $s = 1, \ldots, S$ samples from the posterior distributions of the region-specific bivariate means, $\boldsymbol{\mu}_r^{(s)} = [\mu_{r1}^{(s)}, \mu_{r2}^{(s)}]$, where $\mu_{rc}^{(s)}$ are the weighted averages of the urban and rural estimates, i.e. $\mu_{rc}^{(s)} = \left[(1-q_r) \times \left(\beta_c^{(s)} + g_c^{(s)}\right)\right] + \left[q_r \times \left(\beta_c^{(s)} + \gamma_c^{(s)} + g_c^{(s)}\right)\right]$ with the $\beta_c^{(s)}$, $\gamma_c^{(s)}$, and $g_c^{(s)}$ parameters as defined previously with the (s) superscript denoting the posterior estimate in sample s. Remember, these samples are from models

fit to data with region r removed.

4. Compute the survey-weighted direct estimates of the means for the held out data, $\hat{\tilde{\boldsymbol{y}}}_{rc}$, along with the appropriate design-based covariance $\hat{\boldsymbol{V}}_r^{des}$. See Chapter 3 Section 3.3 for details on this calculation.
5. For each sample s, the posterior predictive distribution of the held out data is $N_C(\boldsymbol{\mu}_r^{(s)}, \hat{\boldsymbol{V}}_r^{des})$, and we evaluate this distribution at the value $\hat{\tilde{\boldsymbol{y}}}_{rc}$. We call these likelihood values $\tilde{\pi}_r^{(s)}$.
6. Average these over all samples and calculate the negative loglikelihood, $\ell_r = -\log\left(\frac{1}{S}\sum_{s=1}^{S}\tilde{\pi}_r^{(s)}\right)$.

Once we have all of these region-level scoring estimates, we average over them and end up with

$$\text{LogScore} = \frac{1}{R}\sum_{r=1}^{R}\ell_r$$

We compare the LogScore values among all candidate models. Since the LogScore is the average of negative loglikelihoods, a lower LogScore indicates a better performing model.

4.3.4 Simulation study

We perform a simulation study to show the validity of this modeling procedure and the explore how model misspecification of the spatial model and shared component affects results. We will generate individual-level data similar to our working example of HAZ and WAZ scores in the 2014 KDHS. Then, we will fit the above four candidate models to the simulated data. We will do this for seven different scenarios each with their own data generating mechanism. In all scenarios, the parameter values used to generate the latent means will be set equal to parameters estimated from the described model fit to the HAZ and WAZ data from the 2014 KDHS. In each scenario, we will compare the bias, variance, mean squared error (MSE), and 95% interval coverage and width for the aggregated area-level means from four different

models.

Data will be generated using a normal likelihood, $Y_{rkic}|\mu_{rkic}, \omega_c^2 \sim N(\mu_{rkic}, \omega_c^2)$. We now describe μ_{rkic} and ω_c^2 for each scenario.

Scenario 1 - IID nonshared, 2014 KDHS parameters: This scenario will generate latent means from the IID nonshared candidate model, with the parameters, including the variance of the Gaussian observations ω_c^2, set equal to those estimated from the IID nonshared model to the 2014 KDHS. Thus, the means will have the form

$$\mu_{rk1} = \beta_1 + z_{rk}\gamma_1 + v_{r1}^* + \epsilon_{rk1}$$
$$\mu_{rk2} = \beta_2 + z_{rk}\gamma_2 + v_{r2}^* + \epsilon_{rk2}$$

Scenario 2 - BYM nonshared, 2014 KDHS parameters: This scenario will generate latent means from the BYM nonshared candidate model, with the parameters, including the variance of the Gaussian observation ω_c^2, set equal to those estimated from the BYM nonshared model to the 2014 KDHS. Thus, the means will have the form

$$\mu_{rk1} = \beta_1 + z_{rk}\gamma_1 + \sigma_1(\sqrt{1-\rho_1}v_{r1} + \sqrt{\rho_1}u_{r1}) + \epsilon_{rk1}$$
$$\mu_{rk2} = \beta_2 + z_{rk}\gamma_2 + \sigma_2(\sqrt{1-\rho_2}v_{r2} + \sqrt{\rho_2}u_{r2}) + \epsilon_{rk2}$$

Scenario 3 - IID shared, 2014 KDHS parameters: This scenario will generate latent means from the IID shared candidate model, with the parameters, including the variance of the Gaussian observation ω_c^2, set equal to those estimated from the IID shared model to the 2014 KDHS. Thus, the means will have the form

$$\mu_{rk1} = \beta_1 + z_{rk}\gamma_1 + v_{r1}^* + \lambda v_{r2} + \epsilon_{rk1}$$
$$\mu_{rk2} = \beta_2 + z_{rk}\gamma_2 + v_{r2}^* + \epsilon_{rk2}$$

Scenario 4 - BYM shared, 2014 KDHS parameters: This scenario will generate latent means from the BYM shared candidate model, with the parameters, including the variance of the Gaussian observation ω_c^2, set equal to those estimated from the BYM shared model to the 2014 KDHS. Thus, the means will have the form

$$\begin{aligned}\mu_{rk1} &= \beta_1 + z_{rk}\gamma_1 + \sigma_1(\sqrt{1-\rho_1}v_{r1} + \sqrt{\rho_1}u_{r1}) \\ &\quad + \lambda(\sigma_2(\sqrt{1-\rho_2}v_{r2} + \sqrt{\rho_2}u_{r2})) + \epsilon_{rk1} \\ \mu_{rk2} &= \beta_2 + z_{rk}\gamma_2 + \sigma_2(\sqrt{1-\rho_2}v_{r2} + \sqrt{\rho_2}u_{r2}) + \epsilon_{rk2}\end{aligned}$$

Scenario 5, BYM shared, larger variance of the individual-level likelihood: This scenario will use the same generating mechanism as Scenario 1 except the standard deviation of the Gaussian observations will be set to be approximately 50% larger than was estimated in the 2014 KDHS data.

Scenario 6, BYM shared, smaller variance of the individual-level likelihood: This scenario will use the same generating mechanism as Scenario 1 except the standard deviation of the Gaussian observations will be set to be approximately 80% smaller than was estimated in the 2014 KDHS data.

Scenario 7, BYM shared, one precise and one imprecise outcome: This scenario will use the same generating mechanism as Scenario 1 except the standard deviation of the Gaussian observations for outcome 1 will be set to be twice as large as the standard deviation for outcome 2. These values will be approximately 15% larger and 35% smaller than those observed in the 2014 KDHS.

Simulation results: Tables of simulation results are presented in Appendix C.1. We graph the bias, absolute bias, variance, and MSE of the estimated area-level latent means for each

outcome, along with the coverage and width of 95% credible intervals, averaged over all regions, in Figure 4.1 for scenarios 1 through 4 and Figure 4.2 for scenarios 5 through 7. Starting with scenarios 1–4, we see little difference in bias among the models. In scenarios 1–3, the variance, MSE, coverage, and width are actually quite comparable, with no pattern emerging of the correctly specified model outperforming the rest. In the BYM data, we do see undercoverage which is indicative of a strong impact of the spatial modeling components which perhaps induce too much shrinkage. In scenario 4, the BYM shared data generating mechanism, we see that the correctly specified model does have the lowest MSE, which is driven by having the lowest variance, but this actually leads to it having the lowest coverage due to the narrowest intervals. This tradeoff exposes a potentially tricky decision when choosing a model, as the best fitting model may not have optimal coverage while providing more precise estimation. For scenario 6, which has very low variance of the Gaussian likelihood, all models perform similarly as to be expected since there is little shrinkage. Scenario 5, with large Gaussian variance, and scenario 7, with one larger and one smaller Gaussian variance, demonstrate similar issues as seen in scenario 4. Notably, however, scenario 7 also shows a markedly lower absolute bias in the outcome with the larger variance, which indicates the utility of a shared component to borrow information from a precise outcome to aid estimation of a more diffuse outcome.

4.4 Unit-level modeling of height- and weight-for-age from the 2014 Kenya DHS

We model HAZ and WAZ scores from the 2014 KDHS as our first working example. For a detailed description of the 2014 KDHS survey as well as the HAZ and WAZ scores, see Chapter 3.

4.4.1 Modeling specifications

HAZ and WAZ are both continuous variables, so we will use the framework described in Section 4.3.1. We will fit candidate models I, II, III, and IV to the data and compare them

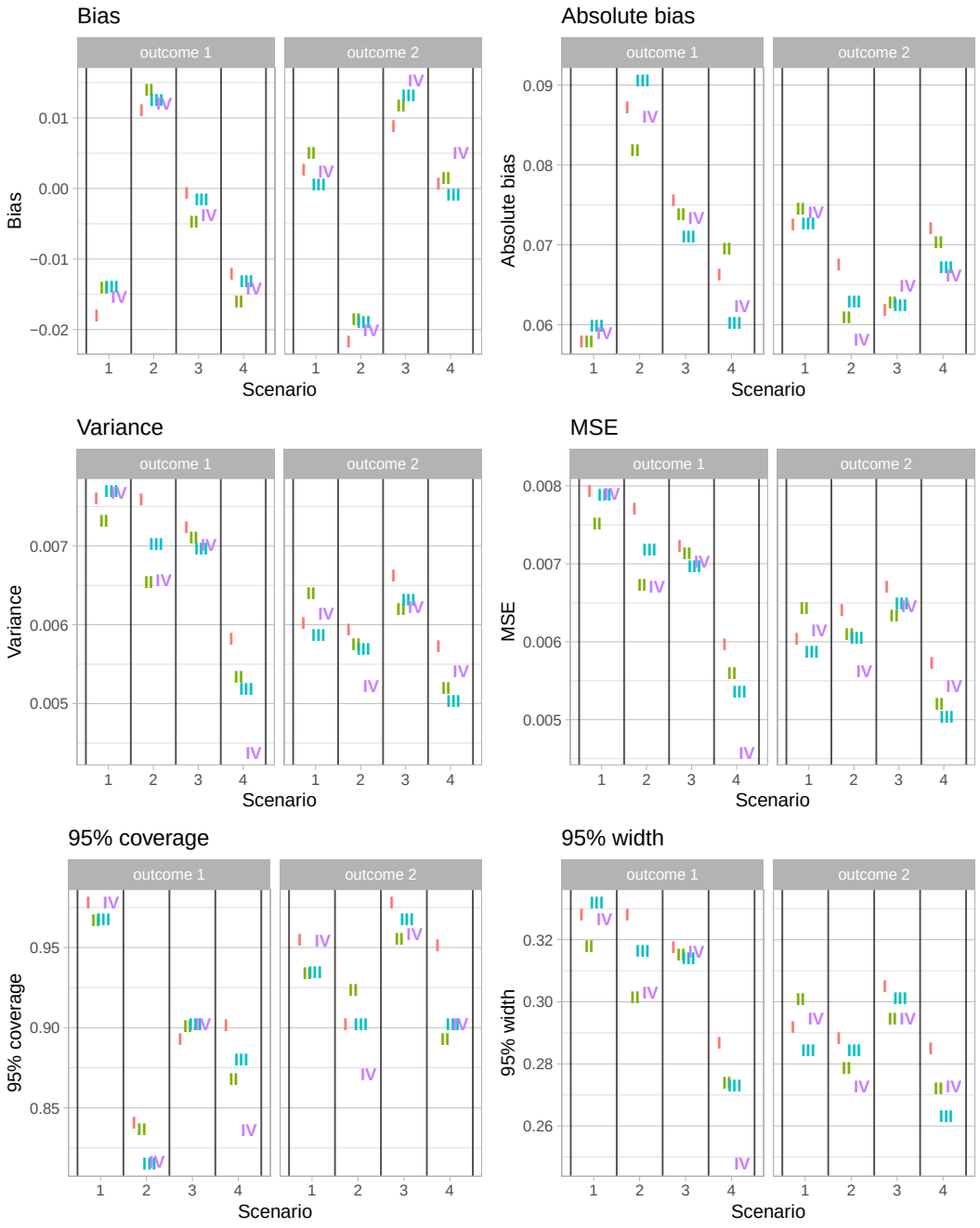


Figure 4.1: Bias, absoulte bias, variance, MSE, 95% interval coverage, and 95% interval width for all models across simulation scenarios 1–4. Scenario descriptions: 1 - BYM shared; 2 - IID shared; 3 - BYM nonshared; 4 - IID nonshared. Model descriptions: I - IID nonshared; II - BYM nonshared; III - IID shared; IV - BYM shared.

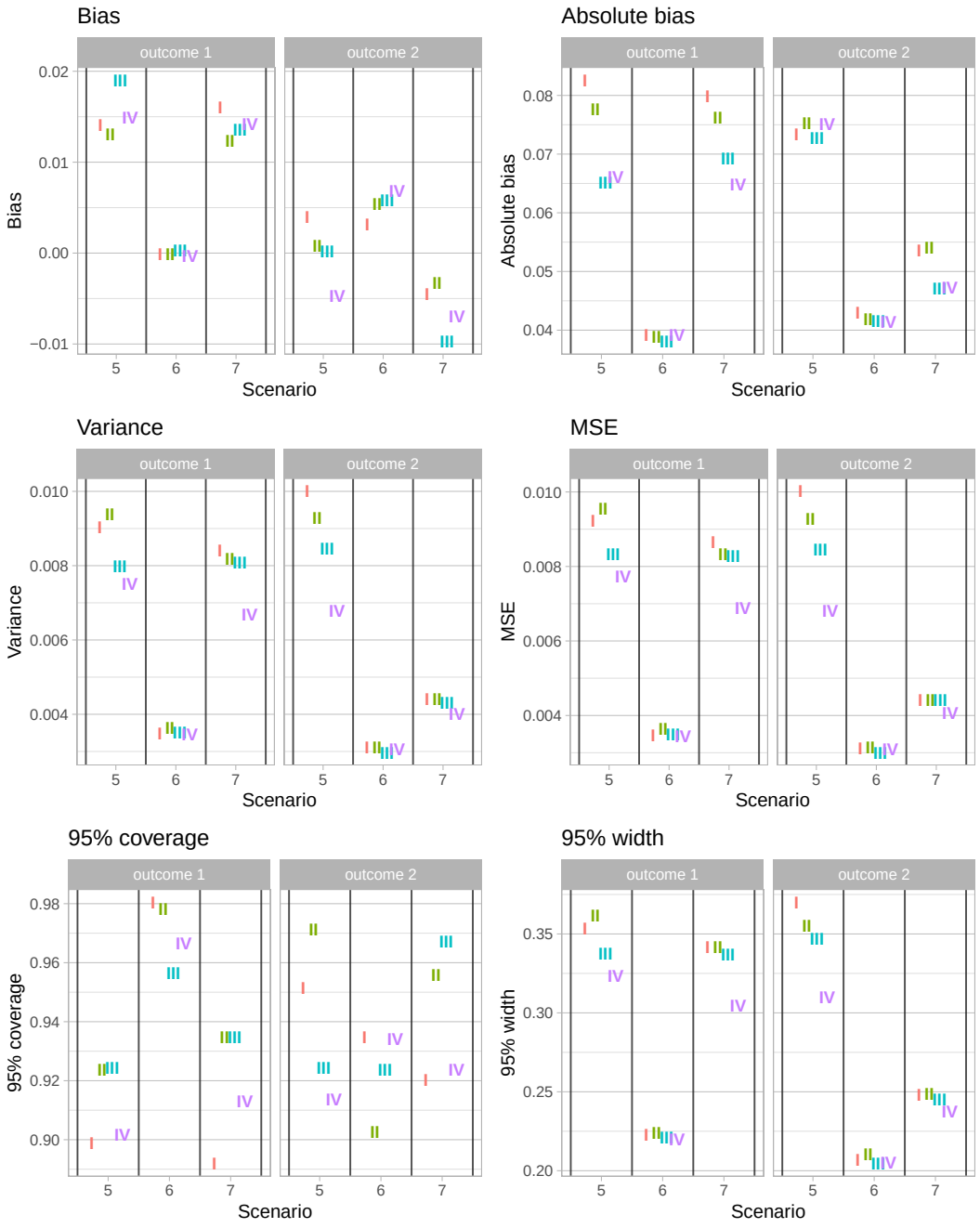


Figure 4.2: Bias, absoulte bias, variance, MSE, 95% interval coverage, and 95% interval width for all models across simulation scenarios 5–7. Scenario descriptions: 5 - BYM shared, small first stage variances; 6 - BYM shared, small first stage variances; 7 - BYM shared, one larger and one smaller first stage variance. Model descriptions: I: IID nonshared; II: BYM nonshared; III: IID shared; IV: BYM shared.

using the LogScore to choose a best-performing model, for which we will present results.

As we are using Bayesian inference, we must specify prior distributions. Since this example has a relatively large amount of data, we use quite noninformative priors. For the fixed effects we use improper flat priors. For the IID random effects in IID-only models (I, III), we use penalized complexity (PC) priors (Simpson et al., 2017) on the standard deviations such that there is a 1% chance to be larger than 1. For BYM random effects, we use PC priors on the total variance of the IID and ICAR random effects such that there is a 1% probability that the total variance is greater than 1. We use Beta(1, 1) prior on ρ_c, the percent of the variation that is spatial, as recommended in Riebler et al. (2016).

For the aggregation of estimates to the area-level, we need to use the proportion of the under-5 population in each county that is rural (in 2014), q_r. These were calculated from WorldPop urban/rural classification and population density rasters ("WorldPop," 2022) in conjunction with admin-1 urban population fractions from the DHS (Kenya National Bureau of Statistics, 2015) using the thresholding method described in Wu et al. (2021).

4.4.2 Results

We compare the posterior medians of the area-level mean estimates of HAZ and WAZ among all models in Figure 4.3. These results show quite strong agreement between all models, which is likely due to the amount of data being modeled. We present a comparison of the posterior standard deviations of the latent means in Figure 4.4. Here, we see that the BYM shared models has the smallest posterior standard deviations. However, as discussed in Section 4.3.4, the posterior variance may be too small leading to uncertainty intervals that are too narrow compared to the nominal level.

To make a final decision on the best performing model, we present a comparison of the LogScore for each of the four candidate models in Table 4.1. Recall that a lower value indicates better performance, since the LogScore is based on the negative log-likelihood. We see that Model IV has the lowest LogScore. This makes sense since HAZ and WAZ have

strong spatial dependence, and additionally there are likely shared unobserved factors that influence both HAZ and WAZ.

Due to this model selection exercise, we present the area-level estimates of mean HAZ and WAZ scores from the 2014 KDHS as the results of model IV. However, we do emphasize that for these results the uncertainty intervals might be too narrow to have nominal coverage (as seen in the previous simulation study) despite it fitting the data better than the other three models. Summaries of the posterior distributions are presented in Table 4.2. Figure 4.5 shows a map of the posterior median HAZ and WAZ scores along with the widths of the 95% credible intervals. Additionally, we plot maps of the total BYM2 random effects, $\sigma_c(\sqrt{1-\rho_c}v_{rc} + \sqrt{\rho_c}u_{rc})$, along with the standardized IID and ICAR components, v_{rc} and u_{rc}, for both HAZ and WAZ in Figure 4.6.

For the posterior median WAZ estimates, we see a more consistent spatial gradient of being lower in the north and higher in the south as compared to HAZ. At least for WAZ, these findings agree with the general knowledge that access to and quality of health care and proper nutrients tend to be worse in the northern regions (Ilinca, Di Giorgio, Salari, & Chuma, 2019). To explore this finding further, we look at the maps of the random effects. This north-south heterogeneity is evident in the shared ICAR random effect, with additional hotspots of high WAZ in the south-center and southwest which arise from the shared IID random effect. This is the complete picture for the WAZ final estimates, but recall that the HAZ final estimates include both the shared random effects and HAZ-specific random effects. Thus, we can interpret the HAZ ICAR random effects as essentially the difference between the WAZ spatial gradient and the HAZ spatial gradient, and similarly the HAZ IID random effects constitute the non-structured differences. With this in mind, it is clear that HAZ has less of a stark north-south gradient since the HAZ random effects are generally in the opposite direction as the shared random effects, especially in the northern regions. For context, the spatial distribution seen here, which arises for other health indicators as well (Godwin & Wakefield, 2021; Mulatya & Ochieng, 2020), is commonly attributed to the rural-urban health divide (Joseph et al., 2020). But the geographical heterogeneity estimated from

Table 4.1: Bivariate LogScore for each model fit to continuous HAZ and WAZ data from the 2014 KDHS. Bold indicates the best performing model.

	Model	LogScore
(III)	IID shared	0.326
(I)	IID nonshared	0.21
(II)	BYM nonshared	-0.348
(IV)	**BYM shared**	**-0.454**

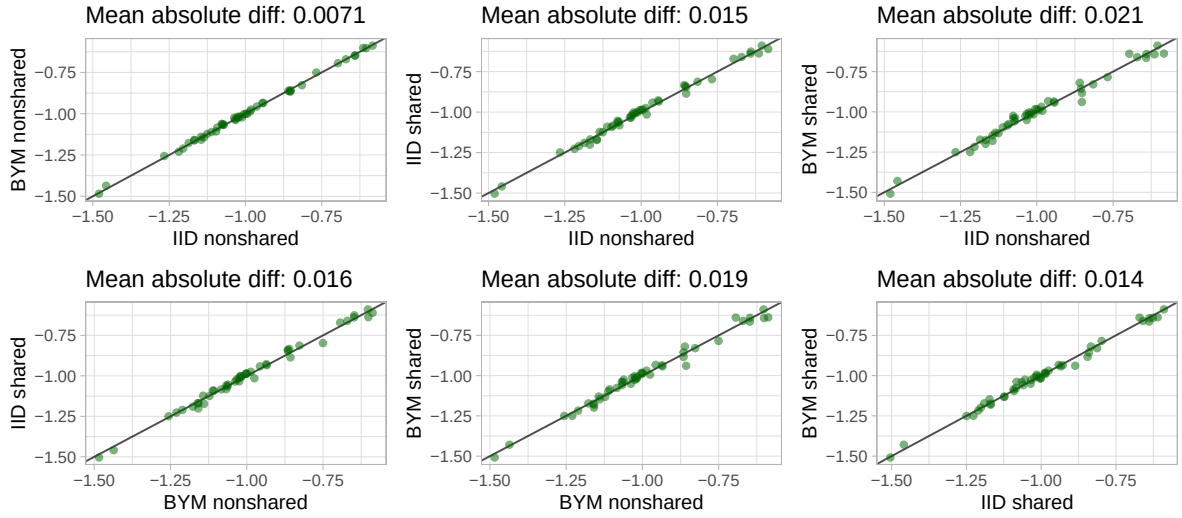
the BYM2 random effects is *in addition to* the strong urban/rural effect that was estimated from our model ($\hat{\gamma}_1 \approx \hat{\gamma}_2 \approx -0.29$). This indicates complex interactions of multiple factors in the spatial distribution of HAZ and, most strongly, WAZ, beyond the urban-rural gap—a finding that is supported in recent literature (Fagbamigbe, Kandala, & Uthman, 2020).

4.5 Unit-level modeling of contraceptive use from the 2014 Kenya DHS

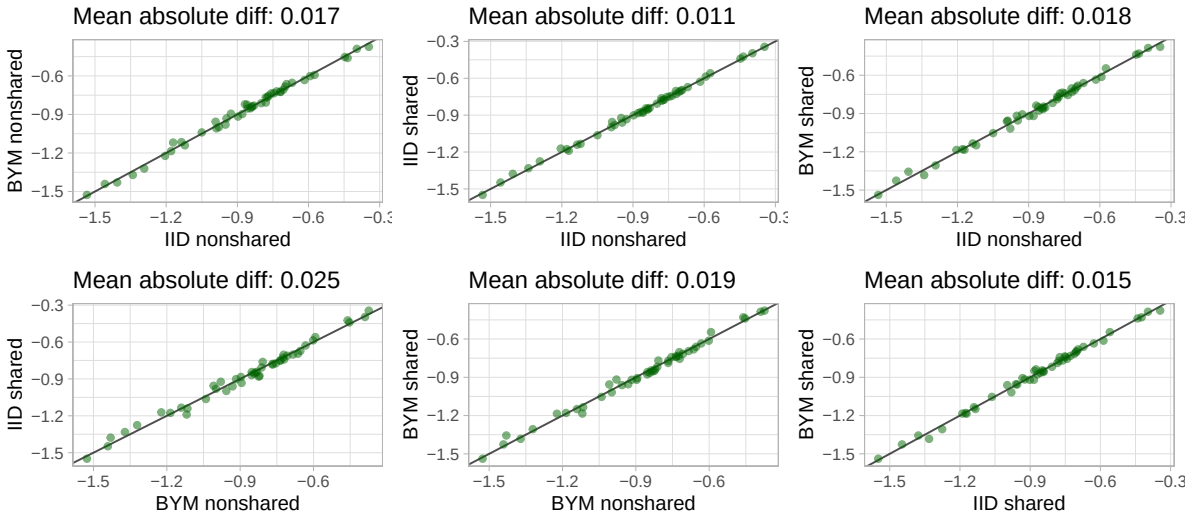
Next, we consider modeling contraceptive use in women age 15-49, again using data from the 2014 KDHS. For a description of this contraceptive use data, see Chapter 3. Importantly, we have three categories of contraceptive use: none, modern methods, or other methods. We will start by detailing the necessary adaptations to our modeling framework for multinomial data. Then, we will fit four candidate models to the contraceptive use data and compare them using the LogScore to choose a best-performing model, for which we will present results.

4.5.1 Adaptation for multinomial outcomes

Our unit-level multivariate modeling framework can also be adapted for use with multinomial outcomes, with some minor changes. We will use the same notation, except now we will let $c = 1, \ldots, C$ denote levels of a categorical variable. For cluster k in region r, define N_{rk} as



(a) HAZ



(b) WAZ

Figure 4.3: Comparison of posterior median area-level mean estimates of HAZ and WAZ from the 2014 KDHS.

Table 4.2: Posterior medians (Est) and credible intervals (CI) for fixed effects and hyper-parameters from the BYM shared model estimating HAZ and WAZ from the 2014 KDHS data.

Parameter	Est	80% CI	95% CI
β_{haz}	-0.77	(-0.81, -0.73)	(-0.83, -0.71)
β_{waz}	-0.66	(-0.69, -0.63)	(-0.71, -0.61)
γ_{haz}	-0.29	(-0.33, -0.25)	(-0.35, -0.23)
γ_{waz}	-0.29	(-0.32, -0.25)	(-0.34, -0.23)
ω_1	1.31	(1.3, 1.32)	(1.3, 1.32)
ω_1	1.14	(1.13, 1.15)	(1.13, 1.15)
σ_1	0.13	(0.11, 0.15)	(0.1, 0.17)
ρ_1	0.71	(0.39, 0.93)	(0.25, 0.97)
σ_2	0.19	(0.17, 0.21)	(0.16, 0.23)
ρ_2	0.67	(0.5, 0.82)	(0.42, 0.88)
$\sigma_{\epsilon 1}$	0.36	(0.35, 0.38)	(0.34, 0.39)
$\sigma_{\epsilon 2}$	0.33	(0.31, 0.34)	(0.31, 0.35)
λ	0.72	(0.55, 0.91)	(0.46, 1.01)

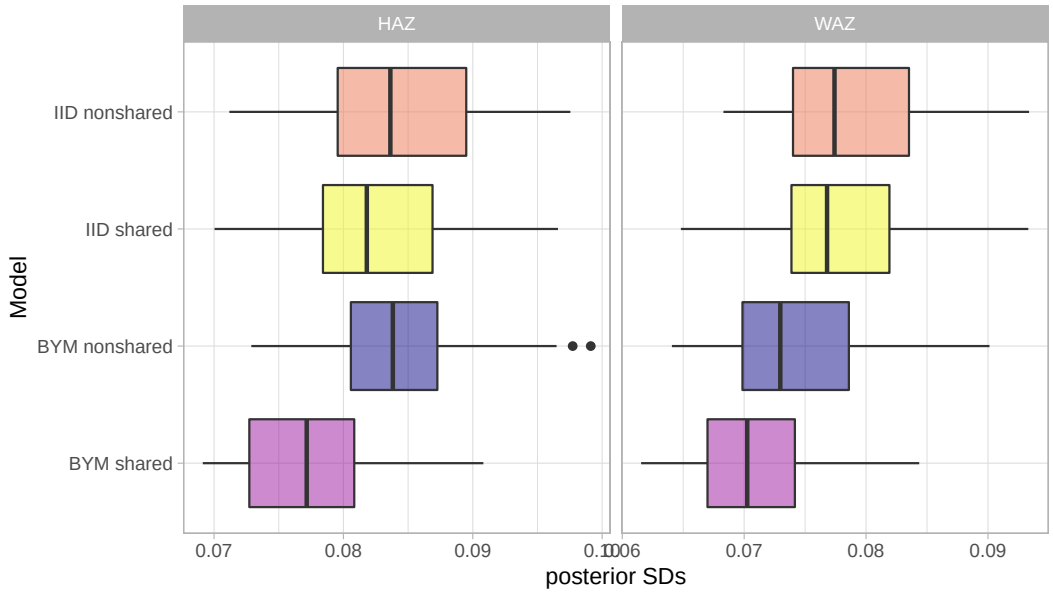


Figure 4.4: Comparison of posterior standard deviations of latent means for each of six models used to estimate HAZ and WAZ from the 2014 KDHS.

the total number of individuals, Y_{rkc} as the number of individuals that are in category c (in vector form $\boldsymbol{Y}_{rk} = [Y_{rk1}, \ldots, Y_{rkC}]'$), with p_{rkc} the associated probability of being in category c (in vector form $\boldsymbol{p}_{rk} = [p_{rk1}, \ldots, p_{rkC}]'$).

First, our cluster-level likelihood is

$$\boldsymbol{Y}_{rk} | N_{rk}, \boldsymbol{p}_{rk} \sim \text{Multinomial}(N_{rk}, \boldsymbol{p}_{rk}). \tag{4.1}$$

We specify a baseline category logistic model (Hartzel et al., 2001) for the $\boldsymbol{p}_{rk}$, with reference category C, such that for outcome $c = 1, \ldots, C-1$ we have

$$\log\left(\frac{p_{rkc}}{p_{rkC}}\right) = \mu_{rkc}$$

For model fitting with the `INLA` package in R, however, there is no multinomial likelihood available. Therefore, we will use the multinomial-Poisson transformation, or the "Poisson

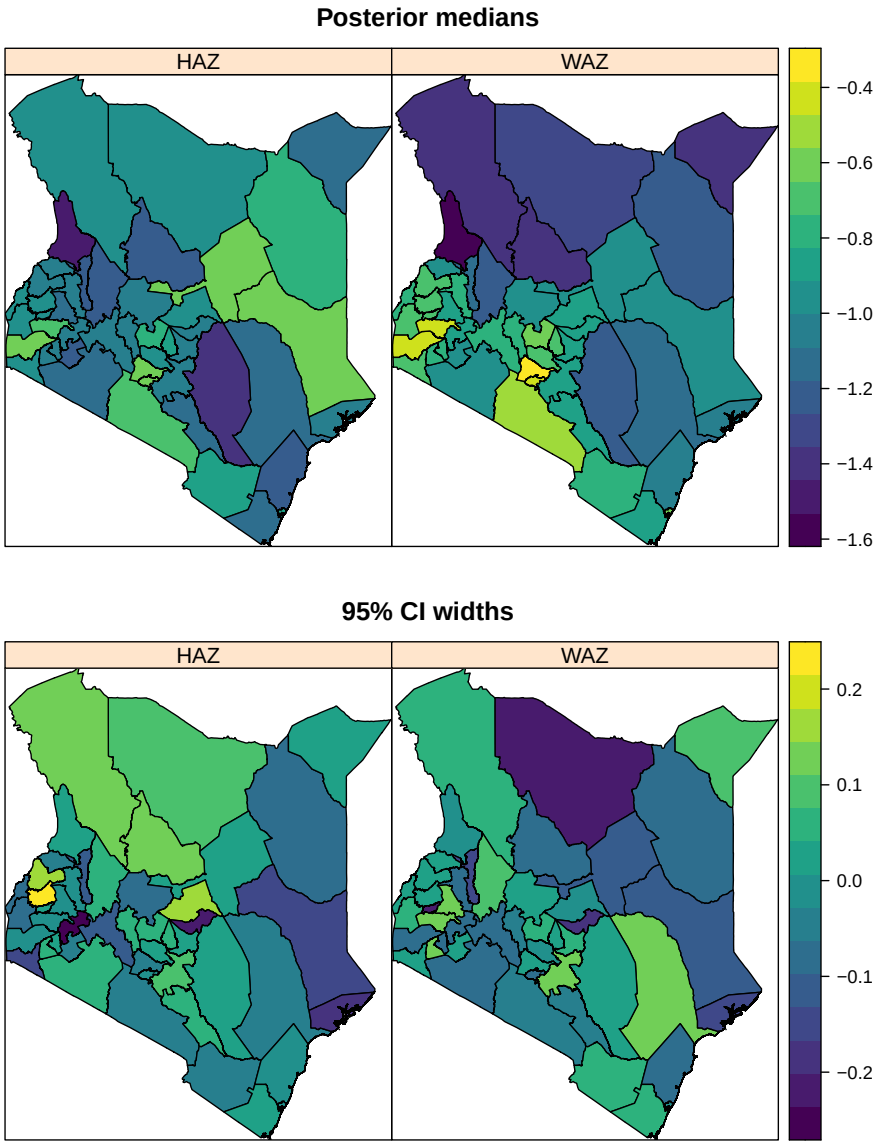


Figure 4.5: Estimated posterior medians and 95% interval widths for HAZ and WAZ estimated from our bivariate shared BYM model from the 2014 KDHS.

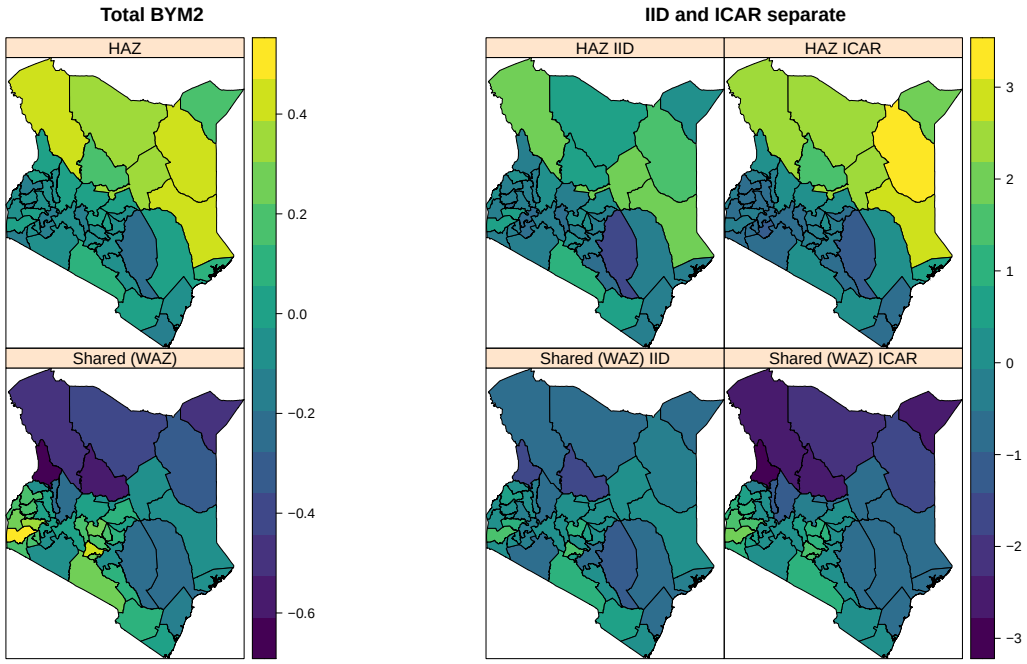


Figure 4.6: Estimated IID and ICAR random effects (normalized to have variance 1) for HAZ and WAZ estimated from our bivariate shared BYM model from the 2014 KDHS.

trick," and use a working Poisson likelihood. Lee et al. (2017) provides a detailed reference for modeling multinomial data via a Poisson distribution from a maximum likelihood perspective. In Appendix C.2, we show that the MLE for the multinomial likelihood is equal to the MLE of a specific Multinomial-Poisson surrogate model.

From the Bayesian perspective, Ghosh, Zhang, & Mukherjee (2006) prove the equivalence of the posterior distributions of our multinomial model and a Poisson model that has additional cluster-specific intercepts with flat priors. To show this, we first drop the r subscripts for simplicity and let $p_{kc} = \frac{\exp(\zeta_{kc})}{\sum_{c'=1}^{C} \exp(\zeta_{kc'})}$ for some formulation of parameters ζ_{kc}. The multinomial likelihood function is

$$\mathcal{L}_M(\boldsymbol{\zeta}) \propto \prod_{k=1}^{K} \prod_{c=1}^{C} \left[\frac{\exp(\zeta_{kc})}{\sum_{c'=1}^{C} \exp(\zeta_{kc'})} \right]^{y_{kc}}.$$

Now, let $\boldsymbol{\phi} = [\phi_1, \ldots, \phi_K]$ be a set of auxiliary parameters, one for each cluster. The equivalent likelihood for the Poisson trick is that $Y_{kc} | \zeta_{kc}, \phi_k \overset{\text{indep}}{\sim} \text{Poisson}(\exp(\phi_k + \zeta_{kc}))$, i.e. each multinomial count is independently Poisson distributed. This likelihood is

$$\mathcal{L}_P(\boldsymbol{\phi}, \boldsymbol{\zeta}) \propto \prod_{k=1}^{K} \prod_{c=1}^{C} [\exp(\phi_k + \zeta_{kc})]^{y_{kc}} [\exp\{-\exp(\phi_k + \zeta_{kc})\}].$$

To see the equivalence between the posterior estimates of $\boldsymbol{\zeta}$ from this likelihood and the multinomial likelihood, we will let $\pi(\boldsymbol{\zeta})$ be a proper prior for $\boldsymbol{\zeta}$ which is independent of $\boldsymbol{\phi}$. Letting $\alpha_k = \exp(\phi_k)$ for all k with improper flat priors $p(\alpha_k) \propto 1/\alpha_k$, we can integrate over the $\boldsymbol{\alpha}$ to get the marginal posterior of $\boldsymbol{\zeta}$ as

$$
\begin{aligned}
\pi(\boldsymbol{\zeta}|\boldsymbol{y}) &\propto \pi(\boldsymbol{\zeta}) \prod_{k=1}^{K} \int_0^{\infty} \left[\alpha_k^{-1} \prod_{c=1}^{C} \{\alpha_k \exp(\zeta_{kc})\}^{y_{kc}} \exp\{-\alpha_k \exp(\zeta_{kc})\} d\alpha_k \right] \\
&= \pi(\boldsymbol{\zeta}) \prod_{k=1}^{K} \left[\prod_{c=1}^{C} \zeta_{kc}^{y_{kc}} \int_0^{\infty} \alpha_k^{\sum_{c=1}^{C} y_{kc} - 1} \exp\left(-\alpha_k \sum_{c=1}^{C} \zeta_{kc} \right) d\alpha_k \right] \\
&\propto \pi(\boldsymbol{\zeta}) \prod_{k=1}^{K} \prod_{c=1}^{C} \left[\frac{\exp(\zeta_{kc})}{\sum_{c=1}^{C} \exp(\zeta_{kc})} \right]^{y_{kc}} \\
&= \pi(\boldsymbol{\zeta}) \mathcal{L}_M(\boldsymbol{\zeta}).
\end{aligned}
$$

which shows the equivalence. The intuition here, as detailed in Lee et al. (2017), is that if we treat the total counts as randomly Poisson distributed, the joint multinomial-Poisson likelihood is the same as if each count were independently Poisson distributed. The additional $\boldsymbol{\phi}$ parameters (there is one for each cluster, i.e., one for each multinomial count) correspond to the the additional randomness of the total counts.

The conclusion of this work is that to model our multinomial data and produce estimates μ_{rkc}, we can instead use a Poisson likelihood such that

$$
\mathbf{Y}_{rkc} | \theta_{rkc} \sim \text{Poisson}(\theta_{rkc})
$$

with $\theta_{rkc} = \exp(\phi_{rk} + \zeta_{rkc})$. Letting $\mu^*_{rkc} = \log(\theta_{rkc})$, we can specify a loglinear model

$$
\begin{aligned}
\log(\theta_{rkc}) &= \mu^*_{rkc} \\
&= \phi_{rk} + \zeta_{rkc}.
\end{aligned}
$$

This relates back to the probabilities that we wish to estimate by the relationship $p_{rkc} = \frac{\theta_{rkc}}{\sum_{c'=1}^{C} \theta_{rkc'}}$.

The first observation here is that the Poisson model requires specifying a cluster-specific intercept with flat priors, ϕ_{rk}. The models we consider in this section, however, already need

cluster-specific intercepts to account for the complex survey design—these were the ϵ_{rkc} nugget terms. We had previously used outcome-specific nuggets, but for modeling multinomial data we will only specify an overall nugget in order to satisfy the posterior distribution equivalency, and these will have flat priors rather than normal priors with an estimated variance hyperparamter.

The second observation is that the Poisson model specifies separate parameters for each category, while the multinomial model only specifies $C - 1$ parameters for the logit probabilities. Thus, we set $\mu^*_{rkC} = 0$ for reference category, which means $\theta_{rkC} = 1$. This also gives the desired relation back to a multinomial model on logit probabilities,

$$\begin{aligned}\mu^*_{rkc} &= \log(\theta_{rkc}) \\ &= \log\left(\frac{\theta_{rkc}}{\theta_{rkC}}\right),\end{aligned}$$

so we see that this Poisson model is essentially still modeling logit probabilities.

With the above considerations, we can now specify the forms of our candidate models for the contraceptive use example where we have $C = 3$ categories. We will use similar candidate models as the continuous outcome case—(I) IID nonshared, (II) BYM nonshared, (III) IID shared, and (IV) BYM shared—and we will specify the same prior distributions as were previously used except the nugget terms will have flat priors. The models are as follows:

I: IID nonshared

$$\begin{aligned}\mu^*_{rk1} &= \beta_1 + z_{rk}\gamma_1 + v^*_{r1} + \epsilon_{rk} \\ \mu^*_{rk2} &= \beta_2 + z_{rk}\gamma_2 + v^*_{r2} + \epsilon_{rk}\end{aligned}$$

II: BYM nonshared

$$\mu^*_{rk1} = \beta_1 + z_{rk}\gamma_1 + \sigma_1(\sqrt{1-\rho_1}v_{r1} + \sqrt{\rho_1}u_{r1}) + \epsilon_{rk}$$
$$\mu^*_{rk2} = \beta_2 + z_{rk}\gamma_2 + \sigma_2(\sqrt{1-\rho_2}v_{r2} + \sqrt{\rho_2}u_{r2}) + \epsilon_{rk}$$

III: IID shared

$$\mu^*_{rk1} = \beta_1 + z_{rk}\gamma_1 + v^*_{r1} + \lambda v_{r2} + \epsilon_{rk}$$
$$\mu^*_{rk2} = \beta_2 + z_{rk}\gamma_2 + v^*_{r2} + \epsilon_{rk}$$

IV: BYM shared

$$\begin{aligned}\mu^*_{rk1} = \beta_1 + z_{rk}\gamma_1 + \sigma_1(\sqrt{1-\rho_1}v_{r1} + \sqrt{\rho_1}u_{r1}) \\ + \lambda(\sigma_2(\sqrt{1-\rho_2}v_{r2} + \sqrt{\rho_2}u_{r2})) + \epsilon_{rk}\end{aligned}$$
$$\mu^*_{rk2} = \beta_2 + z_{rk}\gamma_2 + \sigma_2(\sqrt{1-\rho_2}v_{r2} + \sqrt{\rho_2}u_{r2}) + \epsilon_{rk}$$

Contrary to the area-level models, we will choose the reference category to be "other" contraceptive use. This category has zero counts for two areas, so any models with area-level random effects for this category will not allow the estimation of the value of the random effects in those two areas. While we could estimate them using the posteriors of the other parameters during model fitting, we instead prefer to parameterize the means for the "modern" and "none" categories with random effects and use the other category as the reference. This also results in the random effects being specified for the two categories that have the most observations.

We again fit these as Bayesian models with the `INLA` package in R, specifying the same priors as in the HAZ/WAZ example with the exception that the ϵ_{rk} will be fixed effects with flat priors.

We wish to report area-level probabilities, so we can aggregate our model estimates up to the area-level by

$$p_{rc} = \left[(1-q_r) \times \left(\frac{\theta^u_{rc}}{\sum_{c'=1}^{C}\theta^u_{rc'}}\right)\right] + \left[q_r \times \left(\frac{\theta^l_{rc}}{\sum_{c'=1}^{C}\theta^l_{rc'}}\right)\right]$$

with urban area-level mean estimates $\theta^u_{rc} = \exp(\beta_c + g_{rc})$ and rural area-level mean estimates $\theta^l_{rc} = \exp(\beta_c + \gamma_c + g_{rc})$ for $c = 1, 2$, where g_{rc} are the corresponding (potentially shared) random effects, and $\theta^u_{rc} = \theta^r_{rc} = 1$. For the contraceptive use data, q_r is the proportion of the population of women age 15-49 that is rural in each region. These were calculated by aggregating pixel-level data from WorldPop up to the county-level with the same method that was used for the under-5 population from the HAZ/WAZ example.

For model selection, we use a similar multivariate scoring strategy as described for the multivariate continuous case. The key adaptation is that we perform all calculations on the logit probability scale (with reference category $c = C$), because for validation of the held out data we will be using the direct estimator of the logit probabilities as we did in Chapter 3. Thus, we calculate the posterior predictive distribution of the held out data as $N_{C-1}(\boldsymbol{\mu}_r^{(s)}, \hat{\boldsymbol{V}}_r^{des})$, and we evaluate this distribution at the value of the direct estimator for the logit probabilities in the held out region, $\widehat{\text{logit}(\boldsymbol{p}_r)}$. If any regions do not have observations for one of the outcomes (which happens in two of the regions for the contraceptive use data), then the direct estimator becomes univariate as it can only estimate the logit probability that is not on the boundary. For these regions, we use the univariate posterior distribution of the single logit probability in conjunction with the univariate direct estimate to calculate ℓ_r, the log likelihood contribution.

4.5.2 Results

We compare the posterior medians of the area-level mean estimates of the proportion of contraceptive use in each category among all models in Figure 4.7. These results show quite strong agreement between all models for the modern and none categories, which is likely due to the amount of data being modeled. Notably, we see very minimal amounts of shrinkage, especially comparing the nonshared and shared models. This is different from the area-level modeling results in Chapter 3. For the other category, we do see a lot of differences between the IID and BYM models, as well as some shrinkage comparing the IID

shared and nonshared models, but the BYM nonshared and shared models are fairly similar. The "other" category does have the least amount of data and some fairly small empirical probabilites in a lot of clusters. This stresses the importance of model selection for the smallest categories for which modeling can have a relatively strong impact. We present a comparison of the posterior standard deviations of the latent means in Figure 4.8, separated by contraceptive use category, and the shared BYM model has the lowest overall posterior standard deviations. This agrees with the area-level models which clearly indicated that the shared BYM model had the lowest posterior standard deviations.

A comparison of the LogScore for each of the four candidate models is provided in Table 4.3. Recall that a lower value indicates better performance, since the LogScore is based on the negative log-likelihood. We see that the shared BYM model is the best performing model, which agrees with the area-level models. We proceed with analyzing the results of model IV, the shared BYM model.

Summaries of the posterior distributions are presented in Table 4.4. Figure 4.9 shows a map of the posterior median estimates of the probabilities for each of the three categories along with the width of 95% credible intervals. Additionally, we plot maps of the total BYM2 random effects, $\sigma_c(\sqrt{1-\rho_c}v_{rc} + \sqrt{\rho_c}u_{rc})$, along with the standardized IID and ICAR components, v_{rc} and u_{rc}, for the modern and none categories in Figure 4.10.

One observation is that the final estimated probabilities for "modern" and "none" have opposite north-south spatial gradients with either low or high hotspots in the center of the country, while the "other" category has a north-south spatial gradient without the same central hotspot. This is similar to what was observed in the area-level models, but some additional insight might be that in that this example, there is a very low percentage of observations that use other types of contraceptives. Thus, any spatial random effects between modern and none are already fairly related in opposite directions due to their probabilities needing to sum to a value that's close to 1. However, since modeling is done at the level of the logit probabilities, we can investigate the estimates of the random effects to see how the logit probabilities are related spatially and nonspatially. We see a similar phenomenon

Table 4.3: Bivariate LogScore for each model fit to multinomial contraceptive use data from the 2014 KDHS. Bold indicates the best performing model.

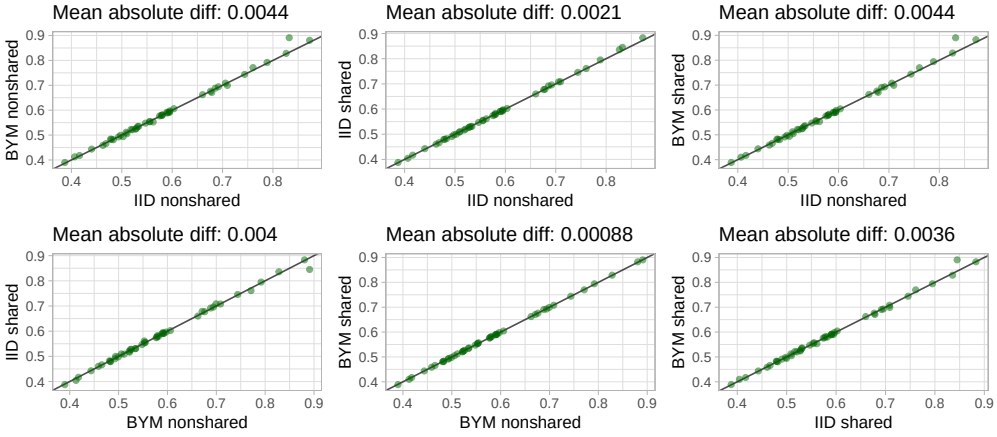
	Model	LogScore
(I)	IID nonshared	2.789
(III)	IID shared	2.684
(II)	BYM nonshared	1.995
(IV)	**BYM shared**	**1.672**

of opposite spatial gradients comparing the ICAR random effects, which shows us that the unit-level and area-level models are capturing similar spatial effects in the data. There is also very little IID nonspatial variability. Looking back at the 95% interval widths for the posterior probabilities, we see a trade-off in the uncertainty interval width, which naturally occurs when modeling dependent probabilities. In counties with low uncertainty for modern contraceptive use, we have higher uncertainty for no contraceptive use and vice versa.

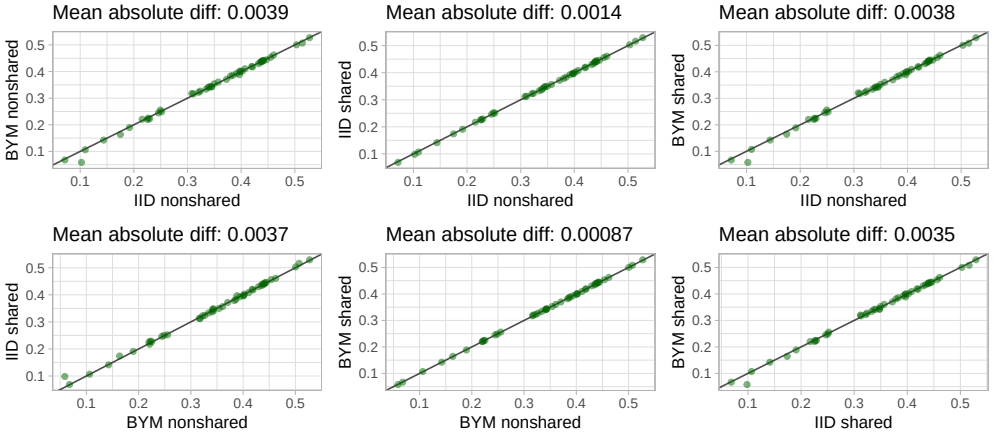
4.6 Discussion

We have developed a class of multivariate unit-level shared component SAE models using complex survey data. We presented simulation results showing how various models in this class compare across multiple scenarios, which showed that in scenarios with lots of data and where there is no true shared component, modeling choice is relatively unimportant; on the other hand, when there is a true shared component and especially when there is also spatial dependence, fitting a model with a shared spatial component is crucial for improving model performance.

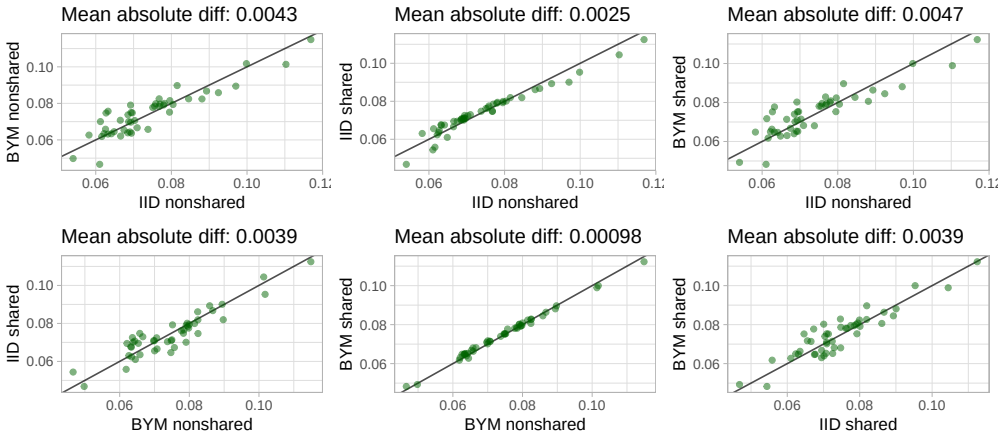
In motivating examples, we jointly modeled height for age and weight for age in children under age 5, as well as contraceptive use in women aged 15-49, using data from the 2014 Kenya DHS survey. In these modeling exercises, we compared suites of models within the



(a) None



(b) Modern



(c) Other

Figure 4.7: Comparison of posterior median area-level mean estimates of contraceptive use from the 2014 KDHS.

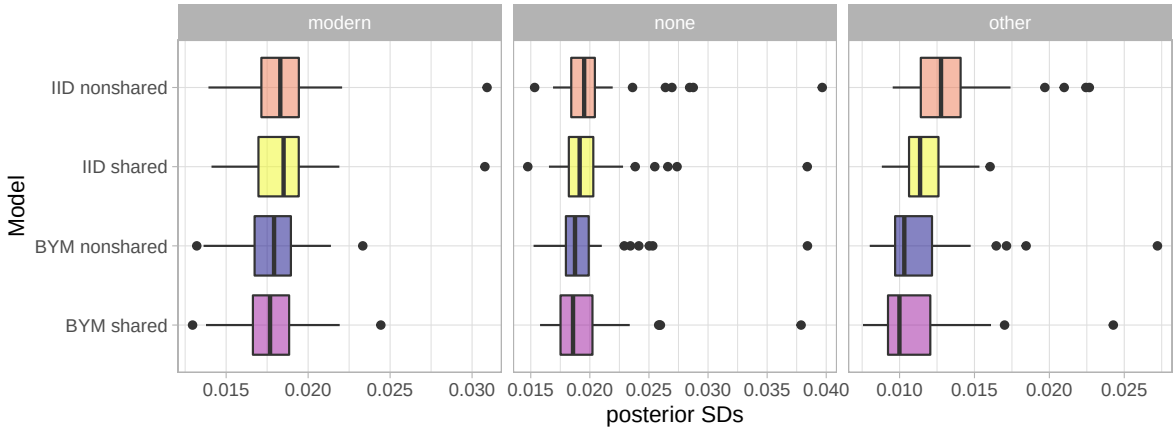


Figure 4.8: Comparison of posterior standard deviations of latent means for each of six models used to estimate contraceptive use from the 2014 KDHS.

Table 4.4: Posterior medians (Est) and credible intervals (CI) for fixed effects and hyper-parameters from the BYM shared model estimating contraceptive use from the 2014 KDHS data.

Parameter	Est	80% CI	95% CI
β_{none}	2.03	(1.94, 2.12)	(1.9, 2.17)
β_{modern}	1.58	(1.5, 1.67)	(1.45, 1.72)
γ_{none}	0.03	(-0.07, 0.13)	(-0.12, 0.18)
γ_{modern}	-0.19	(-0.29, -0.1)	(-0.34, -0.04)
σ_{modern}	0.33	(0.26, 0.4)	(0.23, 0.45)
ρ_{modern}	0.90	(0.72, 0.97)	(0.57, 0.99)
σ_{none}	0.28	(0.21, 0.36)	(0.18, 0.42)
ρ_{none}	0.84	(0.5, 0.97)	(0.31, 0.99)
λ	-0.12	(-0.41, 0.17)	(-0.57, 0.31)

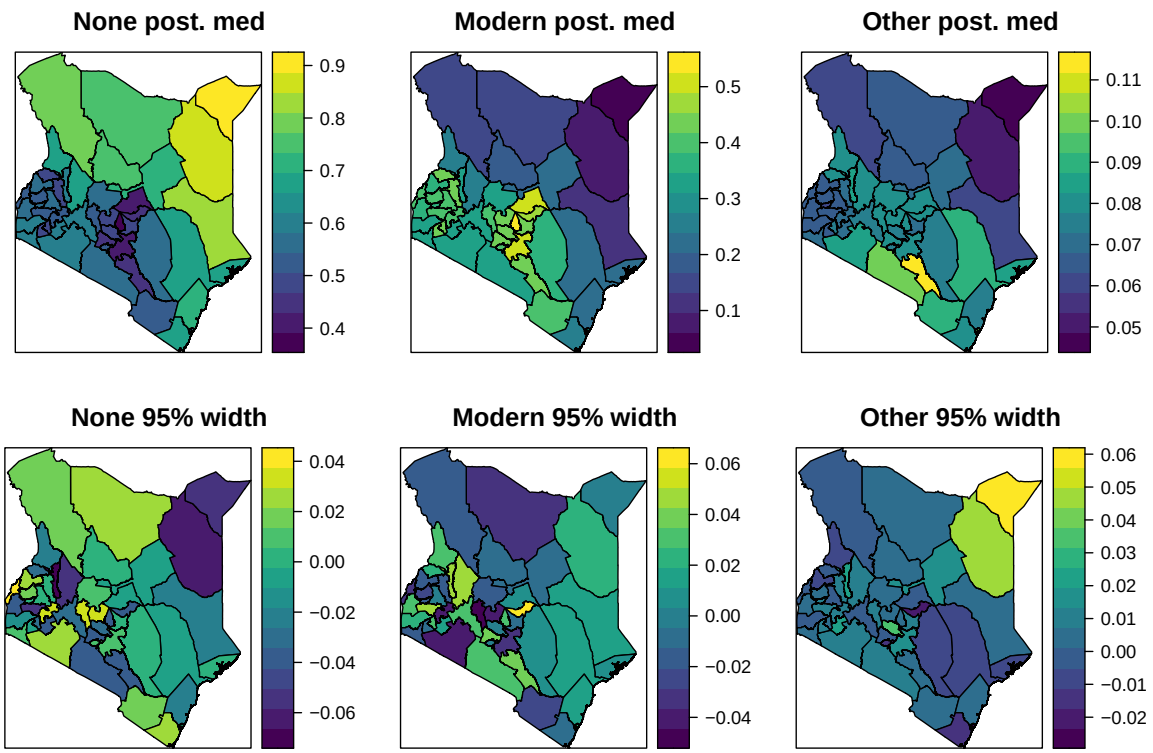


Figure 4.9: Estimated posterior medians and 95% interval widths for contraceptive use proportions estimated with our bivariate shared BYM model from the 2014 KDHS.

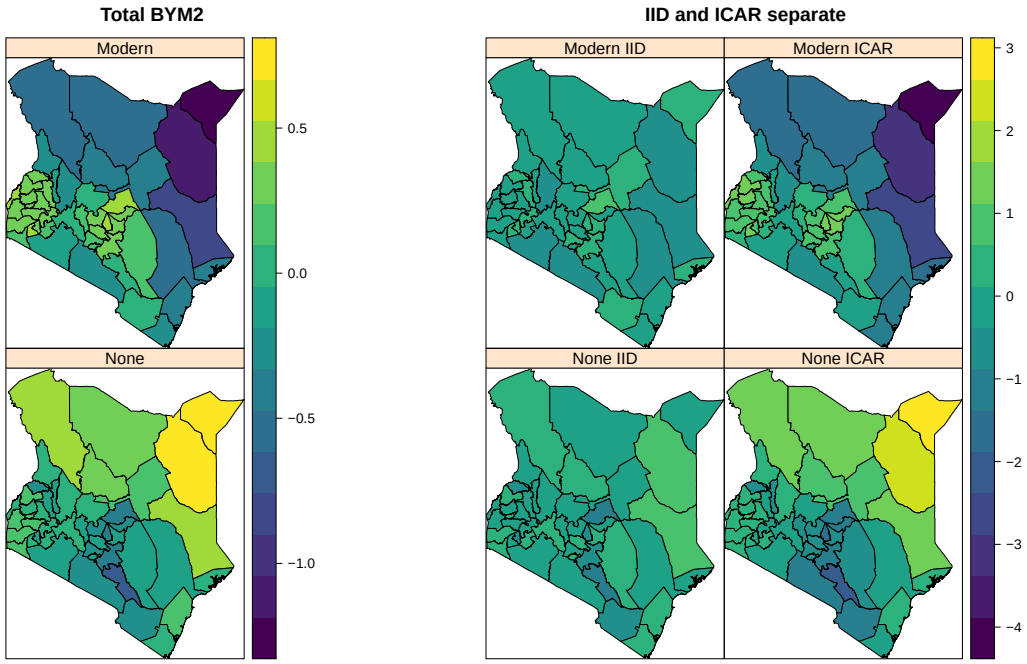


Figure 4.10: Estimated IID and ICAR random effects (normalized to have variance 1) for contraceptive use proportions estimated with our bivariate shared BYM model from the 2014 KDHS.

class we have developed and presented results from the best performing model. This class of models provides a novel method for modeling multivariate outcomes from complex survey data. Unit-level models with shared area-level random effects are enticing for modeling health and demographic data for which we believe underlying unobserved covariates affect multiple outcomes. These models work well for jointly modeling continuous outcomes as well as multinomial outcomes in the examples we explored. Further research into the performance of shared component models for other outcomes is a priority.

Our modeling exercise from the motivating examples use relatively uninformative PC priors because the 2014 KDHS has a relatively large amount of data and sensitivity analyses with more informative prior distributions did not affect estimates greatly. However, in other situations with less data and/or outcomes with more variability, the choice of prior distribution may be important, and further research is needed in order to provide recommendations on informative priors for these models. Research on priors for the shared component are particularly important.

This chapter also restricted the class of models it used to discrete BYM spatial effects which use an ICAR model for spatial dependence. A large portion of spatial modelers prefer continuous spatial models despite the computational burden and the difficulties in aggregating to the area level properly. Our general framework for unit-level modeling does allow for continuous spatial effects. Accordingly, one avenue for future research is to compare continuous and discrete spatial models fit within our framework. Our work can also be extended to include spatiotemporal models. Similarly to the area-level modeling situation, recent developments on spatiotemporal shared component models (Ahmadipanahmehrabadi et al., 2019; Blangiardo et al., 2020; Mahaki et al., 2018; Paradinas et al., 2017) can be explored to figure out what is most applicable in our modeling framework.

In the context of unit-level models, the level of aggregation at which we wish to present results takes special care, particularly with the interplay between aggregation and the desired model to be fit. This chapter solely focused on producing area-level estimates from unit-level models which drove many of our modeling choices. However, different levels of desired

aggregation will lead to different decisions. Further disaggregation for estimates will test the limits of modeling, but these are the exact situations where unit-level models can shine while area-level models cannot produce estimates. Exploring this frontier of research can produce insight into how to effectively specify unit-level models for a finer granularity of target estimates.

Lastly, to reiterate, the differences between area- and unit-level models is an important topic. Deciding between these approaches is crucial for practitioners trying to analyze data and produce estimates. A comparison of area- and unit-level approaches is presented in Chapter 5.

Chapter 5

DISCUSSION AND FUTURE WORK

Jointly modeling multivariate outcomes can be a useful endeavor to produce more reliable estimates of key demographic and health indicators in low- and middle-income countries (LMICs) compared to estimates from separate univariate models. This dissertation developed frameworks for modeling multiple related outcome in three distinct scenarios in order to demonstrate the utility of this statistical approach.

In Chapter 2, we developed a framework for estimating age- and cause-specific child mortality over time using sample registration system (SRS) data. After deriving the modeling framework from first principles, we demonstrated the advantages of our joint modeling approach over other approaches in common use via simulation. Due to the flexibility of our method in terms of allowable model specifications, we provided guidance on how to build models and what issues to consider when doing so. The goal here is to make sure anyone using our framework in their own work has a path to take that is grounded in best practices. Finally, as an illustration of model development with our framework, we demonstrated a way to create a model and validate it in order to estimate age- and cause-specific child mortality over time from the Maternal and Child Health Surveillance System (MCHSS) in China. The methods developed here are widely applicable in many other situations.

Beyond SRS data, Chapters 3 and 4 consider small area estimation (SAE) for multivariate outcomes using complex survey data. In these chapters, we developed both area-level and unit-level shared component models with discrete spatial effects to jointly model multiple continuous outcomes as well as multinomial outcomes. We described how to select between models and provided examples by producing estimates of height-for-age Z scores (HAZ) and weight-for-age Z scores (WAZ) in children as well as contraceptive use in women using data

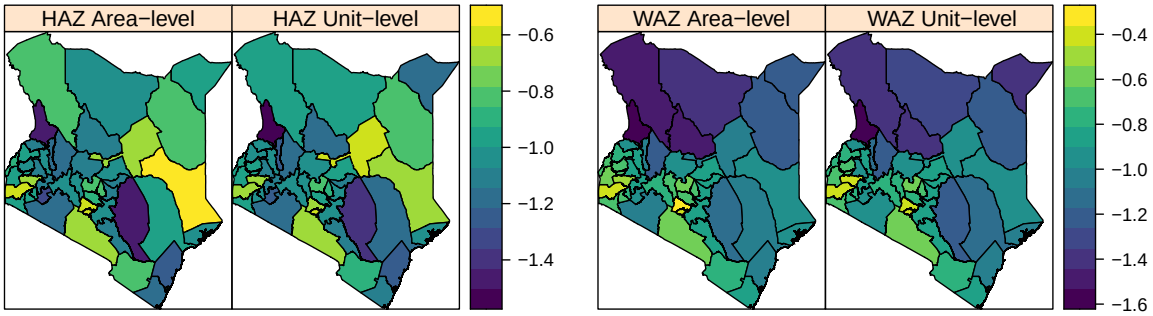


Figure 5.1: Estimated posterior medians from area-level Bivariate shared BYM model and unit-level BYM shared model fit to HAZ and WAZ from the 2014 KDHS.

from the 2014 DHS survey in Kenya.

An important issue that arises is comparing area-level and unit-level models. For the example of modeling HAZ and WAZ, we present maps of the posterior medians for the final Bivariate shared BYM area-level model and the final BYM shared unit-level model in Figure 5.1. These are quite similar, as we see in the comparison scatterplot in Figure 5.2. However, the unit-level model has narrower uncertainty intervals, which is apparent in Figure 5.3. This is consistent with the unit-level modeling simulation, where we saw most scenarios exhibit intervals with low coverage. But this undercoverage was less extreme in the scenario with high observation-level variability. This is a prime issue to consider when choosing what approach to take. If the uncertainty intervals for the area-level models are reasonable and not too prohibitively wide to draw reliable conclusions, then area-level models are preferred. However, when less data is available or data is more noisy, area-level models may exhibit too much posterior uncertainty, in which case unit-level models are preferred. In the example of modeling contraceptive use data, we see that the posterior median estimates from these two models were very similar for the two large categories, but for the category that had small numbers of observations, there were some larger differences. Some of this may be

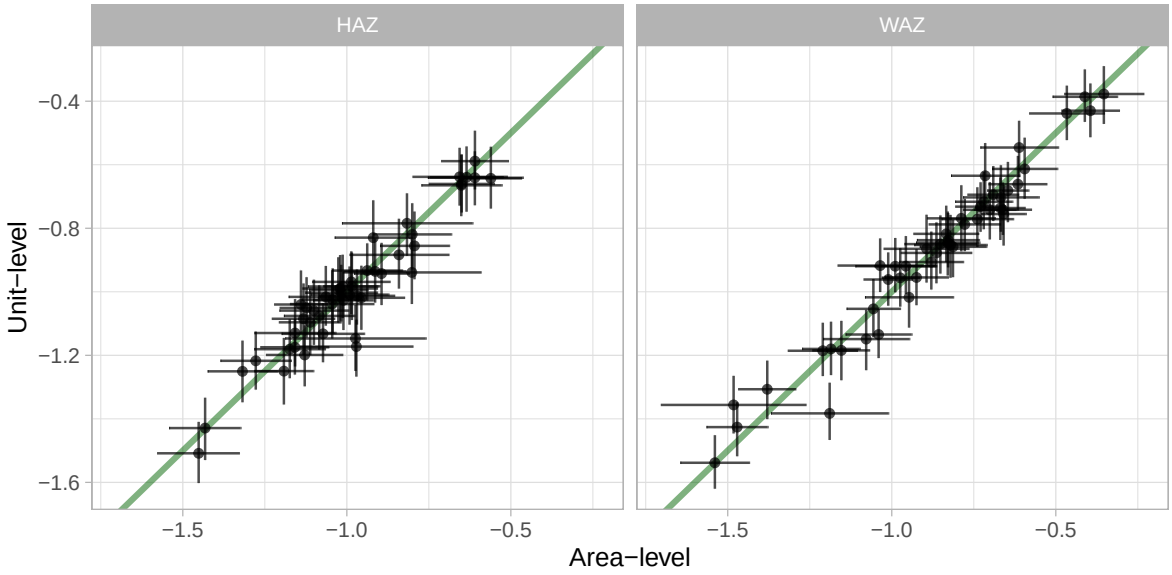


Figure 5.2: Estimated posterior medians and 80% credible intervals from area-level Bivariate shared BYM models vs. unit-level BYM shared models fit to HAZ and WAZ from the 2014 KDHS.

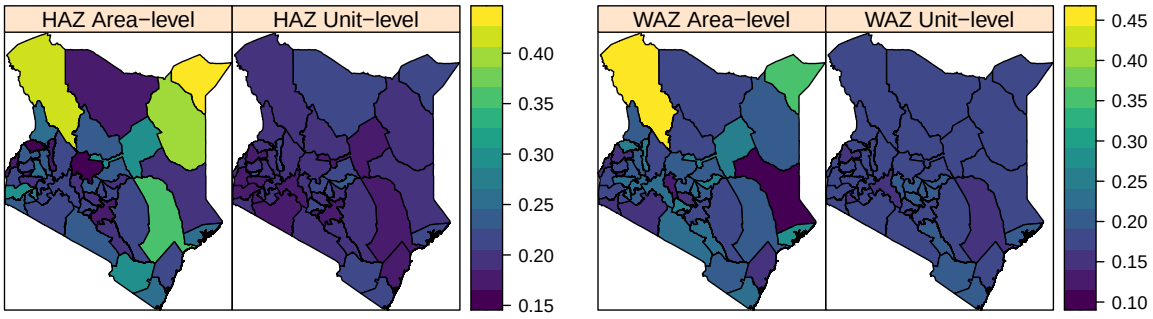


Figure 5.3: Estimated posterior 80% credible interval width from area-level Bivariate shared BYM model and unit-level BYM shared model fit to HAZ and WAZ from the 2014 KDHS.

due to the addition of "phantom" observations in the area-level models, as we see that the regions where these observations were added ended up having somewhat larger differences between the area- and unit-level models. This may lead us to favor unit level models in situations where 0 counts are observed in order to have more consistent estimates in the smallest domains. However, the undercoverage of estimates remains an issue with unit-level models.

The frameworks researched in this dissertation are novel approaches that filled crucial methodological gaps in the literature of modeling health and demographic outcomes in LMICs. However, they are just the first step. The flexible modeling approach in Chapter 2 can be validated and adapted in many other SRS data sources, and can also be used to estimate any count/rate outcomes beyond cause-specific mortality. Figuring out how to best model the different dimensions of the tabulated data (such as cause, age, time, and geography) still needs further exploration. The area- and unit-level SAE models in developed in Chapters 3 and 4 still need to be compared in more detail and in different scenarios. Their utility as the number of outcomes increases also must be demonstrated. Extensions into spatiotemporal models are also an exciting next step due to the constant push for estimates of trends over time, forecasting, and finer spatial disaggregation. Furthermore, some LMICs have access to both SRS and survey data. Developing methods to use both types of data is a very interesting avenue to explore, and the methods covered in this dissertation will be a good starting point.

One of the most important issues to address in future work is the dissemination and uptake of these methods for greater use. One crucial component to this is building capacity for practitioners in LMICs to do their own research, and develop their own estimates. While the underlying sentiment can sometimes be that numbers and/or statistics are objective, this could not be further from the truth. Any statistical modeling or estimation procedure is filled with so many micro-decisions along the way, such as choosing what to model, what data to collect, and what model to use, along with many other issues. Thus, it is imperitive for someone to know the local context of the areas for which they are producing estimates.

This is something that has been especially lacking in global health research to date. Having software available (such as an R package or a Shiny app) that can implement methodology is a good first step, but ultimately the capacity in LMICs needs to be increased—technologically, infrastructurally, educationally—so that methodological development can come from the source. It's time to turn "global health" into "local" health.

Appendix A

APPENDIX FOR CHAPTER 2

A.1 Overview of Sample Registration Systems

Sample registration systems (SRS) are demographic surveillance systems that aim to provide reliable annual estimates of key demographic indicators, notably child mortality, at the national and sub-national levels. The main types of SRS are Health and Demographic Surveillance Systems (HDSS), which cover specific subnational areas, and large national SRS, which cover a nationally representative sample of areas in a country.

HDSS sites are a short- to medium-term measure to provide data on important health and population indicators at regional levels, with possible extrapolation to national levels. They additionally provide lessons that countries can use to set up nationally representative SRS. However, because they do not provide nationally representative information, they should not be seen as a replacement for a nationally representative SRS. For further detail on HDSS sites and examples of the data they provide, see the INDEPTH network (Sankoh & Byass, 2012), the ALPHA network (Maher et al., 2010), and the CHAMPS project (Cunningham et al., 2019).

Large national SRS provide continuous monitoring and reporting. Following a nationally representative sample over time allows ascertainment of population health trends in a relatively timely manner. India and China have both successfully implemented nationally representative SRS (S. Liu et al., 2016; Mahapatra, 2010; Yang et al., 2005). Additionally, China has implemented an SRS focused on maternal and child health—the Maternal and Child Health Surveillance System (MCHSS). Many SRS collect information on all deaths in the covered population and assign causes through health facility death certificates and verbal autopsy surveys. Thus, we can use SRS data to estimate cause-specific mortality rates. This

can be done across age groups, time periods, regions, and other tabulation groups depending on the design of the SRS and availability of data.

A.2 Methods to calculate exposure time in the MCHSS

In order to estimate age- and cause-specific death rates for under-5 in China, we use data from the Chinese Maternal and Child Health Surveillance System (MCHSS). The MCHSS was established in 1996 based on three independent surveillance systems: (1) the Child Mortality Surveillance System that started in 1991; (2) population-based maternal mortality surveillance system; and (3) hospital-based birth defect surveillance system. The database currently available collects information from 1996 to 2015 on births and under-5 deaths by cause, split in six age groups: from 0 to 6 days, from 7 to 27 days, from 1 to 5 months, from 6 to 11 months, from 12 to 23 months, and from 24 to 59 months. The data refers to populations under surveillance from six different strata, corresponding to three regions (East, Mid and West) and two levels of resolution (rural and urban). Additional details on the characteristics of the data can be found in He et al. (2017).

The surveillance contents, case definitions, reporting methods, and quality control are unified across all surveillance sites within MCHSS. However, between 2008 and 2009 there was an expansion of the surveillance sites from 116 (37 urban and 79 rural) to 336 counties/districts (126 urban and 210 rural), which implies a notable increase of the population and number of deaths under surveillance, as observed in Figure A.1. To estimate age-specific death rates one needs deaths and exposures (individuals exposed to the risk of dying). Only births and deaths are reported in the MCHSS, so in order to estimate the exposures for each age category on option is to reconstruct birth cohorts from the available data. This method presents three main challenges:

1. In the MCHSS, data are collected on a yearly scale, but age groups have a finer resolution;

2. Since we are estimating under-5 mortality, data from the previous 5 years are necessary to approximate the exposures of all age category on a given year; and

3. Because of the surveillance expansion, data from 2008 or earlier cannot be used to

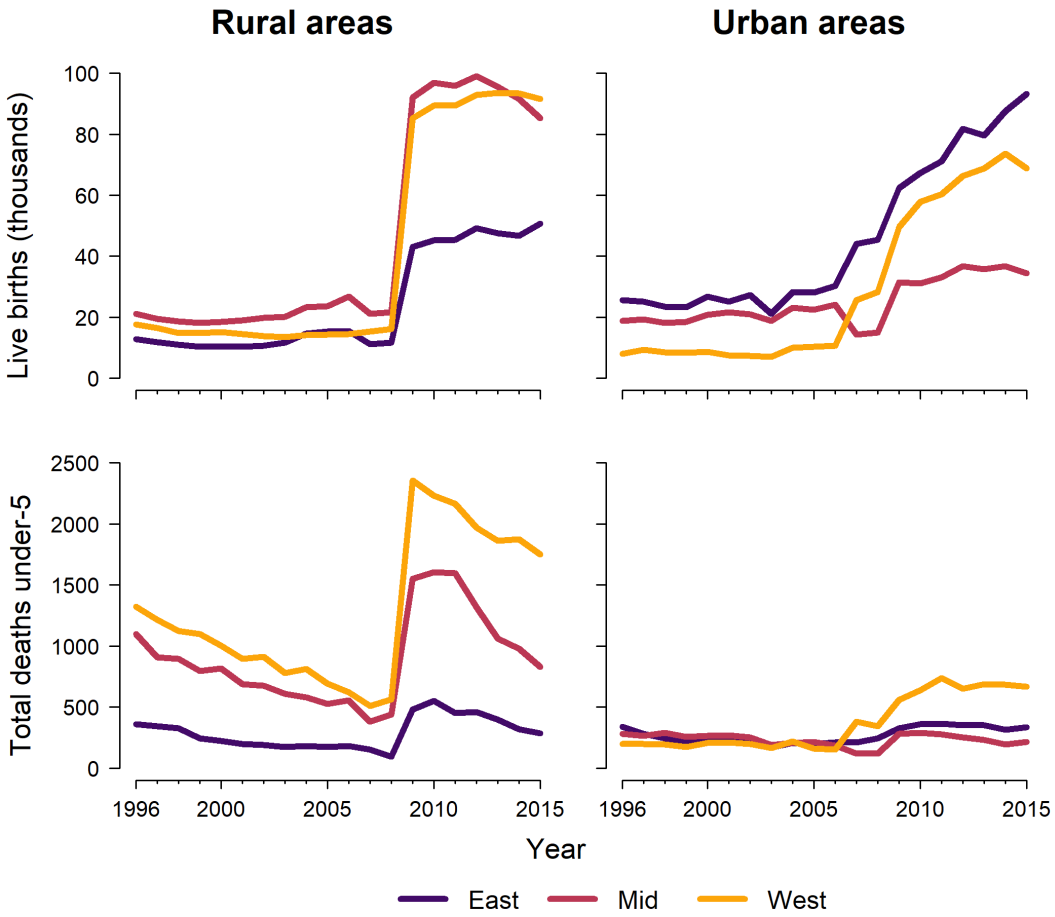


Figure A.1: Total number of births and deaths registered in the MCHSS Chinese data for each fo the six strata, 1996–2015

produce mortality estimates from 2009 onward.

As a result, given the characteristics of the MCHSS, with the "birth-cohort reconstruction method" it is only possible to obtain mortality estimates for years 2001 to 2008, 2014 and 2015 (i.e., missing the first 5 years and the 5 years after the surveillance expansion). Because we wanted to have mortality estimates for as many years as possible to analyze time trends, we opted to build period life tables for 1996–2016 using, of each year, only data for that year. Period life tables are perhaps the most important tool in population studies, and show "what

would happen to a cohort *if* it was subjected for all its life to the mortality conditions of that period" (Preston, Heuveline, & Guillot, 2001). In other words, it summarizes the mortality regime that is inherent in the current pattern of death rates in a population, uninfluenced by fluctuations over time in mortality, fertility or migration.

Life tables are generally calculated based on observed age-specific death rates, computed as the quotient between the observed number of deaths and the mid-year population for a given age and time period. Unfortunately though, the mid-year population for each of the age groups mentioned above are not available in the MCHSS. Only the number of births can be used as *exposures*. Hence, for each year and strata we calculated the different columns of the life table as follows:

1. Use the number of births B_0 as the starting population;

2. Let ${}_{6d}D_0$ be the total number of deaths in the first age group (0 to 6 days), we calculate

$$B_{7d} = B_0 - {}_{6d}D_0 \qquad \text{and} \qquad {}_{6d}q_0 = \frac{{}_{6d}D_0}{B_0} \,,$$

   the number of individuals who reach to the second week, and the probability of dying in the week of life, respectively.

3. Replicating the procedure for older age groups, we get the probabilities of dying between one age group to the following, and obtain the ${}_nq_x$ column of the life table.

Once the death probabilities ${}_nq_x$ are available, there are several existing strategies to estimate age-specific death rates ${}_nm_x$. Perhaps the most common is to use a set ${}_na_x$ values (average person years lived in the interval by those dying in the interval) to estimate the number of person-years lived in the population between ages x and $x + n$ (${}_nL_x$ column of the life table, described in Preston et al. (2001). Here, we assumed that age-specific death rates are constant within each age interval, and estimate them as

$$ {}_nm_x = -\log(1 - {}_nq_x) \,/\, n \ , $$

where n is the length of the corresponding age interval in a yearly scale (Preston et al., 2001). We further carried out a sensitivity analysis to verify that the estimated death rates are not affected by the conversion method from death probabilities. We tested several sets of ${}_na_x$ values—some empirical and other theoretical—and no remarkable differences were identified.

Once the age-specific death rates ${}_nm_x$ and age-specific probabilities of death ${}_nq_x$ are available, it is straightforward to compute the age- and cause-specific rates. Let ${}_nD_x$ denote the total number of deaths between ages x and $x+n$, and ${}_nD_x^i$ the number of deaths by cause i between ages x and $x+1$. The corresponding probabilities of death and death rates are, respectively,

$$ {}_nq_x^i = {}_nq_x \cdot \frac{{}_nD_x^i}{{}_nD_x} \qquad \text{and} \qquad {}_nm_x^i = {}_nm_x \cdot \frac{{}_nq_x^i}{{}_nq_x} \ , $$

as described by Preston et al. (2001).

Lastly, let ${}_nL_x$ be the total person-years lived between ages x and $x+n$, which is the exposure time of interest. We can also write death rates as ${}_nm_x = \frac{{}_nD_x}{{}_nL_x}$ [Preston et al. (2001). Therefore, we calculate the final estimates of exposure time as

$$ {}_nL_x = \frac{{}_nD_x}{{}_nm_x}. $$

A.3 Summary statistics for MCHSS data

	Deaths	Person-years	log mortality
region			
East rural	6111	15955	-7.87
Mid rural	18315	29089	-7.37
West rural	26162	25520	-6.88
East urban	5773	30869	-8.58
Mid urban	5023	17593	-8.16
West urban	7766	19611	-7.83
age			
0-6d	29462	659	-3.11
1-5m	12167	14796	-7.10
12-23m	5784	27356	-8.46
24-59m	7972	79533	-9.21
6-11m	5207	13999	-7.9
7-27d	8558	2295	-5.59
cause			
acute resp. infections	10516	22367	-7.66
birth asphyxia/trauma	10339	2219	-5.37
congenital anomalies	11437	22367	-7.58
diarrhea	2872	22367	-8.96
injuries	10135	22367	-7.70
other group 1	7120	22367	-8.05
other non-communicable	5150	22367	-8.38
prematurity	11581	2219	-5.26
time period			
1996–2000	15993	16200	-6.92
2001–2005	11742	17394	-7.30
2006–2010	17205	37387	-7.68
2011–2015	24210	67656	-7.94

Table A.1: Tabulated deaths, person years, and empirical log mortality in the MCHSS.

A.4 Full derivation of the likelihood for the MCHSS data

Let $i \in \{1, \ldots, n\}$ index individuals and let $c \in \{1, \ldots, C\}$ index causes of death. Let T be a continuous random variable representing survival time and let J be a random variable representing cause of death. We will be parameterizing survival time as age, so it will be time from birth for each individual. Let $\boldsymbol{z}$ be the value of a fixed covariate vector. We will proceed by assuming that this covariate vector is fixed, but this work extends to time-varying covariates in which case we must adjust calculation of the survivor function (defined below) to account for the values of covariates through time (the trajectory of the covariates).

We define the overall hazard rate as

$$\lambda(t|\boldsymbol{z}) = \lim_{\Delta t \to 0} P\big(t \leq T < t + \Delta t | T \geq t, \boldsymbol{z}\big)/\Delta t$$

and define cause-specific failure rates as

$$\lambda_c(t|\boldsymbol{z}) = \lim_{\Delta t \to 0} P\big(t \leq T < t + \Delta t, J = c | T \geq t, \boldsymbol{z}\big)/\Delta t.$$

Assuming mutually exclusive and collectively exhaustive causes of death, we can express the overall hazard rate in terms of the cause-specific hazard rates as

$$\lambda(t|\boldsymbol{z}) = \sum_{c=1}^{C} \lambda_c(t|\boldsymbol{z}).$$

We will define the overall survivor function as

$$S(t|\boldsymbol{z}) = \exp\left(-\Lambda(t|\boldsymbol{z})\right)$$

where $\Lambda(t|\boldsymbol{z}) = \int_0^t \lambda(u|\boldsymbol{z})du$ is the cumulative risk by integrating the overall hazard. The function $S(t|\boldsymbol{z})$ gives the probability of surviving all causes of death until time t.

Similarly, we define the cause-specific analogy

$$S_c(t|\boldsymbol{z}) = \exp\left(-\Lambda_c(t|\boldsymbol{z})\right)$$

where $\Lambda_c(t|\boldsymbol{z}) = \int_0^t \lambda_c(u|\boldsymbol{z})du$ is the integrated or cumulative hazard for cause c. We will note that in general $S_c(t|\boldsymbol{z})$ does not have the interpretation of a survivor function when $C > 1$ due to the competing risks. Since $\Lambda(t|\boldsymbol{z}) = \sum_{c=1}^{C} \Lambda_c(t|\boldsymbol{z})$, then if $\lambda_c(t|\boldsymbol{z}) \neq 0$ for two or more causes, then

$$\begin{aligned} S(t|\boldsymbol{z}) &= \exp[-\Lambda(t|\boldsymbol{z})] \\ &< \exp[-\Lambda_c(t|\boldsymbol{z})] \\ &= S_c(t|\boldsymbol{z}) \end{aligned}$$

with strict inequality since at least two causes have nonzero hazards. Then, since by definition of a survivor function $\lim_{t\to\infty} S(t|\boldsymbol{z}) = 0$, we have $\lim_{t\to\infty} S_c(t|\boldsymbol{z}) > 0$ and thus $S_c(t|\boldsymbol{z})$ are not survivor functions.

Note the relationship that

$$\begin{aligned} S(t|\boldsymbol{z}) &= \exp\left(-\Lambda(t|\boldsymbol{z})\right) \\ &= \exp\left(-\int_0^t \lambda(u|\boldsymbol{z})du\right) \\ &= \exp\left(-\int_0^t \sum_{c=1}^{C} \lambda_c(u|\boldsymbol{z})du\right) \\ &= \exp\left(-\sum_{c=1}^{C} \int_0^t \lambda_c(u|\boldsymbol{z})du\right) \\ &= \prod_{c=1}^{C} \exp\left(-\int_0^t \lambda_c(u|\boldsymbol{z})du\right) \\ &= \prod_{c=1}^{C} \exp\left(-\Lambda_c(t|\boldsymbol{z})du\right) \\ &= \prod_{c=1}^{C} S_c(t|\boldsymbol{z}). \end{aligned}$$

Define the probability function for time to death and cause of death, known as the cause-specific densities, as

$$\begin{aligned} f_c(t|\boldsymbol{z}) &= \lim_{\Delta t \to 0} P\big(t \leq T < t + \Delta t, J = c|\boldsymbol{z}\big)/\Delta t \\ &= \lambda_j(t|\boldsymbol{z}) S(t|\boldsymbol{z}) \end{aligned}$$

which represents the unconditional risk of dying at time t of cause j. This can be interpreted as the probability of surviving up until time t ($S(t|\boldsymbol{z})$) multiplied by the instantaneous risk of dying from cause c. By the law of total probability, we have

$$f(t|\boldsymbol{z}) = \sum_{c=1}^{C} f_c(t|\boldsymbol{z}).$$

We also define the cumulative incidence function as

$$I_c(t|\boldsymbol{z}) = \int_0^t f_c(u|\boldsymbol{z}) du$$

which represents the probability that a death from cause c has occurred by time t.

Following Prentice et al. (1978), suppose we have data $(t_i, c_i, \delta_i, \boldsymbol{z}_i)$ from n subjects, where t_i is the time of observation, c_i is the cause of death, δ_i is a death indicator with $\delta_i = 1$ if a deatg is observed and $\delta_i = 0$ otherwise (censored), and $\boldsymbol{z}_i$ is a vector of covariates for subject i. Up to proportionality, the likelihood function under an independent censoring mechanism (i.e. censoring of observations is independent of survival time given the covariates) can be written as

$$\begin{aligned} \mathcal{L} &= \prod_{i=1}^{n} \lambda_{c_i}(t_i|\boldsymbol{z}_i)^{\delta_i} S(t_i|\boldsymbol{z}_i) \\ &= \prod_{i=1}^{n} \lambda_{c_i}(t_i|\boldsymbol{z}_i)^{\delta_i} \prod_{c=1}^{C} S_c(t_i|\boldsymbol{z}_i) \\ &= \prod_{i=1}^{n} \lambda_{c_i}(t_i|\boldsymbol{z}_i)^{\delta_i} \prod_{c=1}^{C} \exp\left(-\Lambda_j(t_i|\boldsymbol{z}_i)\right) \\ &= \prod_{i=1}^{n} \left[\lambda_{c_i}(t_i|\boldsymbol{z}_i)^{\delta_i} \prod_{c=1}^{C} \exp\left(- \int_0^{t_i} \lambda_c\left(u|\boldsymbol{z}_i\right) du \right) \right]. \end{aligned}$$

Here, an individual who is censored at time t_i contributes to the likelihood the probability of surviving up until this time,

$$S(t_i|\boldsymbol{z}_i),$$

and an individual who is observed to die from cause c_i at time t_i contributes to the likelihood the probability of surviving up until time t_i, $S(t_i|\boldsymbol{z}_i)$, multiplied by the instantaneous hazard of death from cause c_i at time t_i, $\lambda_{c_i}(t_i|\boldsymbol{z}_i)$,

$$\lambda_{c_i}(t_i|\boldsymbol{z}_i)S(t_i|\boldsymbol{z}_i),$$

which is the same as the cause-specific density.

Note that additionally,

$$\mathcal{L} = \prod_{i=1}^{n} f_{c_i}(t_i|\boldsymbol{z}_i).$$

Now, let's redefine the data such that $d_{ic} = 1$ if individual i dies from cause c and $d_{ic} = 0$ otherwise. Because each individual can only die from a single cause, we have $\delta_i = \sum_{c=1}^{C} d_{ic} = 1$. We can then rewrite the likelihood as

$$\mathcal{L} = \prod_{i=1}^{n}\prod_{c=1}^{C}\left[\lambda_c(t_i|\boldsymbol{z}_i)^{d_{ic}} \exp\left(-\int_0^{t_i} \lambda_c\left(u|\boldsymbol{z}_i\right) du\right)\right].$$

We observe two important facts:

1. The overall likelihood function is a product of C likelihoods, one for each cause of death.
2. The likelihood involving a specific cause of death is exactly the same likelihood you would obtain by treating all other causes of death as censored observations.

Additionally note that we can also use different covariates for each cause, which means we can define $\boldsymbol{z}_{ic}$ as the covariate vector for individual i and cause c, and the resulting likelihood is

$$\mathcal{L} = \prod_{i=1}^{n} \prod_{c=1}^{C} \left[\lambda_c(t_i|\boldsymbol{z}_{ic})^{d_{ic}} \exp\left(- \int_0^{t_i} \lambda_c\left(u|\boldsymbol{z}_{ic}\right) du \right) \right].$$

Next, suppose that rather than individual-level data, we instead have data that is tabulated into death counts and total person-time experienced across all people in K age groups; again, observation time is parameterized as the age of each individual rather than calendar time. We will follow the standard argument in Holford (1976), Holford (1980), and Laird & Olivier (1981), and define K intervals with breakpoints $0 = \tau_0 < \tau_1 < \cdots < \tau_K (= \infty)$.

Define $d_{ikc} = 1$ if individual i dies of cause c in age group k and define t_{ik} as the total time that individual i is observed in age group k (i.e. $t_{ik} = \min(t_i - \tau_{k-1}, \tau_k - \tau_{k-1})$. We note that $d_{ic} = \sum_{k=1}^{K} d_{ikc}$ because a person can only die in one age group, and $t_i = \sum_{k=1}^{K} t_{ik}$. With this newly parameterized data, we can rewrite the individual-level likelihood as

$$\begin{aligned}
\mathcal{L} &= \prod_{i=1}^{n} \prod_{c=1}^{C} \left[\lambda_c(t_i|\boldsymbol{z}_{ic})^{d_{ic}} \exp\left(- \int_0^{t_i} \lambda_c\left(u|\boldsymbol{z}_{ic}\right) du \right) \right] \\
&= \prod_{i=1}^{n} \prod_{c=1}^{C} \left[\lambda_c(t_i|\boldsymbol{z}_{ic})^{d_{ic}} \exp\left(- \sum_{k=1}^{K} \int_{\tau_{k-1}}^{\tau_{k-1}+t_{ik}} \lambda_c\left(u|\boldsymbol{z}_{ic}\right) du \right) \right] \\
&= \prod_{i=1}^{n} \prod_{c=1}^{C} \left[\lambda_c(t_i|\boldsymbol{z}_{ic})^{d_{ic}} \prod_{k=1}^{K} \exp\left(- \int_{\tau_{k-1}}^{\tau_{k-1}+t_{ik}} \lambda_c\left(u|\boldsymbol{z}_{ic}\right) du \right) \right] \\
&= \prod_{i=1}^{n} \prod_{c=1}^{C} \left[\prod_{k=1}^{K} \lambda_c(t_i|\boldsymbol{z}_{ic})^{d_{ikc}} \prod_{k=1}^{K} \exp\left(- \int_{\tau_{k-1}}^{\tau_{k-1}+t_{ik}} \lambda_c\left(u|\boldsymbol{z}_{ic}\right) du \right) \right] \\
&= \prod_{i=1}^{n} \prod_{k=1}^{K} \prod_{c=1}^{C} \left[\lambda_c(t_i|\boldsymbol{z}_{ic})^{d_{ikc}} \exp\left(- \int_{\tau_{k-1}}^{\tau_{k-1}+t_{ik}} \lambda_c\left(u|\boldsymbol{z}_{ic}\right) du \right) \right].
\end{aligned}$$

The second quality holds because $\int_{\tau_{k-1}}^{\tau_{k-1}+0} \lambda_c\left(u|\boldsymbol{z}_{ic}\right) du = 0$ which is the case for all age groups subsequent to the final age group of observation for individual i. The second to last equality holds because $\lambda_c(t_i|\boldsymbol{z}_{ic})^{d_{ikc}} = 1$ for all age groups k in which no death was observed.

Now, define the observed data as $y_{kc} = \sum_{i=1}^{n} d_{ikc}$, the number of deaths in age group k from cause c, and $t_k = \sum_{i=1}^{N} t_{ik}$, the total person-time observed in age group k.

We will assume that each age group has constant cause-specific hazards. Additionally for simplicity, we will proceed assuming no covariates such that $\lambda_c(t|\boldsymbol{z}_{ic}) = \lambda_c(t)$; however, we can allow for $\lambda_c(t)$ to depend on covariates and the following derivation still holds. We will account for covariates later in defining a functional form for the constant cause-specific hazards.

With these assumptions, we have

$$\lambda_c(t_i) = \lambda_{kc}, \text{ for } t_i \in [\tau_{k-1}, \tau_k).$$

Now, we can write the tabulated likelihood in terms of the observed data as

$$\begin{aligned}
\mathcal{L} &= \prod_{i=1}^{n}\prod_{k=1}^{K}\prod_{c=1}^{C}\left[\lambda_c(t_i|\boldsymbol{z}_{ic})^{d_{ikc}} \exp\left(-\int_{\tau_{k-1}}^{\tau_{k-1}+t_{ik}} \lambda_c\left(u|\boldsymbol{z}_{ic}\right) du\right)\right] \\
&= \prod_{i=1}^{n}\prod_{k=1}^{K}\prod_{c=1}^{C}\left[\left(\lambda_{kc}\right)^{d_{ikc}} \exp\left(-\int_{\tau_{k-1}}^{\tau_{k-1}+t_{ik}} \left(\lambda_{kc}\right) du\right)\right] \\
&= \prod_{k=1}^{K}\prod_{c=1}^{C}\left[\left(\lambda_{kc}\right)^{\sum_{i=1}^{N} d_{ikc}} \exp\left(-\sum_{i=1}^{N}\int_{\tau_{k-1}}^{\tau_{k-1}+t_{ik}} \left(\lambda_{kc}\right) du\right)\right] \\
&= \prod_{k=1}^{K}\prod_{c=1}^{C}\left[\left(\lambda_{kc}\right)^{y_{kc}} \exp\left(-\sum_{i=1}^{N}\left(t_{ik}\lambda_{kc}\right)\right)\right] \\
&= \prod_{k=1}^{K}\prod_{c=1}^{C}\left[\left(\lambda_{kc}\right)^{y_{kc}} \exp\left(-\lambda_{kc}t_k\right)\right]
\end{aligned}$$

We see that this is the kernel of the likelihood that would arise if

$$y_{kc}|\lambda_{kc}, t_k \sim \text{Poisson}(\lambda_{kc}t_k),$$

and thus we can make likelihood-based inference using separate Poisson distributions for each cause and age group.

Now, suppose that the data we observe is tabulated into both K age groups and H additional subgroups indexed by $h = 1, \ldots, H$. We will first let these be time periods. Notably, an individual may have observation time spent across multiple strata, e.g. observed across different time periods.

First, we will again extend the indexing of the individual-level data. Define $d_{ihkc} = 1$ if individual i dies of cause c in age group k and time period h, and define t_{ihk} as the total time that individual i is observed in age group k and time period h. We note that $d_{ic} = \sum_{k=1}^{K} \sum_{h=1}^{H} d_{ihkc}$ because a person can only die in one age group and one time period, and $t_i = \sum_{k=1}^{K} \sum_{h=1}^{H} t_{ihk}$. Thus we can rewrite the individual-level likelihood, following the same arguments as the previous section, as

$$\mathcal{L} = \prod_{i=1}^{n} \prod_{h=1}^{H} \prod_{k=1}^{K} \prod_{c=1}^{C} \left[\lambda_c(t_i|\boldsymbol{z}_{ic})^{d_{ihkc}} \exp\left(- \int_{\tau_{k-1}}^{\tau_{k-1}+t_{ihk}} \lambda_c\left(u|\boldsymbol{z}_{ic}\right) du \right) \right].$$

Now, define the observed data as $y_{hkc} = \sum_{i=1}^{n} d_{ihkc}$, the number of deaths in age group k and time period h from cause c, and $t_{hk} = \sum_{i=1}^{N} t_{ihk}$, the total person-time observed from individual in age group k and time period h.

We will now assume that each age-period tabulation group has constant cause-specific hazards, and again we will proceed assuming no covariates for simplicity. Therefore, we have

$$\lambda_c(t_i) = \lambda_{hkc}, \text{ for } t_i \in [\tau_{k-1}, \tau_k) \text{ and individual } i \text{ observed in time period } h.$$

Now, we can write the tabulated likelihood in terms of the observed data as

$$\begin{aligned}
\mathcal{L} &= \prod_{i=1}^{n} \prod_{h=1}^{H} \prod_{k=1}^{K} \prod_{c=1}^{C} \left[\lambda_c(t_i|\boldsymbol{z}_{ic})^{d_{ihkc}} \exp\left(- \int_{\tau_{k-1}}^{\tau_{k-1}+t_{ihk}} \lambda_c\left(u|\boldsymbol{z}_{ic}\right) du \right) \right] \\
&= \prod_{h=1}^{H} \prod_{k=1}^{K} \prod_{c=1}^{C} \left[(\lambda_{hkc})^{\sum_{i=1}^{N} d_{ikc}} \exp\left(- \sum_{i=1}^{N} \left(t_{ihk}\lambda_{hkc}\right) du \right) \right] \\
&= \prod_{h=1}^{H} \prod_{k=1}^{K} \prod_{c=1}^{C} \left[(\lambda_{hkc})^{y_{hkc}} \exp\left(-\lambda_{hkc} t_{hk}\right) \right]
\end{aligned}$$

We see that this is the kernel of the likelihood that would arise if

$$y_{hkc}|\lambda_{hkc}, t_{hk} \sim \text{Poisson}(\lambda_{hkc}t_{hk}),$$

and thus we can again make likelihood-based inference using separate Poisson distributions for each cause and age group.

This argument extends to any additional tabulations. For example, if we have data further tabulated into regions $r = 1, \ldots, R$ and we assume constant cause-specific hazards in each age-period-region tabulation group ($\lambda_c(t_i|\boldsymbol{z}_{ic}) = \lambda_{rhkc}$ for individual i while in region r, period h, and age group k), then the kernel of the likelihood becomes the same as if

$$y_{rhkc}|\lambda_{rhkc}, t_{rhk} \sim \text{Poisson}(\lambda_{rhkc}t_{rhk}),$$

with y_{rhkc} the number of deaths from cause c and t_{rhk} the total person-time observed in region r, period h, and age group k.

A.5 Full description and results from a simulation study comparing a unified model to a multistage model with a Multinomial logistic regression to estimate cause fractions

A.5.1 Data generating mechanism

Let $i \in \{1, \ldots, n = 720\}$ index strata (i.e. 6 age groups, 6 regions, 20 years) and let $c \in \{1, \ldots, 8\}$ index cause, i.e. only 3 causes. Let y_{ic} denote the death counts in strata i from cause c, with N_i the total exposure time in strata i. We generate the cause-specific death counts as follows:

$$\begin{aligned} y_{ic} &\sim \text{Poisson}\left(N_i \lambda_{ic}\right) \\ \log(\lambda_{ic}) &= \theta_{ic} \\ &= \alpha + \beta_c + \epsilon_{ic} \\ \epsilon_{ic} &\sim \text{Normal}(0, \sigma_\epsilon^2) \end{aligned}$$

For identifiability we set the coefficient for the reference cause $\beta_1 = 0$.

Aditionally, let $y_{i+} = \sum_c y_{ic}$ be the all-cause death counts, let $\lambda_{i+} = \sum_c \lambda_{ic}$ be the all-cause mortality rate, let $\theta_{i+} = \sum_c \theta_{ic}$ be the sum of the log mortality rates, and let $\tilde{\theta}_{i+} = \log(\lambda_{i+})$, the log all cause mortality rate.

We simulate 100 datasets for each of three different magnitudes of exposure time, $N_i \in \{1000, 10000, 100000\}$, noting that the median exposure time in the MCHSS data is approximately 9000. For all simulations, we set $\alpha = -4$, $\beta_1 = \beta_2 = 0.5$, and $\sigma_\epsilon^2 = 0.1$.

A.5.2 Multistage model

For stage 1, we fit a Poisson GLM on the all-cause mortality rate such that

$$\begin{aligned} y_{i+} &\sim \text{Poisson}\left(N_i \hat{\lambda}_{i+}\right) \\ \log(\hat{\lambda}_{i+}) &= \hat{\tilde{\theta}}_{i+} \\ &= \hat{\alpha} + \hat{\gamma}_i \\ \hat{\gamma}_i &\sim \text{Normal}(0, \hat{\sigma}^2_\gamma) \end{aligned}$$

We use the default diffuse priors in INLA to fit this model. We then take 1000 draws from the posterior predicted mortality rates λ_{i+}.

For the second model, we fit a Multinomial logistic regression. We specify this regression in INLA using the "Poisson Trick" (Lee et al., 2017), which means we fit the model

$$\begin{aligned} y_{ic} &\sim \text{Poisson}\left(N_i \hat{\lambda}_{ic}\right) \\ \log(\hat{\lambda}_{ic}) &= \hat{\theta}_{ic} \\ &= \hat{\alpha} + \hat{\beta}_c + \hat{\phi}_c + \hat{\xi}_{ic} \\ \hat{\xi}_{ic} &\sim \text{Normal}(0, \hat{\sigma}^2_\xi) \\ \hat{\phi}_c &\sim \text{Normal}(0, 1000), \end{aligned}$$

with $\hat{\beta}_1 = 0$ and a sum-to-zero constraint on the ϕ_c. We then take 1000 draws from the posterior distribution of the CSMFs, $\frac{\hat{\theta}_{ic}}{\hat{\theta}_{i+}}$.

To combine the two stages, we pair each of the 1000 all-cause mortality draws with a draw of CSMFs and multiply them and take the log to estimate the log CSMR $\hat{\lambda}_{ic}$.

A.5.3 Unified model

For comparison, we fit a correctly-specified Poisson regression fit in INLA with the default prior distributions.

A.5.4 Comparison between multistage and unified models

We calculate the relative bias, coverage of nominal 95% credible intervals, and width of these intervals averaged over all of the estimated $\hat{\lambda}_{ic}$. Results are presented in Figure A.2. We see

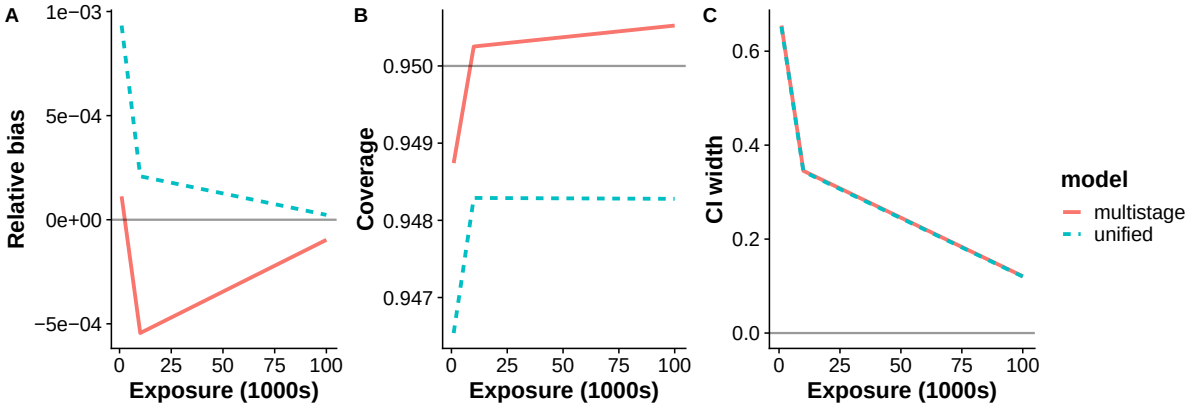


Figure A.2: Comparison of relative bias, coverage of 95% credible intervals, and width of 95% credible intervals for predicted log mortality rates between multistage (Poisson-Multinomial) and unified modeling approaches. Data was generated with fixed intercepts for each cause along with IID Normal random effects for each observation.

comparable results from both methods.

A.6 Full description of the simulation study comparing a multistage model and a unified model in a scenario with observation-level IID Normal random effects

A.6.1 Data generating mechanism

Let $i \in \{1, \ldots, n = 720\}$ index strata (i.e. 6 age groups, 6 regions, 20 years) and let $c \in \{1, \ldots, C\}$ index cause. Let y_{ic} denote the death counts in strata i from cause c, with N_i the total exposure time in strata i. We generate the cause-specific death counts as follows:

$$\begin{aligned} y_{ic} &\sim \text{Poisson}\left(N_i \lambda_{ic}\right) \\ \log(\lambda_{ic}) = \theta_{ic} & \\ &= \alpha + \sum_{j=2}^{C} \beta_j \mathbb{1}_{[j=c]} + \epsilon_{ic} \\ \epsilon_{ic} &\sim \text{Normal}(0, \sigma_\epsilon^2) \end{aligned}$$

Aditionally, let $y_{i+} = \sum_c y_{ic}$ be the all-cause death counts, let $\lambda_{i+} = \sum_c \lambda_{ic}$ be the all-cause mortality rate, let $\theta_{i+} = \sum_c \theta_{ic}$ be the sum of the log mortality rates, and let $\tilde{\theta}_{i+} = \log(\lambda_{i+})$, the log all cause mortality rate.

A.6.2 Multistage model

Stage 1: all-cause mortality model

To estimate all-cause mortality rates, we use a Poisson GLMM with an overall intercept and an IID Normal random effect on each strata. The model is

$$\begin{aligned} y_{i+} &\sim \text{Poisson}(N_i \hat{\lambda}_{i+}) \\ \log(\hat{\lambda}_{i+}) = \widehat{\tilde{\theta}}_{i+} & \\ &= \widehat{\tilde{\alpha}}_{+} + \hat{\gamma}_i \\ \hat{\gamma}_i &\sim \text{Normal}(0, \hat{\sigma}_\gamma^2). \end{aligned}$$

We use the default INLA priors. We then sample from the posterior for each $\hat{\lambda}_i$ to get 1000 draws of each.

Stage 2: cause-specific mortality fraction model

To estimate the CSMFs, we use separate Poisson GLMMs with an overall intercept and an IID Normal random effect on each strata. For each of the C causes, the model is

$$\begin{aligned} y_{ic} &\sim \text{Poisson}(N_i \hat{\lambda}_{ic}) \\ \log(\hat{\lambda}_{ic}) &= \hat{\theta}_{ic} \\ &= \hat{\alpha}_c + \hat{\xi}_{ic} \\ \hat{\xi}_{ic} &\sim \text{Normal}(0, \hat{\sigma}^2_{\xi_c}). \end{aligned}$$

We use the default INLA priors. We take 1000 samples from the posterior distribution of the $\hat{\lambda}_{ic}$ from each regression model. We then calculate 1000 draws of the CSMFs, which we calculate as

$$\begin{aligned} \hat{p}_{ic} &= \frac{\hat{\lambda}_{ic}}{\sum_c \hat{\lambda}_{ic}} \\ &= \frac{\exp(\hat{\theta} ic)}{\sum_c \exp(\hat{\theta}_{ic})} \end{aligned}$$

with one draw of CSMR from each cause-specific model resulting in 1000 samples of CSMFs.

Stage 3: combining stages 1 and 2

Lastly, we pair each of the 1000 draws of all-cause mortality with one of the 1000 draws of a set of CSMFs and thus calculate log CSMRs as

$$\begin{aligned} \hat{\theta}_{ic} &= \log(\hat{\lambda}_{i+} \times \hat{p}_{ic}) \\ &= \log\left(\exp(\hat{\tilde{\theta}}_{i+}) \times \frac{\exp(\hat{\theta} ic)}{\sum_c \exp(\hat{\theta}_{ic})}\right) \end{aligned}$$

We then calculate the average relative bias, average coverage of 95% credible interval, and average width of these intervals for all $\hat{\theta}_{ic}$ estimates, which we average over all 100 simulated data sets.

A.6.3 Unified model

We fit a model with the actual data generating mechanism correctly specified. We use the default priors in INLA.

A.7 Explanation of the nonequivalence between a unified model and a multistage model with cause-specific regressions to estimate cause fractions

A simple version of a multistage modeling approach uses three stages: (1) all cause mortality rates are estimated; (2) cause-specific mortality fractions (CSMFs) are estimated; (3) all-cause mortality rates are multiplied by CSMFs to calculate cause-specific mortality rates. This section will show that the case where the all-cause mortality in stage 1 is estimated with a Poisson likelihood and the CSMFs in stage 2 are estimated by separately modeling each cause with a Poisson likelihood is not consistent with jointly estimating cause-specific mortality rates with Poisson likelihoods in the presence of IID Normal random effects.

Let y_{ic} and λ_{ic} be the death count and mortality rate, respectively, for strata $i \in \{1, \ldots, n\}$ and cause $c \in \{1, \ldots, C\}$. Let N_i be the exposure time for strata i. We will specify the data generating mechanism is as follows:

$$\begin{aligned} y_{ic} &\sim \text{Poisson}(N_i \lambda_{ic}) \\ \log(\lambda_{ic}) &= \theta_{ic} \\ &= \alpha_c + \epsilon_{ic} \\ \epsilon_{ic} &\sim \text{Normal}(0, \sigma^2). \end{aligned}$$

For simplicity we are defining cause-specific intercepts α_c, however, the conclusions from this section are also true for the "corner-point" specification in Section A.6. This data generating mechanism reflects a simple scenario in which each cause has its own intercept, and each strata-cause combination has it's own random Normal perterbation. Even in this simple scenario, the multistage model was shown to have poor coverage as compared to the unified model. To show why, we will first note the following:

$$\begin{aligned} \theta_{ic} &\sim \text{Normal}(\alpha_c, \sigma^2) \\ \lambda_{ic} &\sim \text{Lognormal}(\alpha_c, \sigma^2) \end{aligned}$$

We will define $y_{i+} = \sum_j y_{ic}$, $\theta_{i+} = \sum_c \theta_{ic}$, $\lambda_{i+} = \sum_j \lambda_{ic}$, and $\alpha_+ = \sum_c \alpha_c$. Then λ_{i+} is the sum of Lognormals which has no closed form distribution, and $\theta_{i+} \sim \text{Normal}(\alpha_+, m\sigma^2)$.

Then, we will define $\tilde{\lambda}_{i+} = \exp(\theta_{i+}) \sim \text{Lognormal}(\alpha_+, m\sigma^2)$. We then calculate the expectations and variances of λ_{i+} and $\tilde{\lambda}_{i+}$, using properties of Normal and Lognormal distributions, as

$$\begin{aligned}
\mathbb{E}[\lambda_{i+}] &= \exp(\sigma^2/2) \sum_c \exp(\alpha_c) \\
\mathbb{E}[\tilde{\lambda}_{i+}] &= \exp(m\sigma^2/2) \exp(\alpha_+) \\
\text{Var}[\lambda_{i+}] &= \left[\exp(\sigma^2) - 1)\right] \exp(\sigma^2) \sum_c \exp(2\alpha_c) \\
\text{Var}[\tilde{\lambda}_{i+}] &= \left[\exp(m\sigma^2) - 1\right] \exp(\sigma^2) \exp(2\alpha_+).
\end{aligned}$$

Therefore, we see that $\mathbb{E}[\lambda_{i+}] \neq \mathbb{E}[\tilde{\lambda}_{i+}]$ and $\text{Var}[\lambda_{i+}] \neq \text{Var}[\tilde{\lambda}_{i+}]$. Additionally, define $\tilde{\theta}_{i+} = \log(\lambda_{i+})$ and note that $\tilde{\theta}_{i+} \neq \theta_{i+}$, and the corresponding expectations and variances are not equal either.

Results from our simulation showed that using our unified framework to fit a model that correctly specifies the data generating mechanism leads to unbiased estimates with nominal coverage of 95% credible intervals using noninformative prior distributions. This makes sense because we estimate the log mortality rates using consistent estimators for all parameters. Our simulation further showed that the multistage model is unbiased, but it does not have proper coverage of 95% credible intervals, which we will show analytically. Section A.6 describes the parameters, data generating mechanism, and the multistage model fit. We can calculate the posterior log CSMRs to be

$$\begin{aligned}
\hat{\theta}_{ic} &= \log\left(\exp(\widehat{\tilde{\theta}}_{i+}) \times \frac{\exp(\hat{\theta}_{ic})}{\sum_c \exp(\hat{\theta}_{ic})}\right) \\
&= \log(\exp(\widehat{\tilde{\theta}}_{i+})) + \log(\exp(\hat{\theta}_{ic})) - \log\left(\sum_c \exp(\hat{\theta}_{ic})\right) \\
&= \widehat{\tilde{\theta}}_{i+} + \hat{\theta}_{ic} - \widehat{\tilde{\theta}}^{\star}_{i+}
\end{aligned}$$

Now, since the estimators in both stages of the regression models are unbiased for the , we

can calculate

$$\begin{aligned}
\mathbb{E}\left[\hat{\theta}_{ic}\right] &= \mathbb{E}\left[\widehat{\tilde{\theta}}_{i+} + \hat{\theta}_{ic} - \widehat{\tilde{\theta}}^{\star}_{i+}\right] \\
&= \mathbb{E}\left[\widehat{\tilde{\theta}}_{i+}\right] + \mathbb{E}\left[\hat{\theta}_{ic}\right] - \mathbb{E}\left[\widehat{\tilde{\theta}}^{\star}_{i+}\right] \\
&= \mathbb{E}\left[\hat{\theta}_{ic}\right] \\
&= \theta_{ic}
\end{aligned}$$

and so the CSMR are unbiased. However, we calculate the variance as

$$\begin{aligned}
\mathrm{Var}\left[\hat{\theta}_{ic}\right] &= \mathrm{Var}\left[\widehat{\tilde{\theta}}_{i+} + \hat{\theta}_{ic} - \widehat{\tilde{\theta}}^{\star}_{i+}\right] \\
&= \mathrm{Var}\left[\widehat{\tilde{\theta}}_{i+}\right] + \mathrm{Var}\left[\hat{\theta}_{ic}\right] + \mathrm{Var}\left[\widehat{\tilde{\theta}}^{\star}_{i+}\right] \\
&= \mathrm{Var}\left[\widehat{\tilde{\theta}}_{i+}\right] + \mathrm{Var}\left[\hat{\theta}_{ic}\right] + \mathrm{Var}\left[\log\left(\sum_{c}\exp(\hat{\theta}_{ic})\right)\right]
\end{aligned}$$

This is a complicated combination of the variance of estimators of normals, log normals, and the log of sums of lognormals, and furthermore we have the previously shown results that $\mathrm{Var}[\theta_{i+}] \neq \mathrm{Var}[\tilde{\theta}_{i+}]$. We would want the variance of our estimator to be equal to the true variance σ^2 which is clearly not the case. Thus, the results of our simulation study are consistent with this analytical derivation.

A.8 Full description of the simulation study comparing a multistage model and a unified model in a scenario with correlated cause-specific mortality rates

Let $i \in \{1, \ldots, n = 720\}$ index strata (i.e. 6 age groups, 6 regions, 20 years), let $c \in \{1, 2\}$ index cause, let $a \in \{1, \ldots, 6\}$ index age, and let $r \in \{1, \ldots, 6\}$ index region. Let y_{ic} and λ_{ic} be the death count and mortality rate, respectively, and let N_i be the exposure time for observation i. We will denote $a[i]$ and $r[i]$ as the age and region, respectively, for observation i.

The data generating mechanism is as follows:

$$\begin{aligned} y_{ic} &\sim \text{Poisson}(N_i \lambda_{ic}) \\ \log(\lambda_{ic}) &= \theta_{ic} \\ &= \alpha + \beta^C \mathbb{1}_{[c=2]} + \sum_{j=2}^{6} \beta_j^A \mathbb{1}_{[j=a[i]]} + \sum_{k=2}^{6} \beta_k^R \mathbb{1}_{[k=r[i]]} + b_{ic} \\ \mathbf{b}_i &= \{b_{i1}, b_{i2}\} \\ &\sim \text{Normal}_2(\mathbf{0}, \mathbf{\Sigma}). \end{aligned}$$

In this set up, $\mathbf{\Sigma}$ controls the levels of cause-specific overdispersion (as compared to the Poisson distribution) via its diagonal entries and the levels of correlation between cause-specific mortality rates via its off-diagonal entries. Performing this simulation study for a range of diagonal and off-diagonal values will explore how overdispersion and correlations affect a simplified multi-stage modeling procedure.

For this simulation study, we set the exposure time N_i equal to 8749.465, the median observed in the MCHSS data, for all i. We let $\mathbf{\Sigma} = \big(\sigma \mathcal{I}_2\big)\mathbf{\Omega}\big(\sigma \mathcal{I}_2\big)$, with $\mathcal{I}_2$ the two dimensional identity matrix, and $\mathbf{\Omega}$ a correlation matrix with diagonal entries equal to 1 and off-diagonal entries equal to the correlation ρ. We set $\alpha = -4$, $\beta = 0.5$. We perform simulations for 9 scenarios: each combination of $\sigma \in \{0.01, 0.1, 1\}$ and $\rho \in \{-0.5, 0, 0.5\}$. For each scenario, we simulate 100 datasets. For each simulation, we fit the multistage model described in Section A.6 and also a unified model that correctly specified the data generating mechanism

using the default priors in INLA. Across all simulations, we calculate the average relative bias of the estimated log cause-specific mortality rate compared to the true θ_{ic} and calculate what percent of the 95% confidence intervals actually contain this marginal parameter. We also calculate the width of the confidence intervals.

A.9 Residual analyses for MCHSS data to determine which strata to share random walks

In order to see which dimensions of the MCHSS data we could collapse and fit the same temporal trend via random walks, we fit a suite of generalized linear models to the MCHSS data and plot the residuals as a function of time in order to look for patterns.

Each Poisson GLM we fit had the log person years as the offset, fixed intercepts for each age, region, and cause, all pairwise interactions between age, region, and cause, and a linear and a quadratic time trend. Each model varied which dimensions the linear and quadratic time trends varied over. With the same definitions of indices, data, and parameters as the Section 4 in the main text, we fit a models of the form

$$\begin{aligned}\theta_{r,a,t,c} = \alpha + \beta_r^R + \beta_a^A + \beta_c^C \\ \beta_{a,c}^{AC} + \beta_{r,c}^{RC} + \beta_{a,r}^{AR} + \\ \beta^{T1} t + \beta^{T2} t^2 + \\ g(\boldsymbol{\eta}),\end{aligned}$$

where $g(\boldsymbol{\eta})$ is some set of regression coefficients that include the interaction of the linear and quadratic terms. We fit GLMs with the following 8 interaction terms:

1. $g(\boldsymbol{\eta}) = 0$
2. $g(\boldsymbol{\eta}) = \eta_{at}^{AT}(\mathbb{1}_{[a]} \times t \times t^2)$
3. $g(\boldsymbol{\eta}) = \eta_{rt}^{RT}(\mathbb{1}_{[r]} \times t \times t^2)$
4. $g(\boldsymbol{\eta}) = \eta_{ct}^{CT}(\mathbb{1}_{[c]} \times t \times t^2)$
5. $g(\boldsymbol{\eta}) = \eta_{act}^{ACT}(\mathbb{1}_{[ac]} \times t \times t^2)$
6. $g(\boldsymbol{\eta}) = \eta_{art}^{ART}(\mathbb{1}_{[ar]} \times t \times t^2)$

7. $g(\boldsymbol{\eta}) = \eta_{rct}^{RCT}(\mathbb{1}_{[rc]} \times t \times t^2)$

8. $g(\boldsymbol{\eta}) = \eta_{arct}^{ARCT}(\mathbb{1}_{[arc]} \times t \times t^2)$

We present plots of these residuals in Figure A.3 through Figure A.10. We examined these plots for patterns similar between age groups, cause groups, and regions.

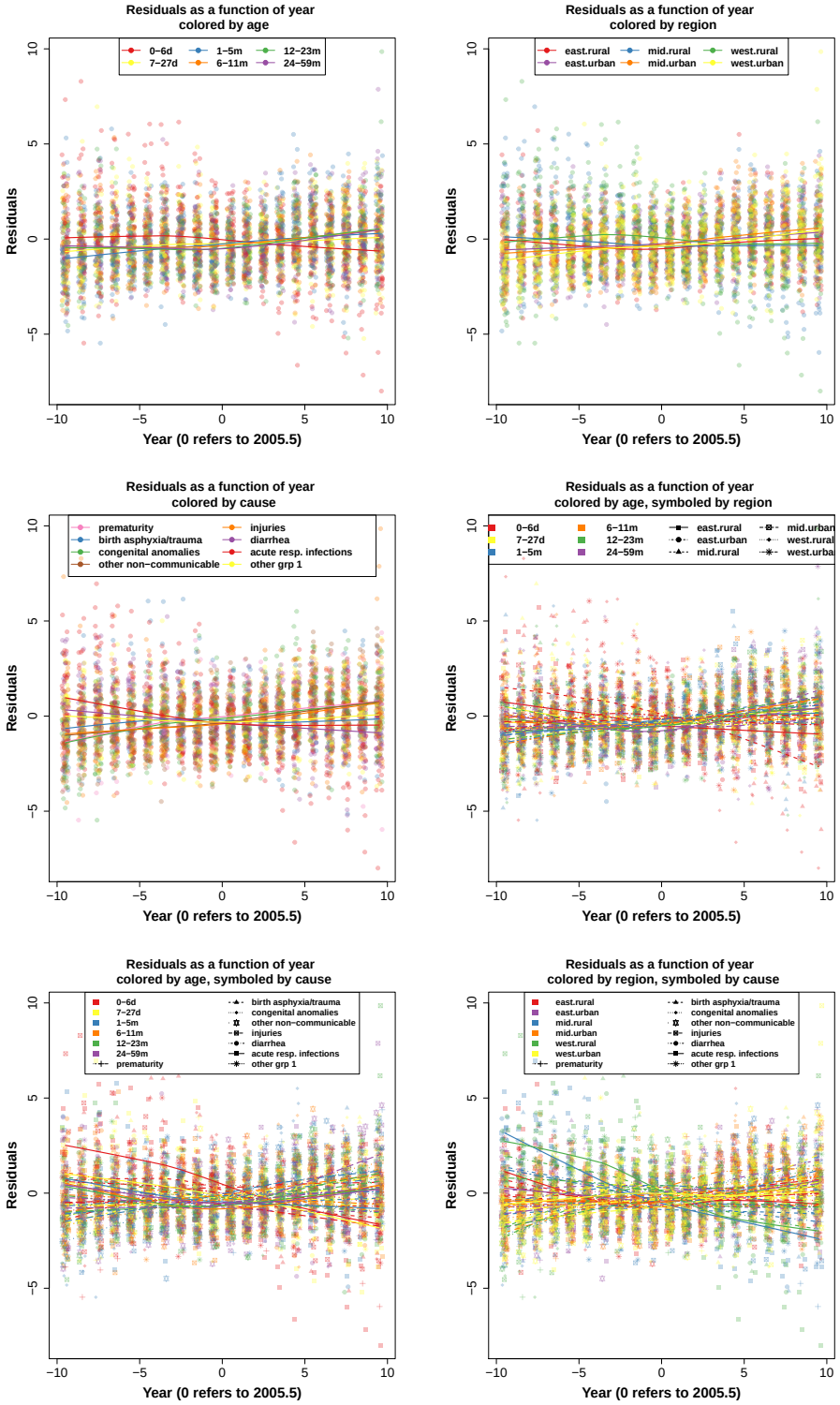


Figure A.3: Plots of residuals for a Poisson GLM with no interactions between time trends and any other variable.

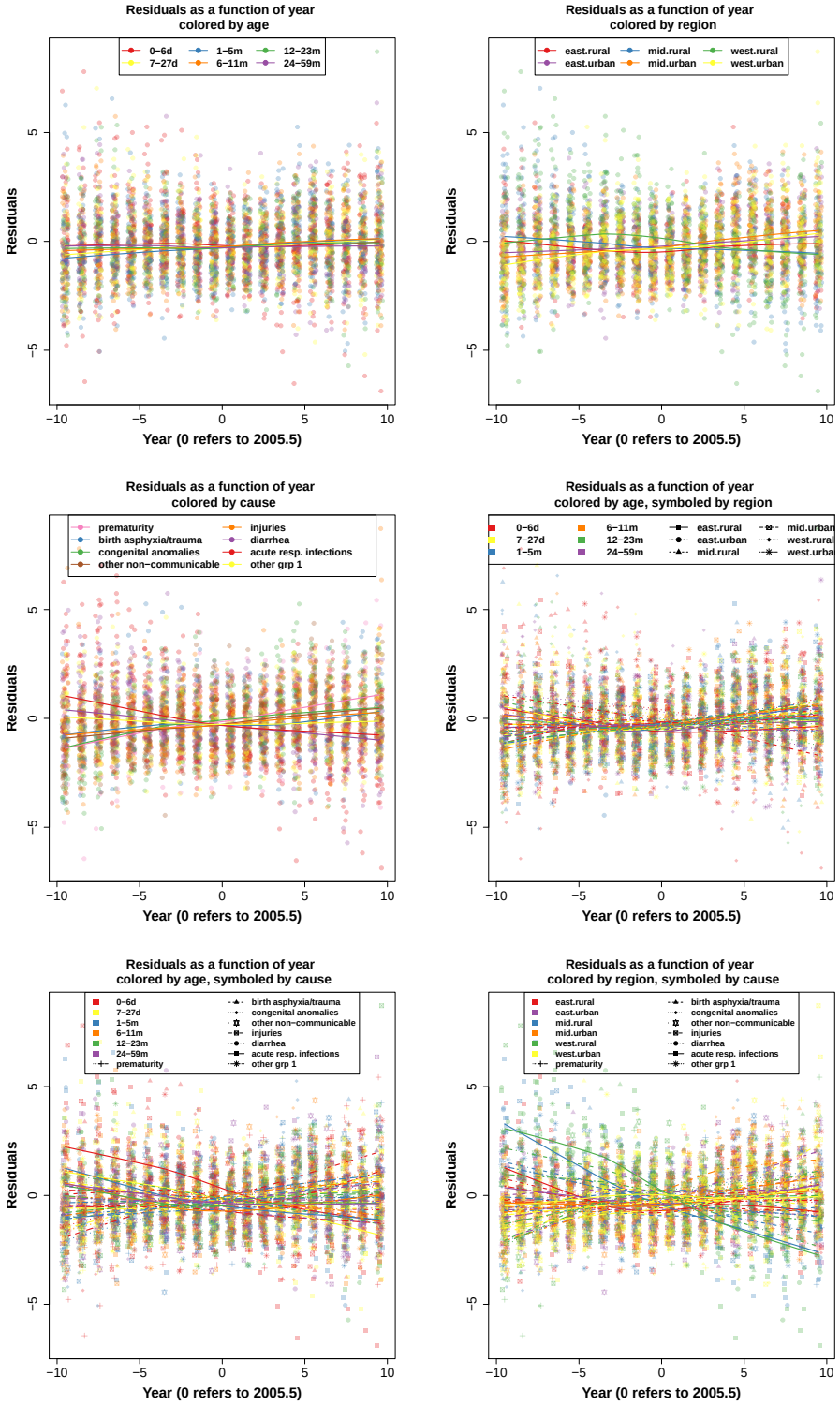


Figure A.4: Plots of residuals for a Poisson GLM with interactions between time trends and age.

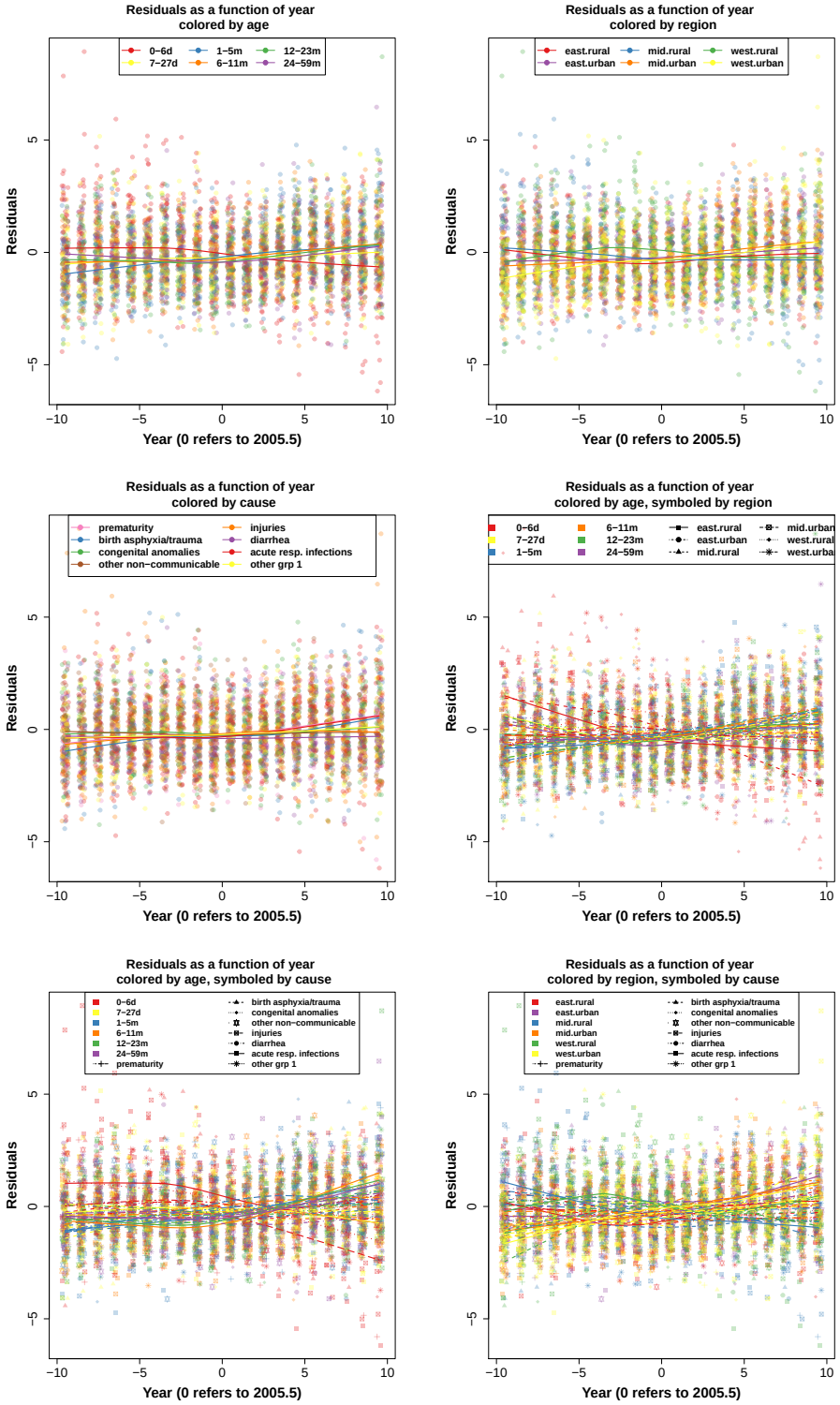


Figure A.5: Plots of residuals for a Poisson GLM with interactions between time trends and cause.

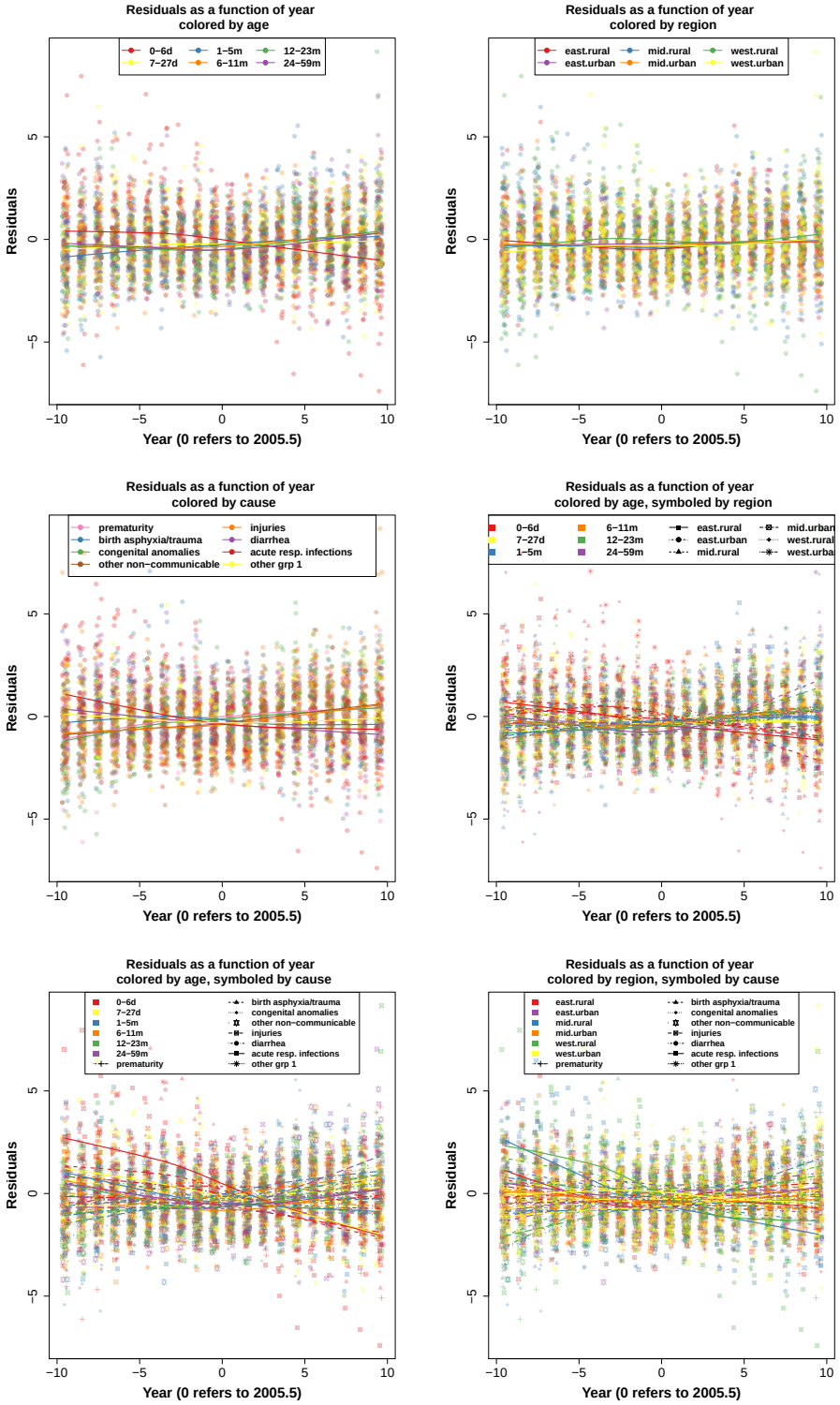


Figure A.6: Plots of residuals for a Poisson GLM with interactions between time trends and region.

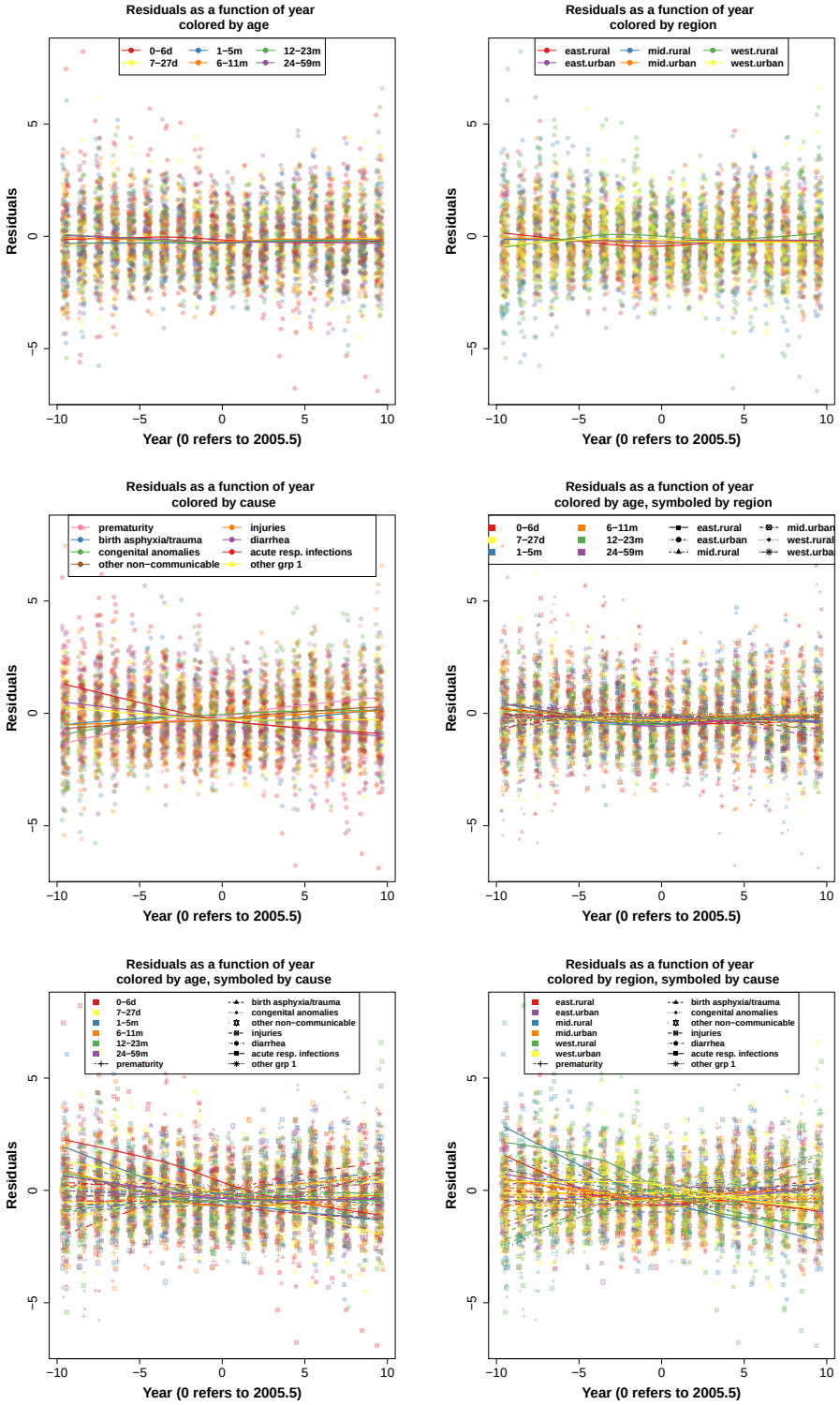


Figure A.7: Plots of residuals for a Poisson GLM with interactions between time trends and age-region.

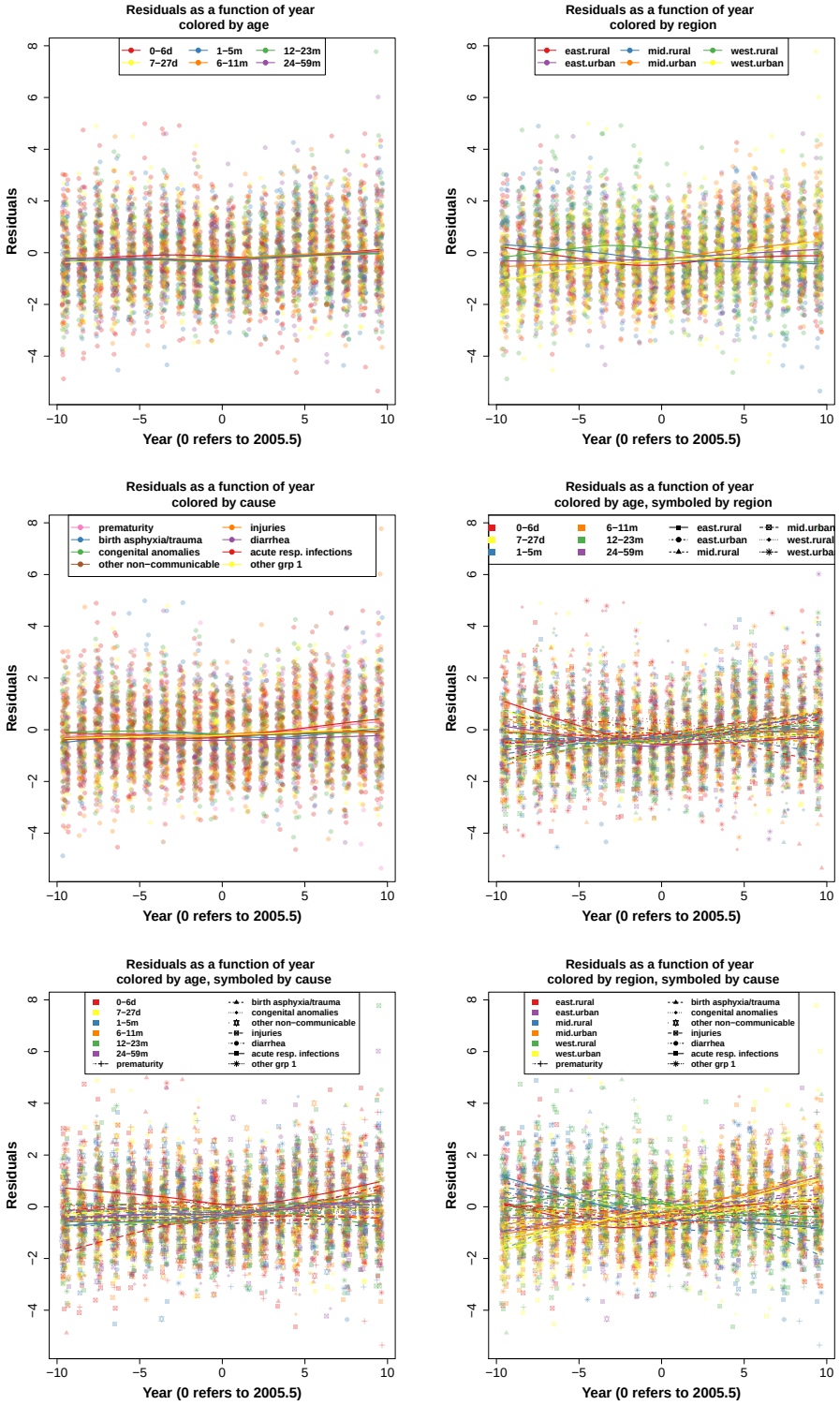


Figure A.8: Plots of residuals for a Poisson GLM with interactions between time trends and age-cause.

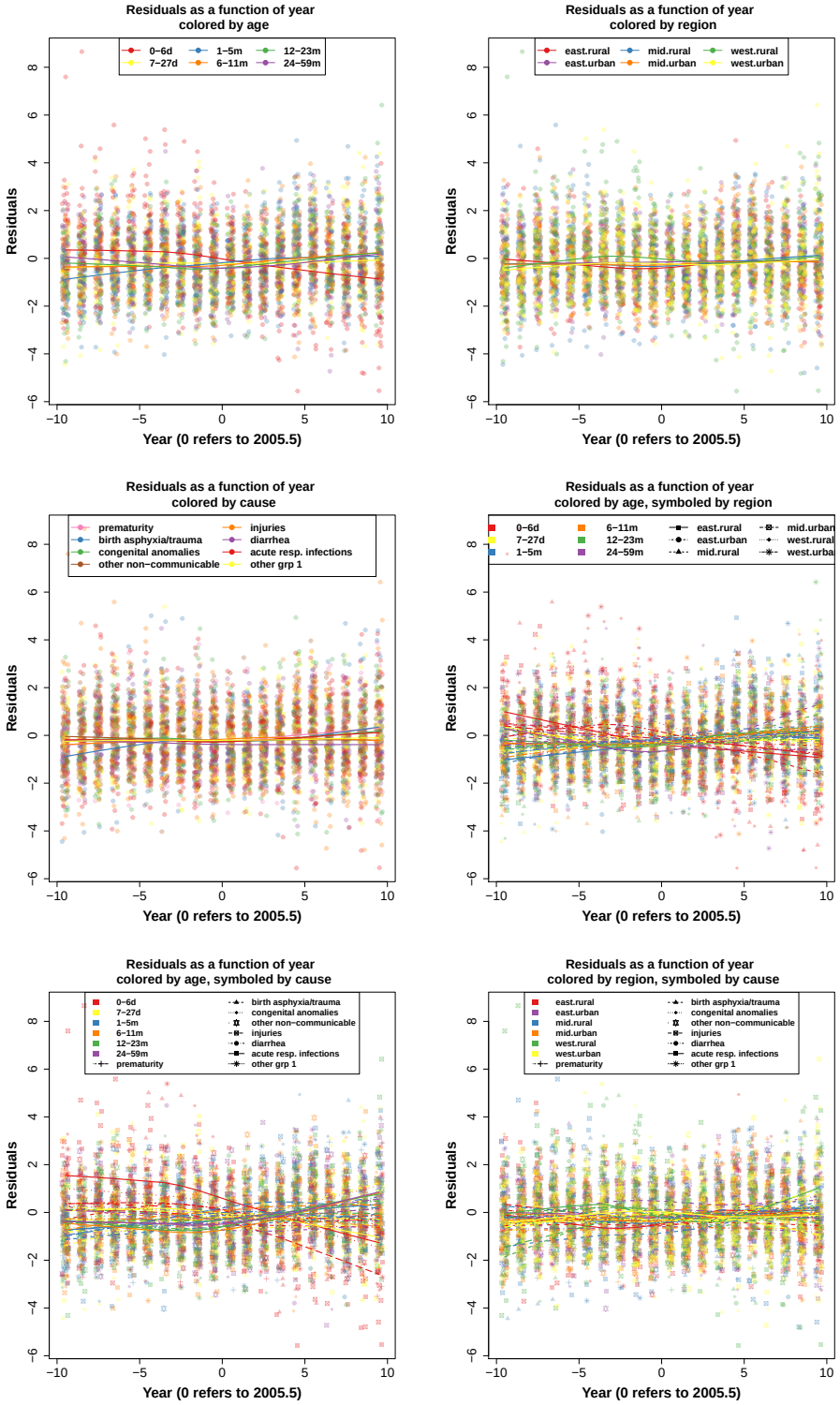


Figure A.9: Plots of residuals for a Poisson GLM with interactions between time trends and region-cause.

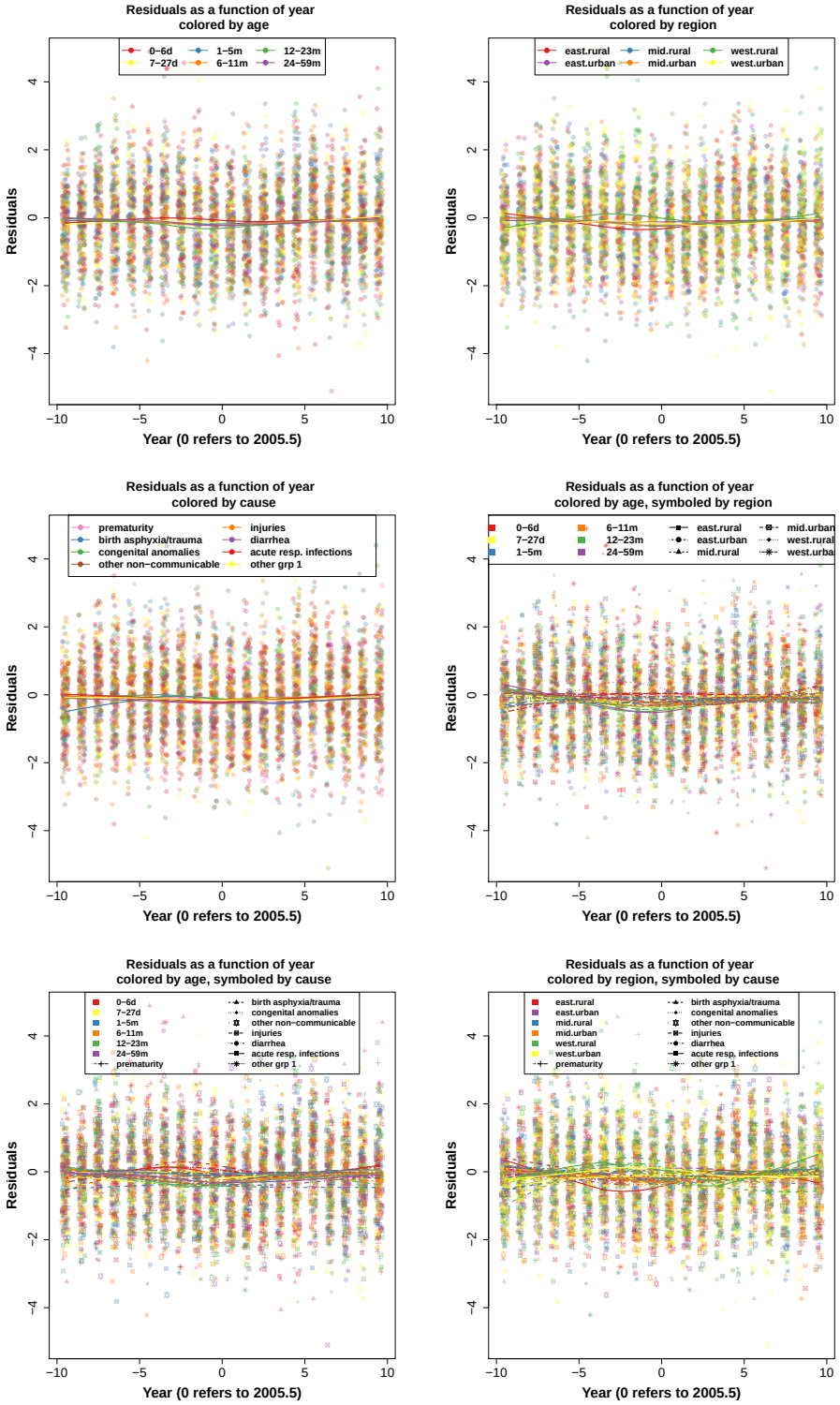


Figure A.10: Plots of residuals for a Poisson GLM with interactions between time trends and age-region-cause.

A.10 Methods and results for deciding random walk variance parameter specification in the MCHSS data

We decided to model random walks for combinations of 2 age groups age (0-6d or not), all 6 regions, and 6 causes (collapsing other communicable and diarrhea, as well as collapsing congenital anomalies and other non-communicable).

In each random walk modeling strata, we fit a model with a sub-strata specific intercept (so if there were multiple substrata of the 6 age groups, 6 regions, and 8 causes within a random walk modeling strata, each had its own intercept) and a random walk whose variance parameter was estimated and plotted. The plots are shown below in Figure A.11.

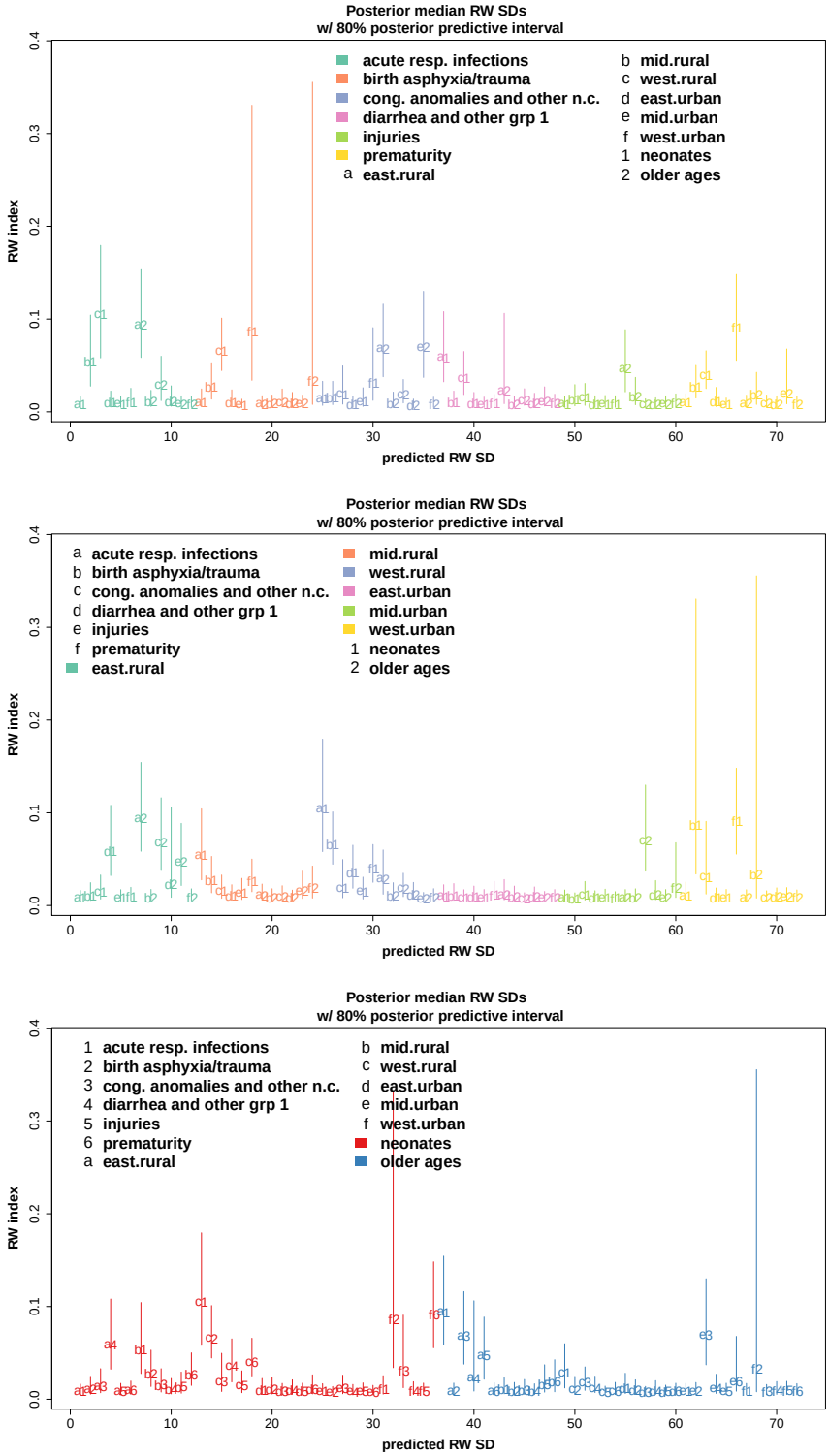
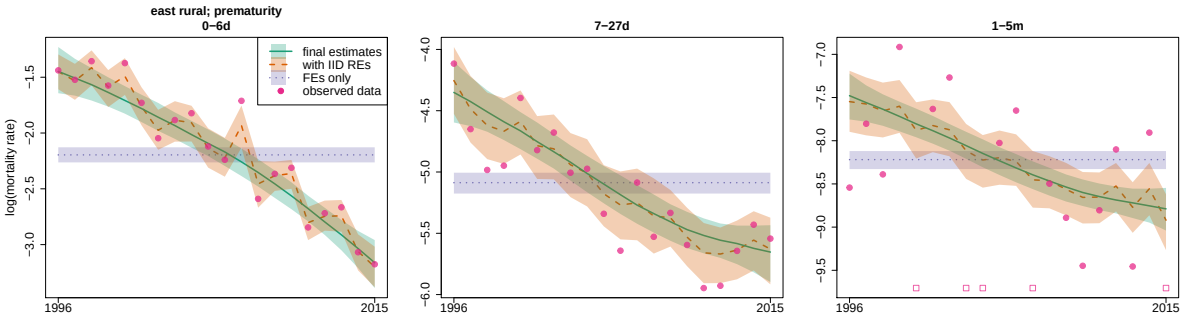
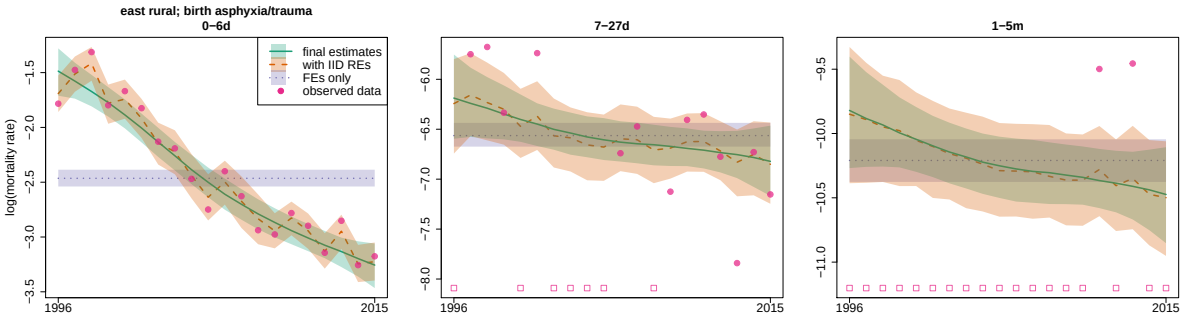
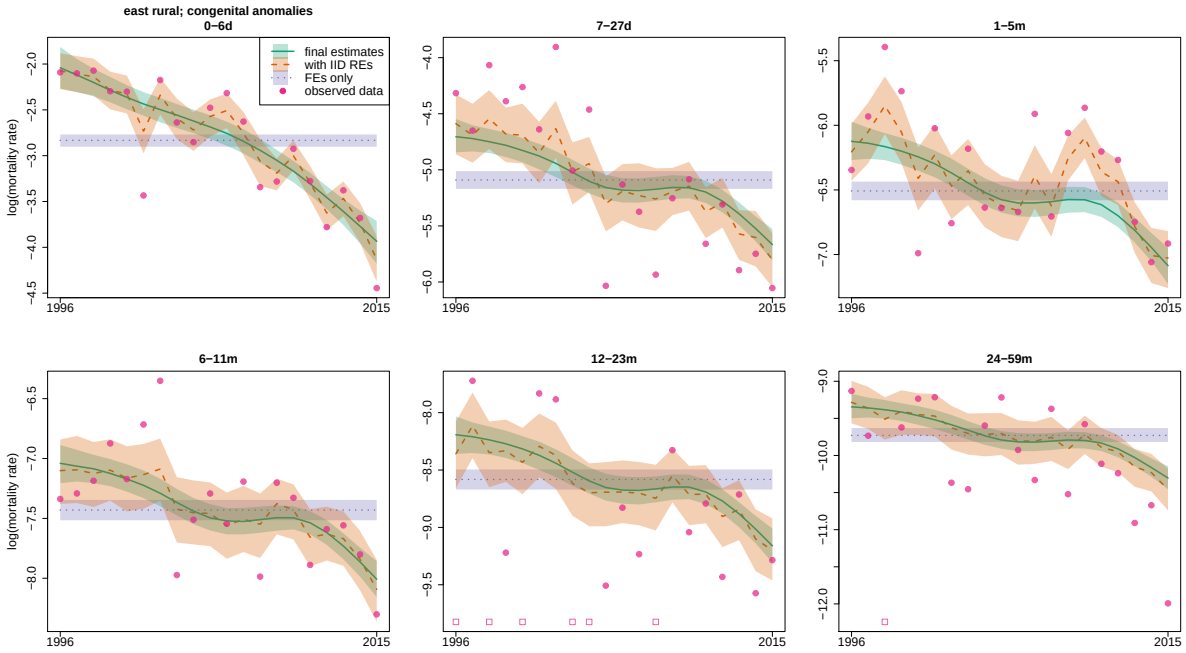


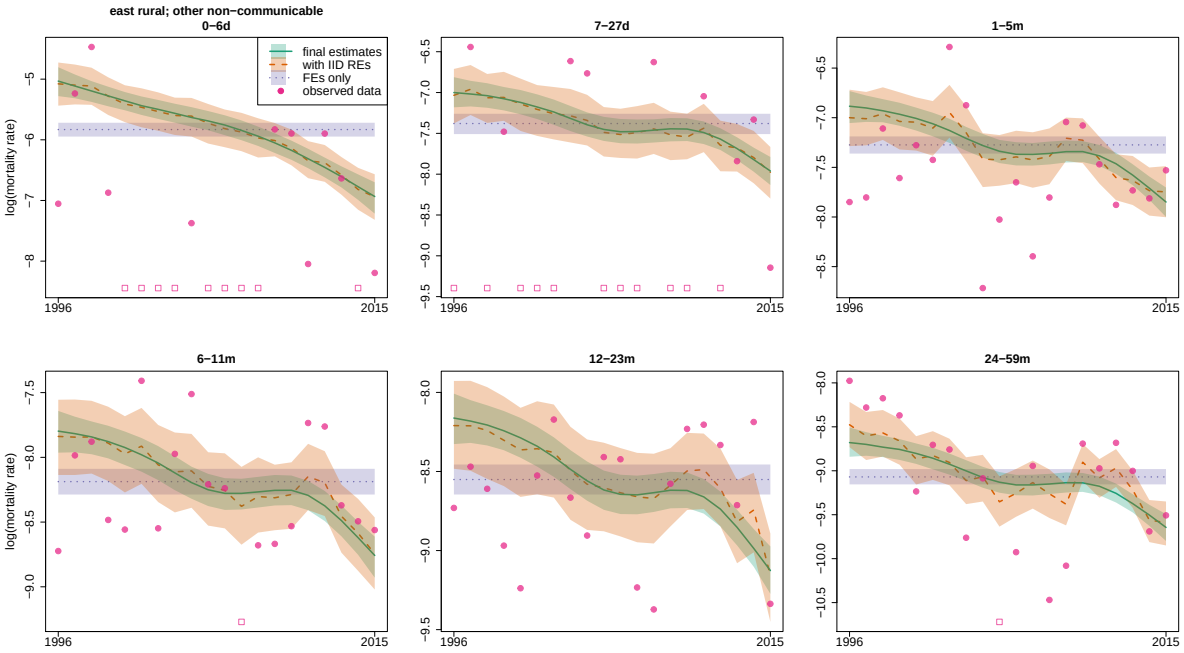
Figure A.11: Plots of estimated random walk standard deviation parameters for Poisson regressions with strata-specific intercepts and a second order random walk in time across all groups for which we fit a random walk in the final analysis of the MCHSS data.

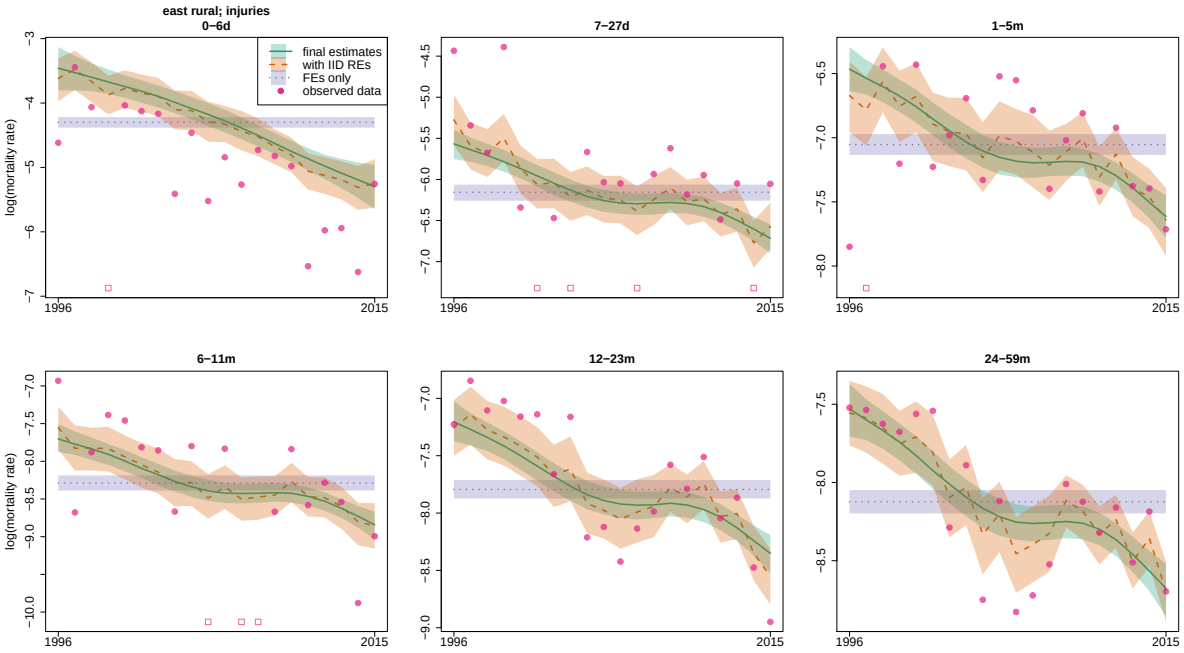
A.11 Graphs of all estimated posterior median log mortality rates with 80% credible intervals from MCHSS data for the final model presented in the main text

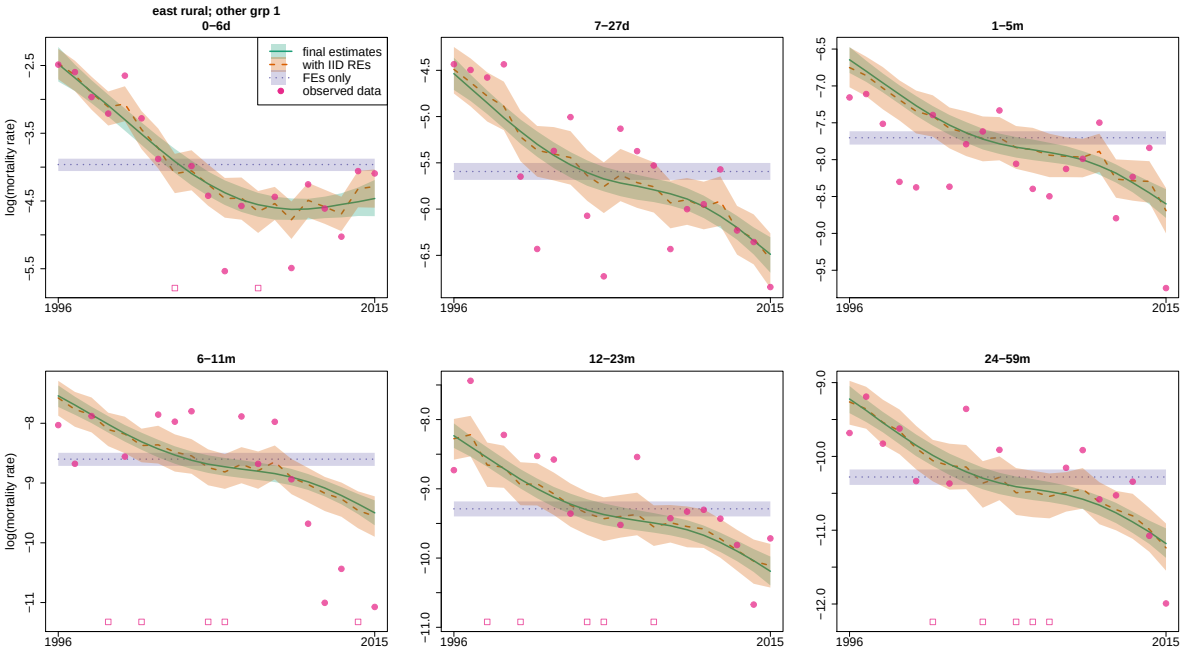

east rural; prematurity
0–6d
7–27d
1–5m
final estimates
with IID REs
FEs only
observed data
log(mortality rate)
1996
2015

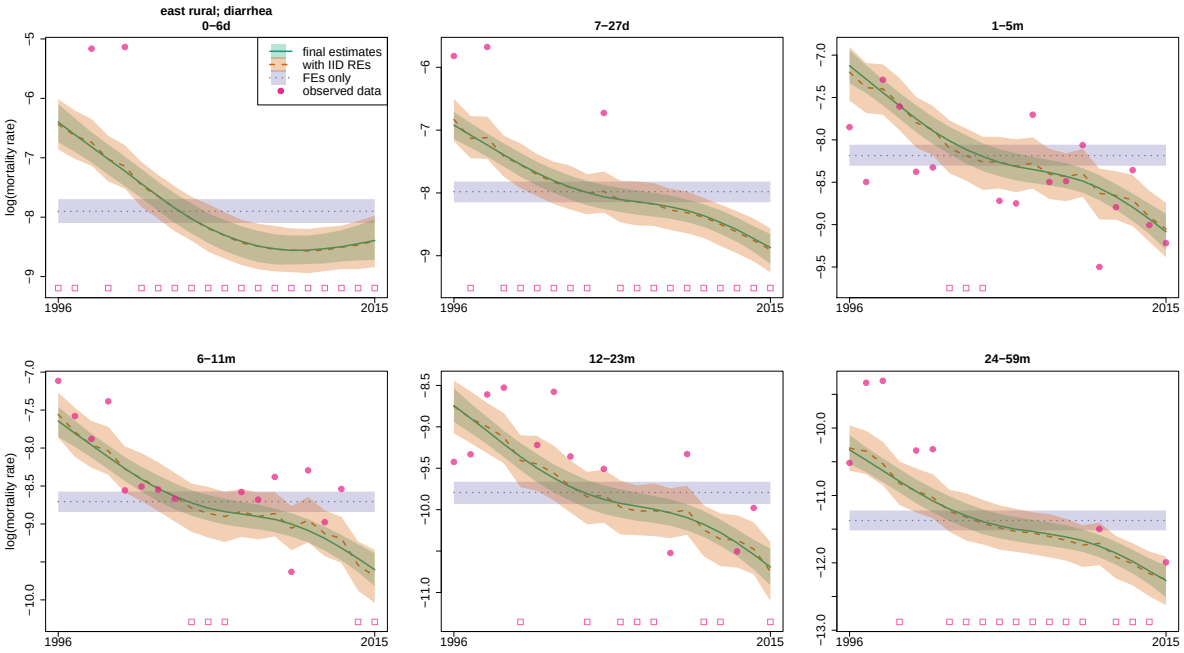

east rural; birth asphyxia/trauma
0–6d
7–27d
1–5m
final estimates
with IID REs
FEs only
observed data
log(mortality rate)
−1.5
−2.0
−2.5
−3.0
−3.5
−6.0
−6.5
−7.0
−7.5
−8.0
−9.5
−10.0
−10.5
−11.0
1996
2015

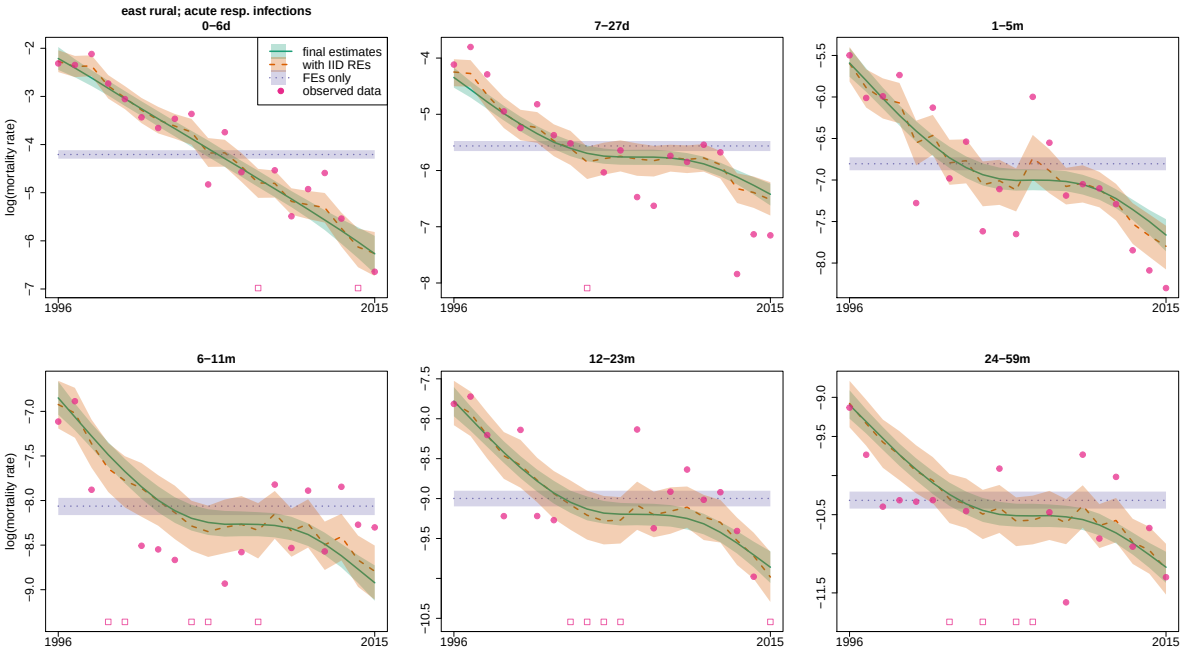

east rural; congenital anomalies
0–6d
7–27d
1–5m
6–11m
12–23m
24–59m
final estimates
with IID REs
FEs only
observed data
log(mortality rate)
1996
2015

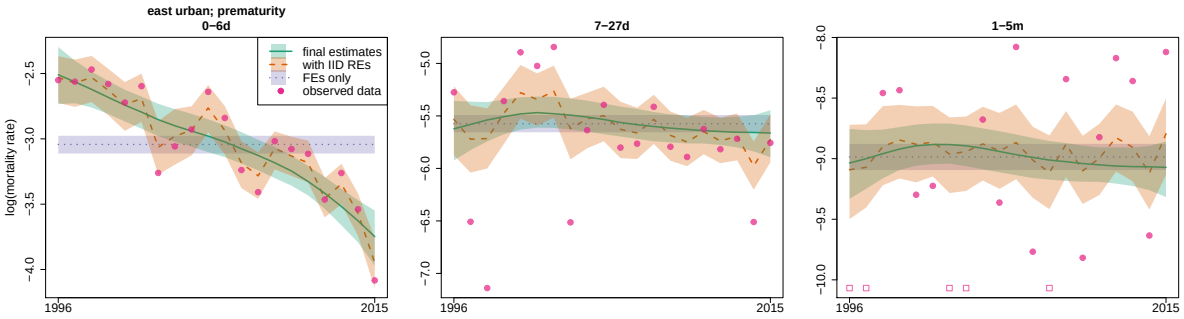

east rural; other non−communicable
0−6d
7−27d
1−5m
6−11m
12−23m
24−59m
final estimates
with IID REs
FEs only
observed data
log(mortality rate)
1996
2015

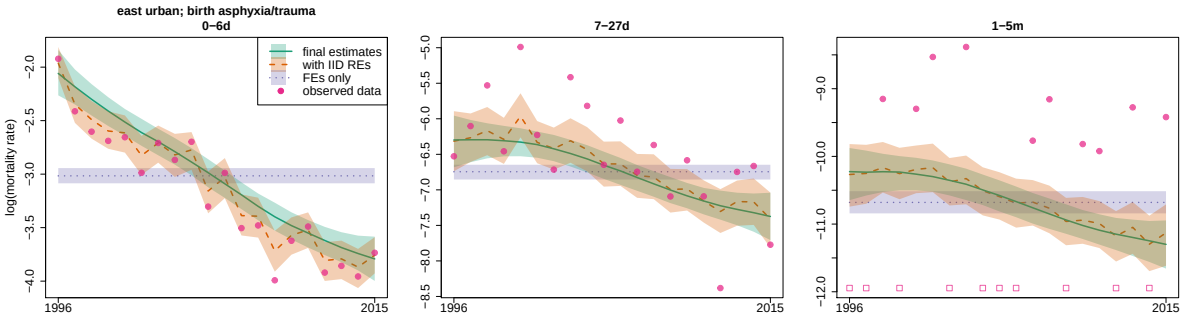

east rural; injuries
0–6d
7–27d
1–5m
6–11m
12–23m
24–59m
final estimates
with IID REs
FEs only
observed data
log(mortality rate)
1996
2015

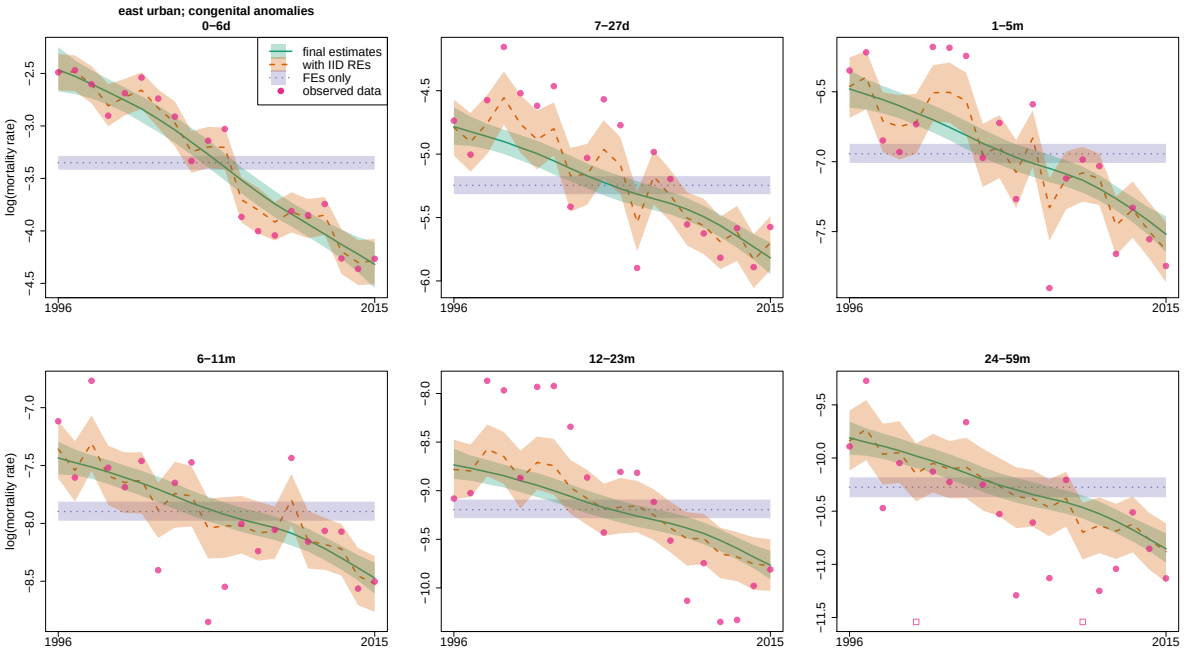

east rural; other grp 1
0–6d
7–27d
1–5m
6–11m
12–23m
24–59m
final estimates
with IID REs
FEs only
observed data
log(mortality rate)
1996
2015

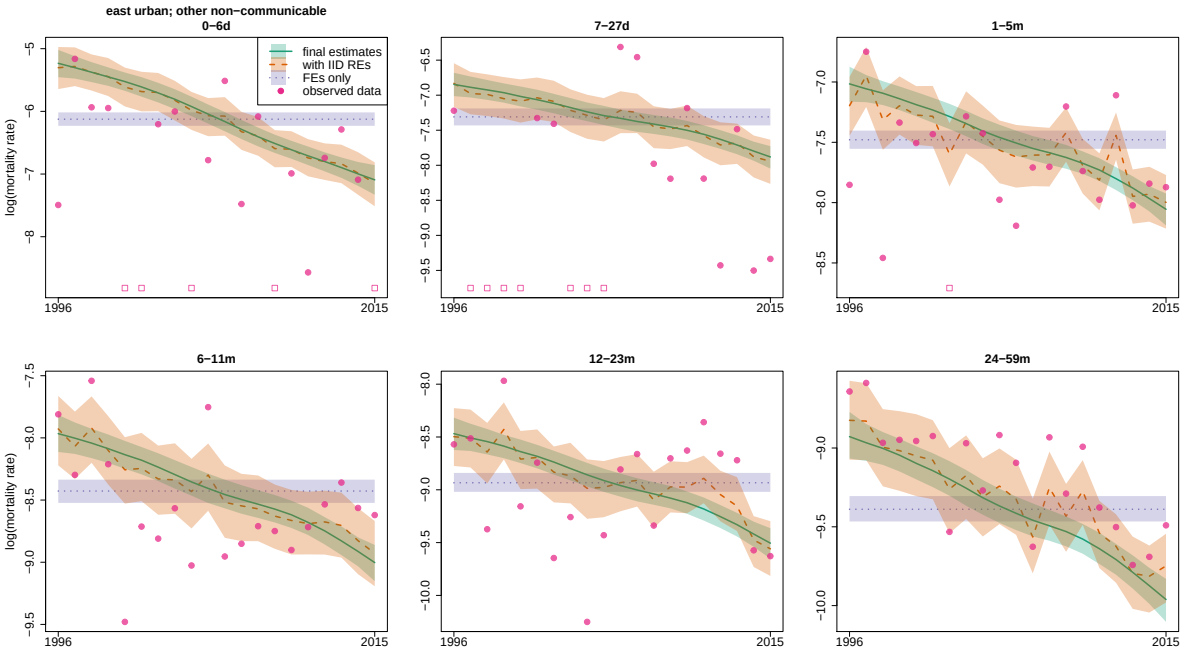

east rural; diarrhea
0–6d
7–27d
1–5m
6–11m
12–23m
24–59m
final estimates
with IID REs
FEs only
observed data
log(mortality rate)
1996
2015

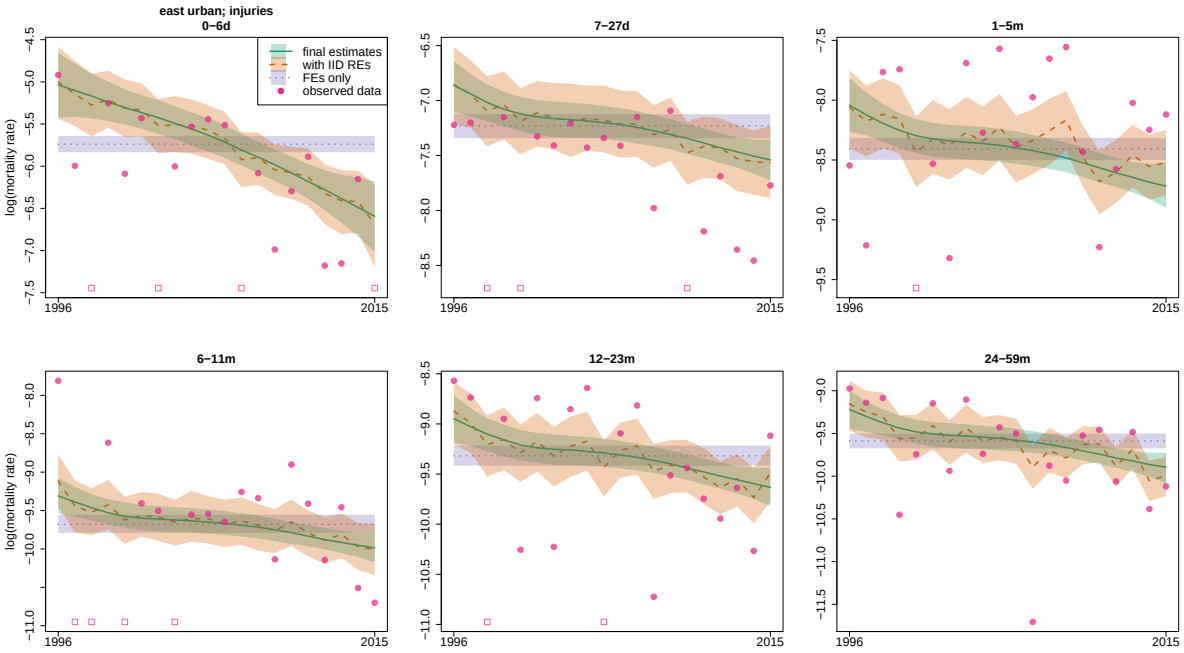

east rural; acute resp. infections
0–6d
7–27d
1–5m
6–11m
12–23m
24–59m
final estimates
with IID REs
FEs only
observed data
log(mortality rate)
1996
2015

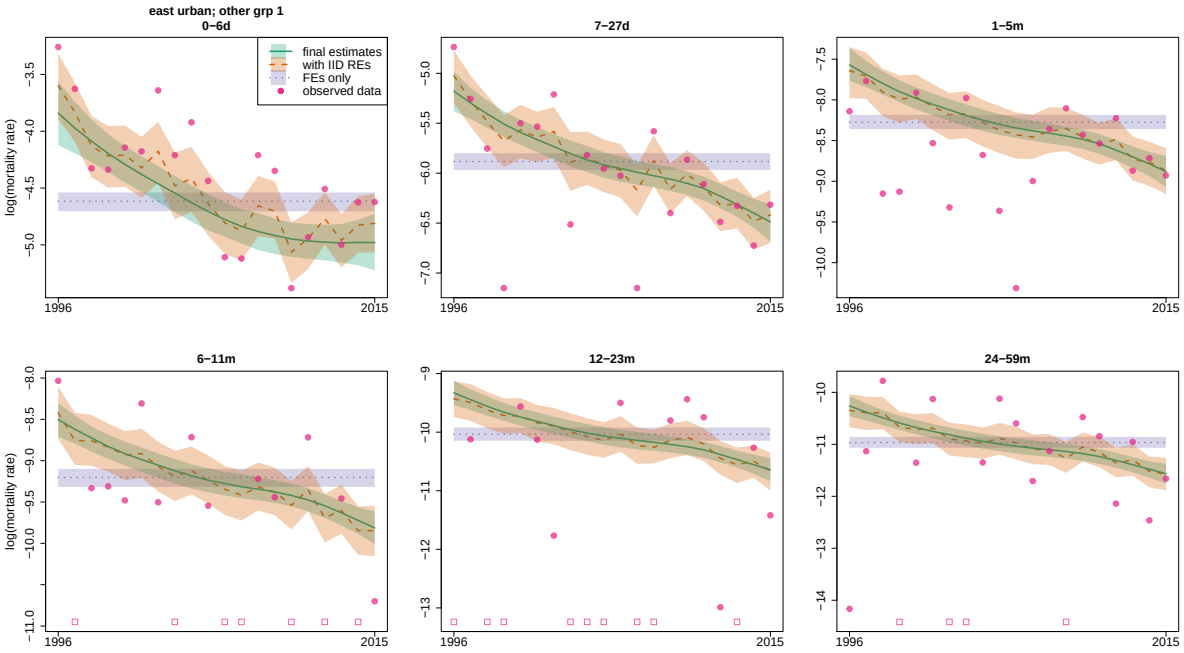

east urban; prematurity
0–6d
7–27d
1–5m
final estimates
with IID REs
FEs only
observed data
log(mortality rate)
−2.5
−3.0
−3.5
−4.0
−5.0
−5.5
−6.0
−6.5
−7.0
−8.0
−8.5
−9.0
−9.5
−10.0
1996
2015

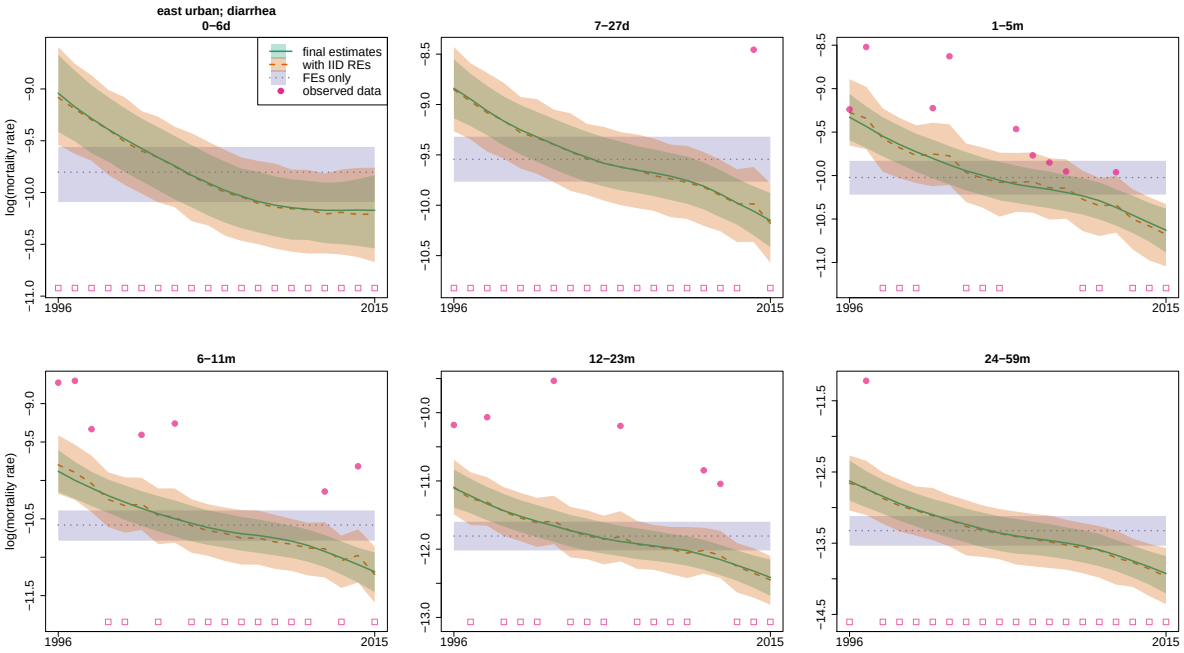

east urban; birth asphyxia/trauma
0–6d
7–27d
1–5m
final estimates
with IID REs
FEs only
observed data
log(mortality rate)
1996
2015

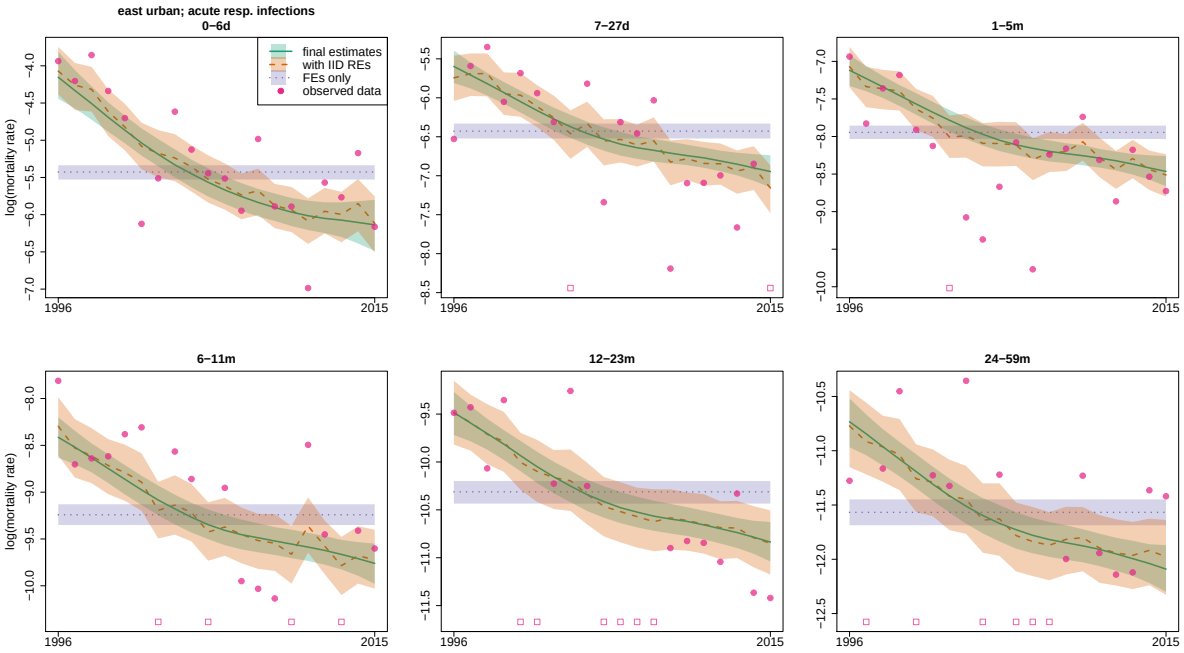

east urban; congenital anomalies
0–6d
7–27d
1–5m
6–11m
12–23m
24–59m
final estimates
with IID REs
FEs only
observed data
log(mortality rate)
1996
2015

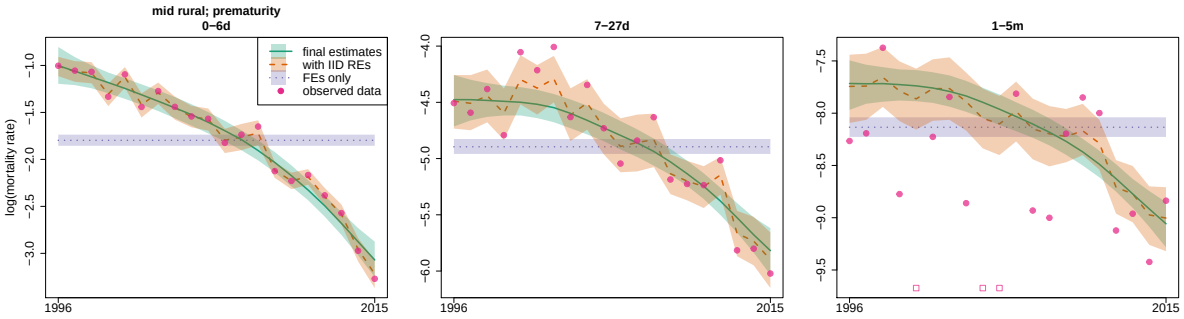

east urban; other non−communicable
0−6d
7−27d
1−5m
6−11m
12−23m
24−59m
final estimates
with IID REs
FEs only
observed data
log(mortality rate)
1996
2015

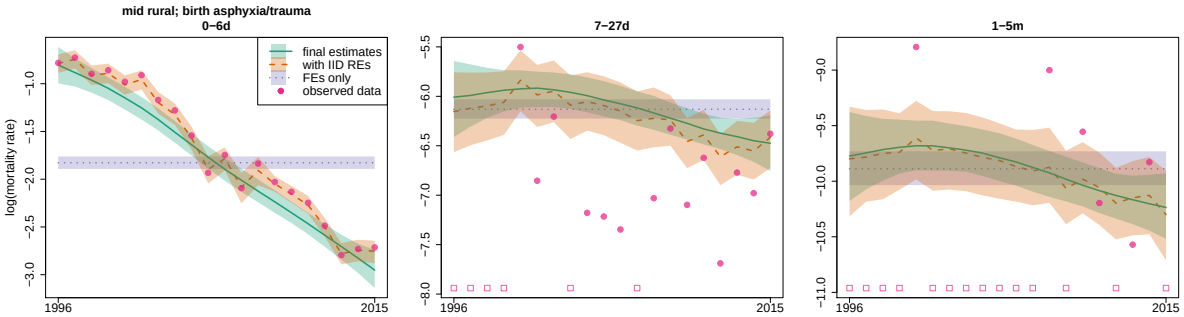

east urban; injuries
0–6d
7–27d
1–5m
6–11m
12–23m
24–59m
final estimates
with IID REs
FEs only
observed data
log(mortality rate)
1996
2015

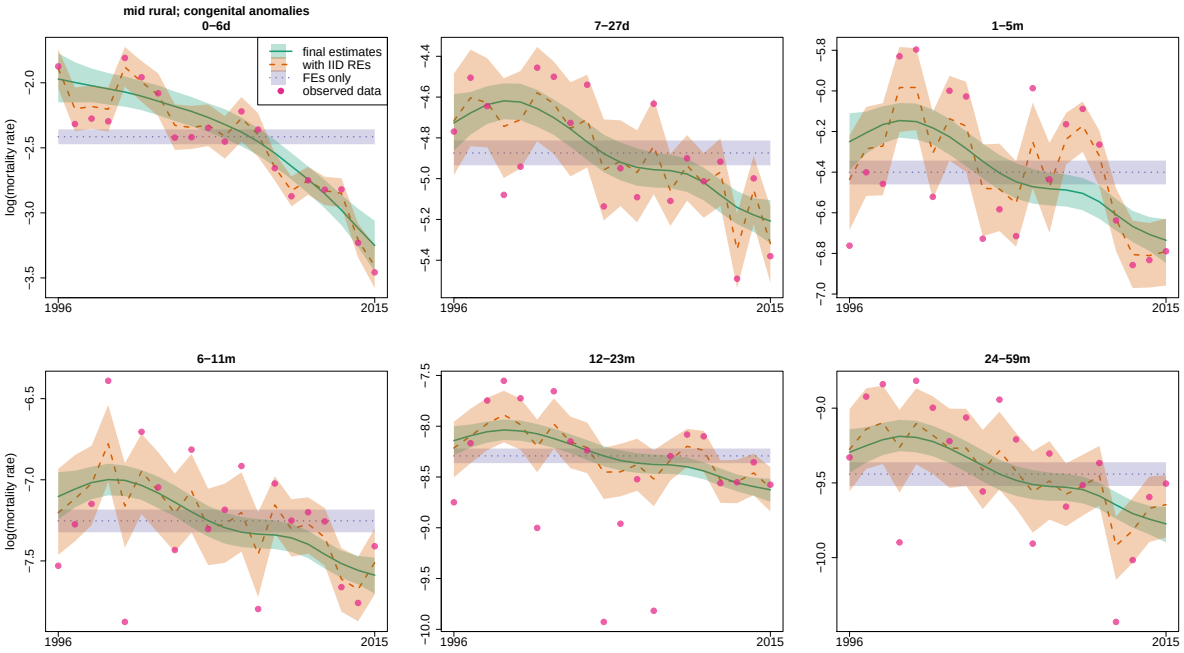

east urban; other grp 1
0–6d
7–27d
1–5m
6–11m
12–23m
24–59m
final estimates
with IID REs
FEs only
observed data
log(mortality rate)
1996
2015

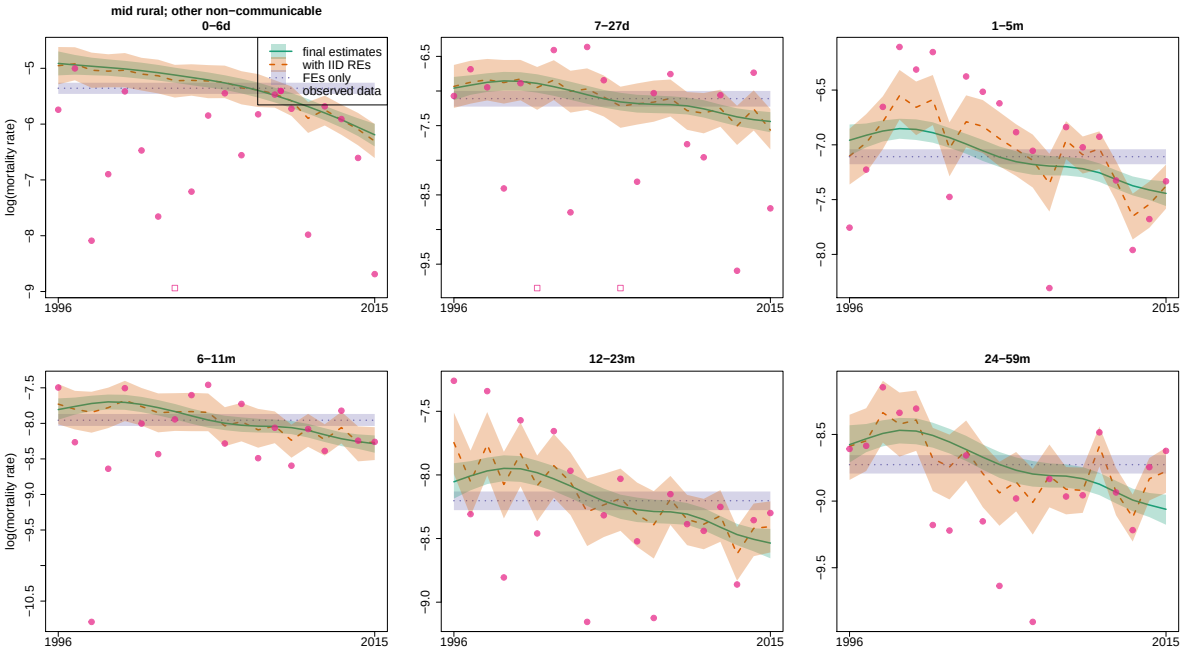

east urban; diarrhea
0–6d
7–27d
1–5m
6–11m
12–23m
24–59m
final estimates
with IID REs
FEs only
observed data
log(mortality rate)
1996
2015

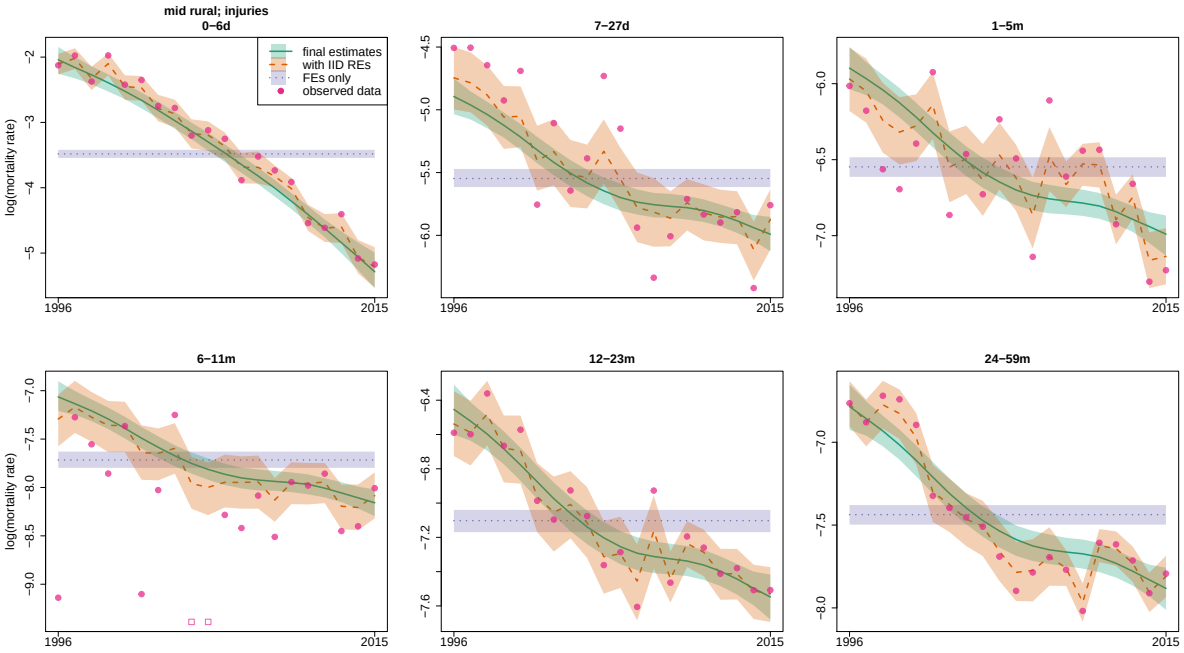

east urban; acute resp. infections
0–6d
7–27d
1–5m
6–11m
12–23m
24–59m
final estimates
with IID REs
FEs only
observed data
log(mortality rate)
1996
2015

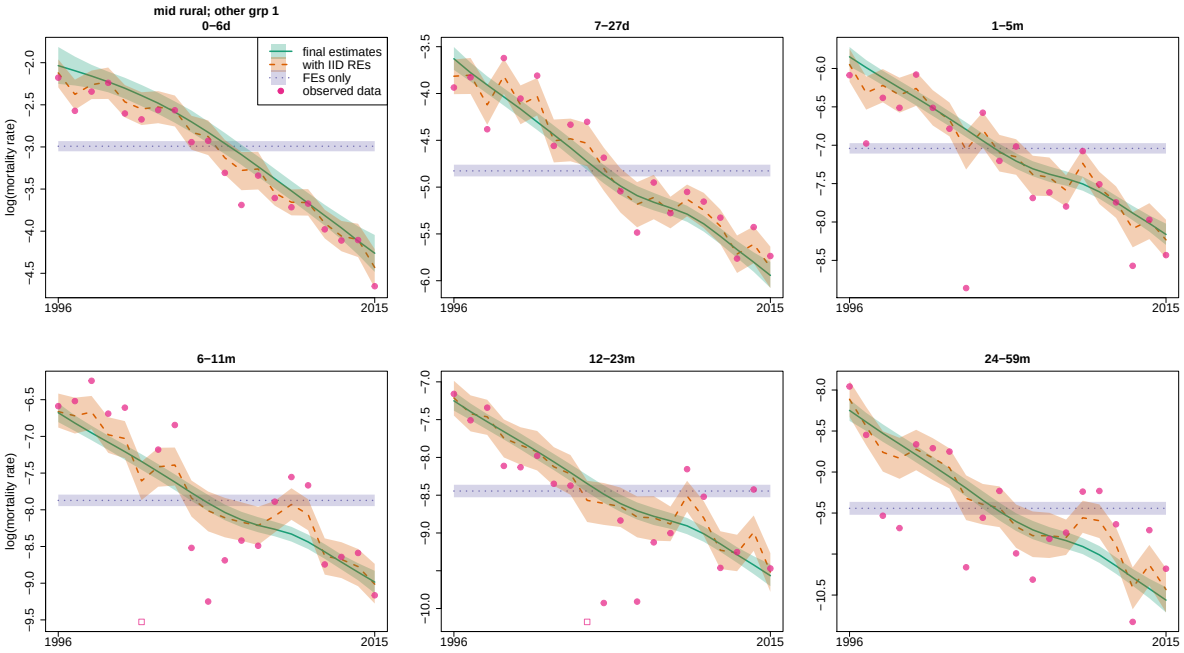

mid rural; prematurity
0–6d
7–27d
1–5m
final estimates
with IID REs
FEs only
observed data
log(mortality rate)
1996
2015

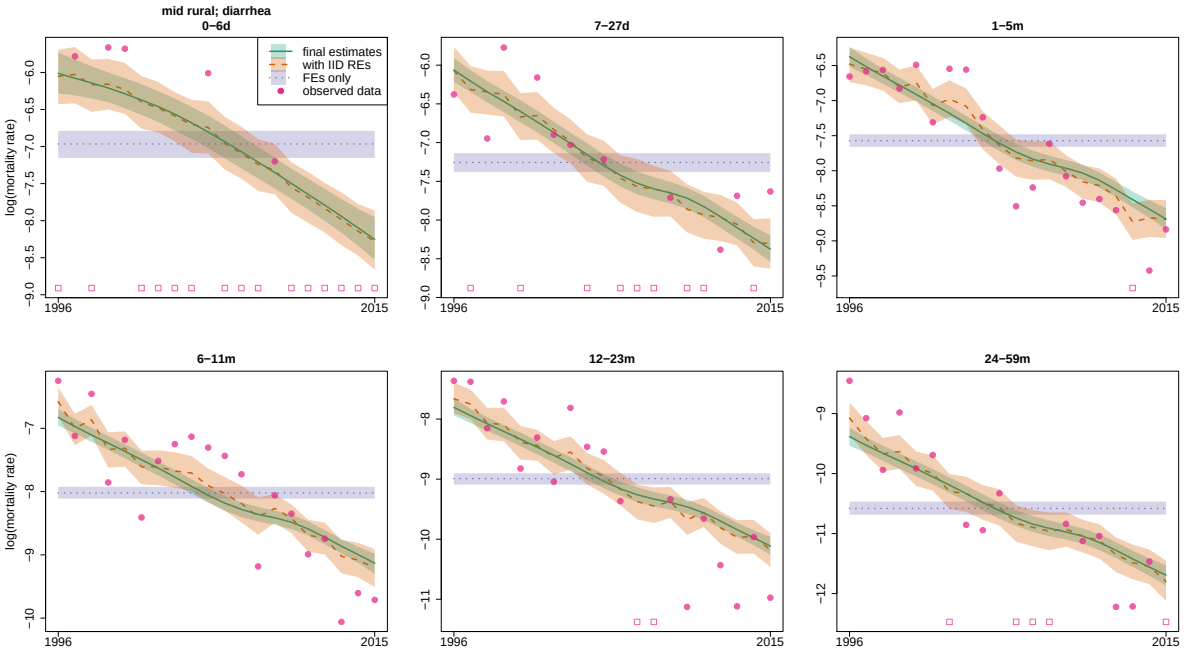

mid rural; birth asphyxia/trauma
0–6d
7–27d
1–5m
final estimates
with IID REs
FEs only
observed data
log(mortality rate)
1996
2015

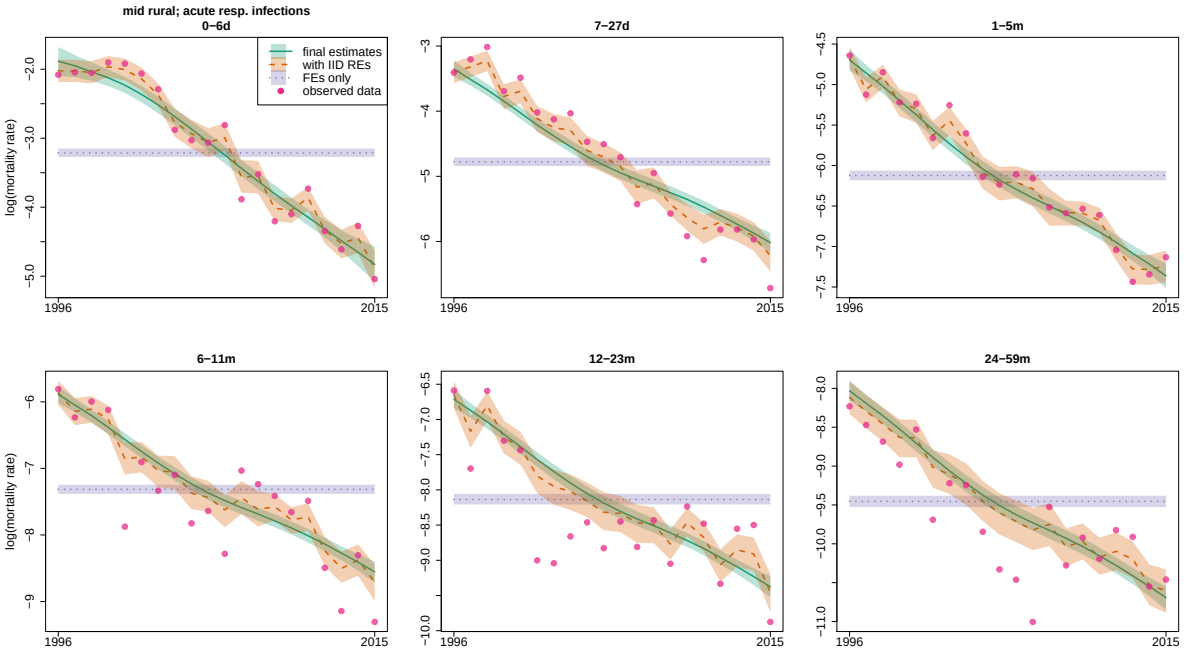

mid rural; congenital anomalies
0–6d
7–27d
1–5m
6–11m
12–23m
24–59m
final estimates
with IID REs
FEs only
observed data
log(mortality rate)
1996
2015

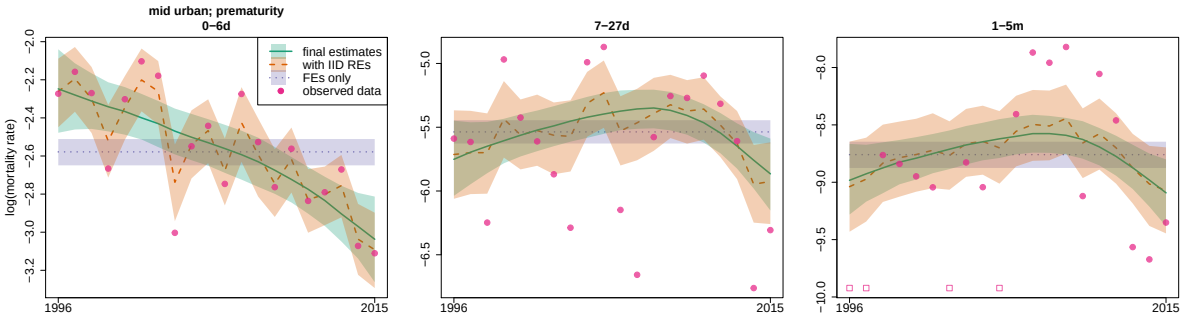

mid rural; other non−communicable
0−6d
7−27d
1−5m
6−11m
12−23m
24−59m
final estimates
with IID REs
FEs only
observed data
log(mortality rate)
1996
2015

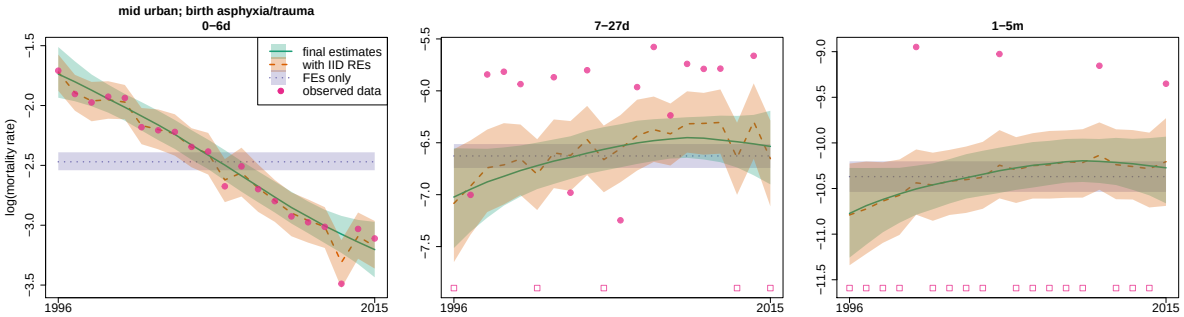

mid rural; injuries
0–6d
7–27d
1–5m
6–11m
12–23m
24–59m
final estimates
with IID REs
FEs only
observed data
log(mortality rate)
1996
2015

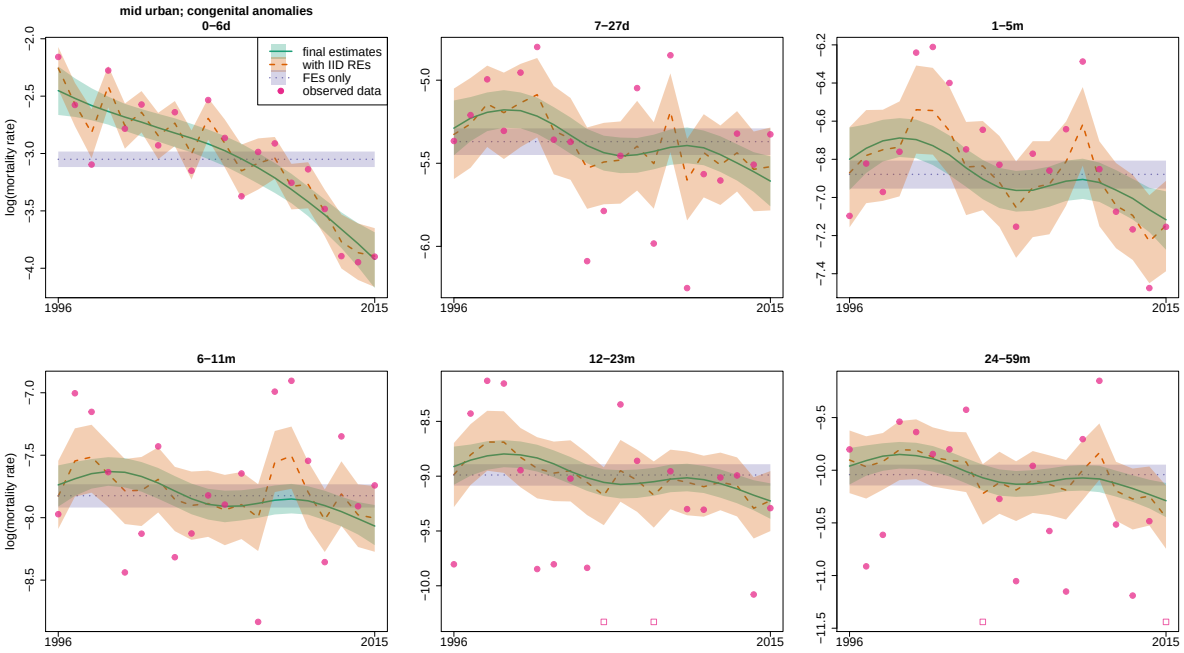

mid rural; other grp 1
0–6d
7–27d
1–5m
6–11m
12–23m
24–59m
final estimates
with IID REs
FEs only
observed data
log(mortality rate)
1996
2015

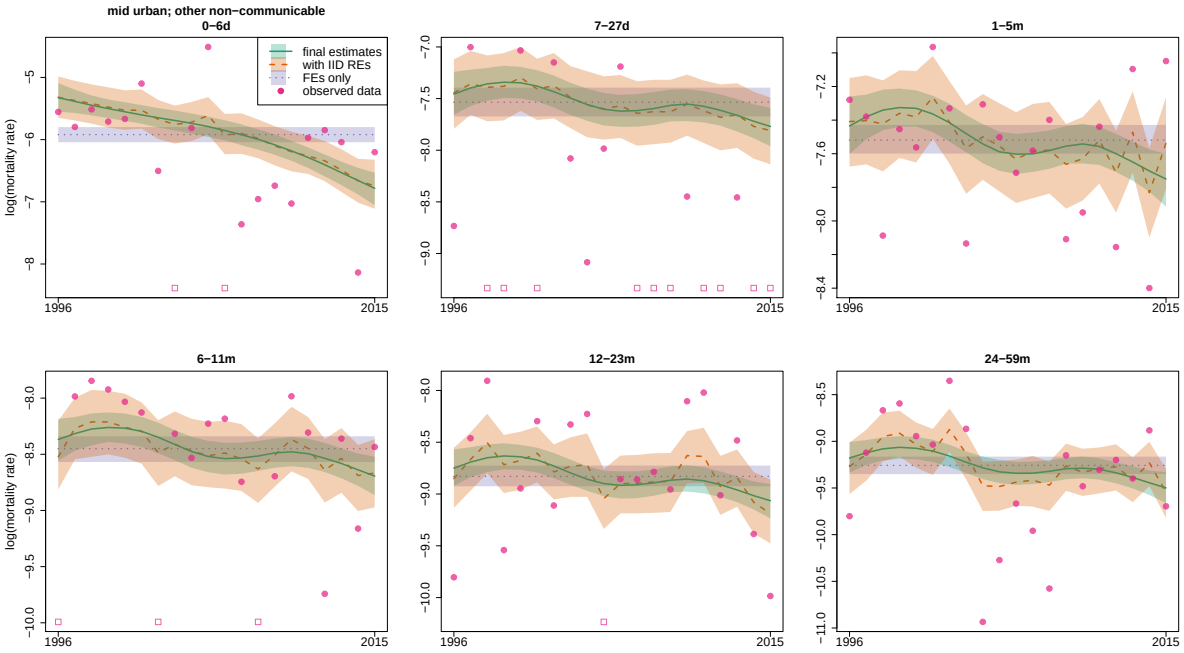

mid rural; diarrhea
0–6d
7–27d
1–5m
6–11m
12–23m
24–59m
final estimates
with IID REs
FEs only
observed data
log(mortality rate)
1996
2015

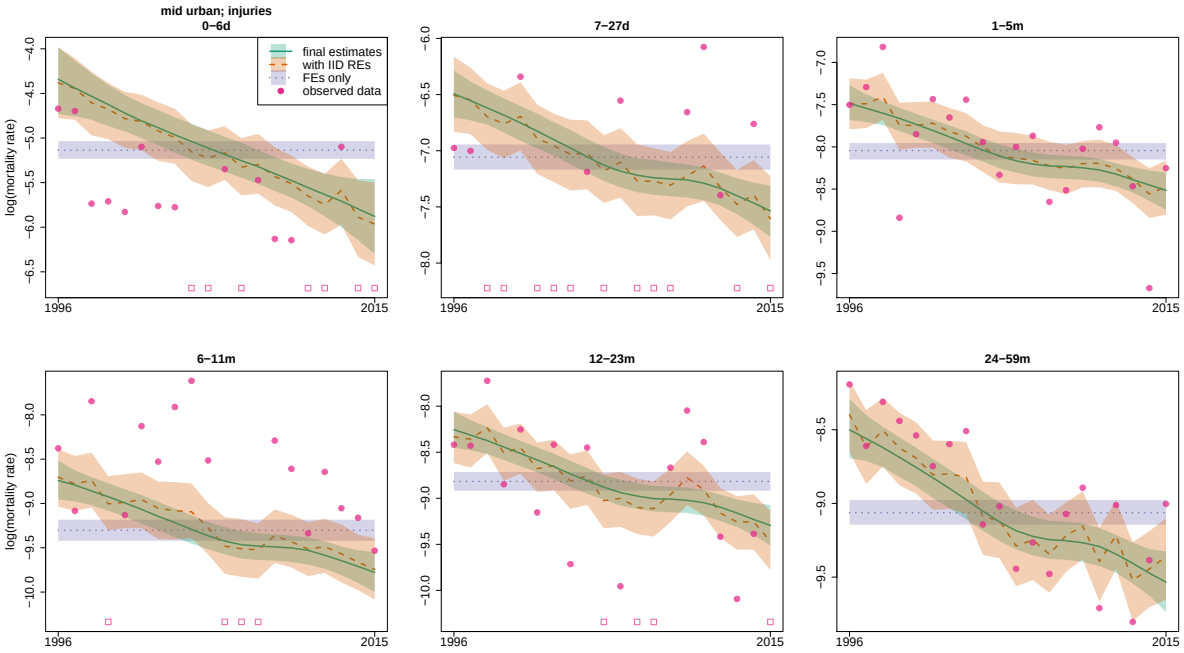

mid rural; acute resp. infections
0–6d
7–27d
1–5m
6–11m
12–23m
24–59m
final estimates
with IID REs
FEs only
observed data
log(mortality rate)
1996
2015

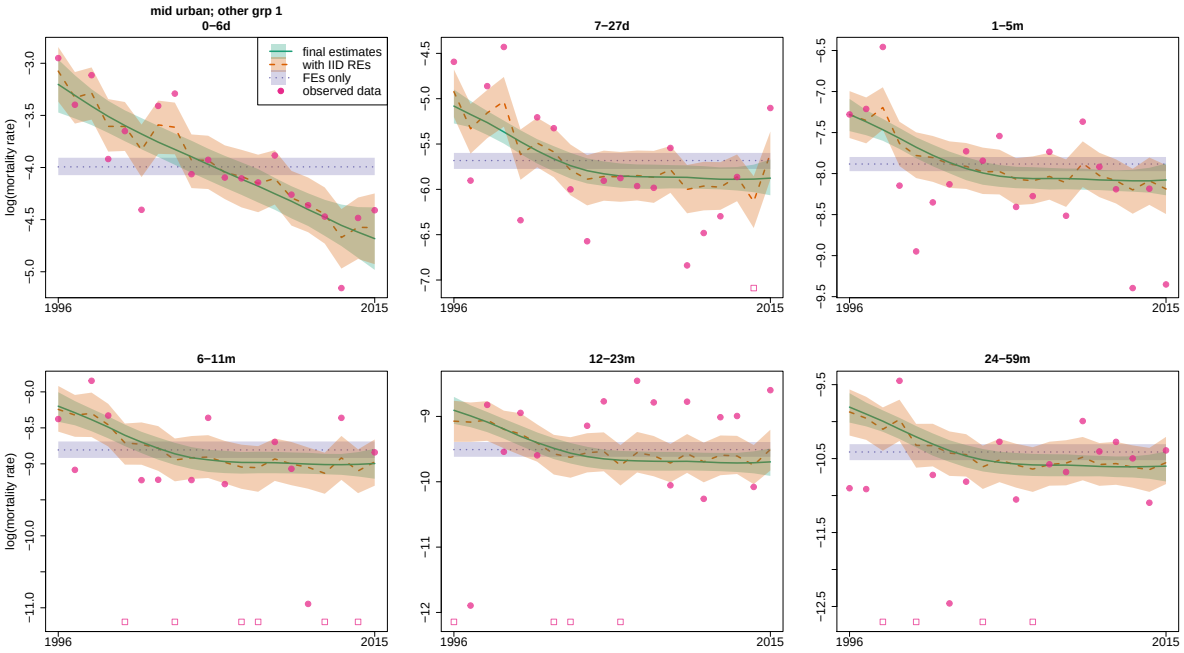

mid urban; prematurity
0−6d
7−27d
1−5m
final estimates
with IID REs
FEs only
observed data
log(mortality rate)
−2.0
−2.2
−2.4
−2.6
−2.8
−3.0
−3.2
−5.0
−5.5
−6.0
−6.5
−8.0
−8.5
−9.0
−9.5
−10.0
1996
2015

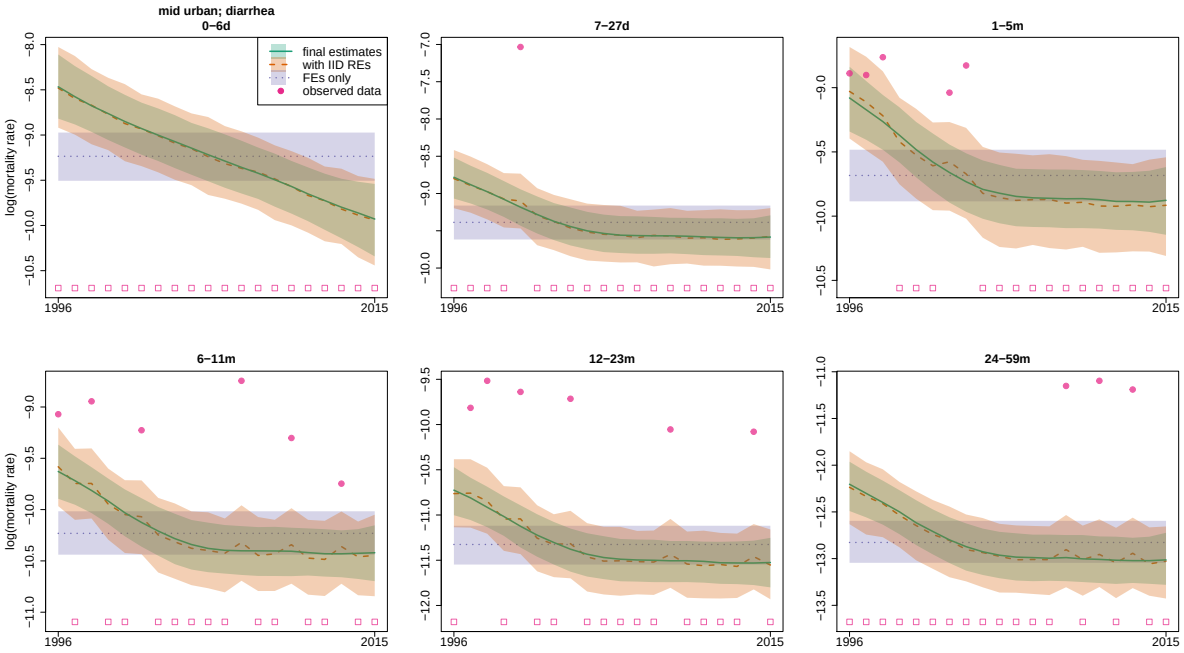

mid urban; birth asphyxia/trauma
0–6d
7–27d
1–5m
final estimates
with IID REs
FEs only
observed data
log(mortality rate)
1996
2015

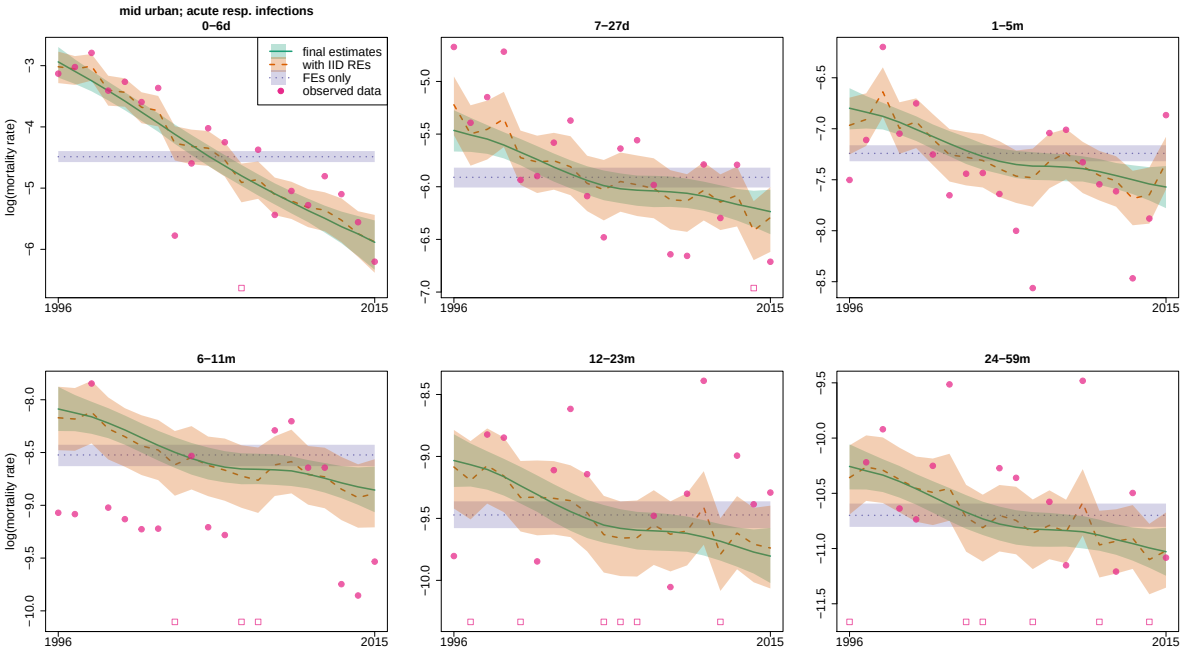

mid urban; congenital anomalies
0–6d
7–27d
1–5m
6–11m
12–23m
24–59m
final estimates
with IID REs
FEs only
observed data
log(mortality rate)
1996
2015

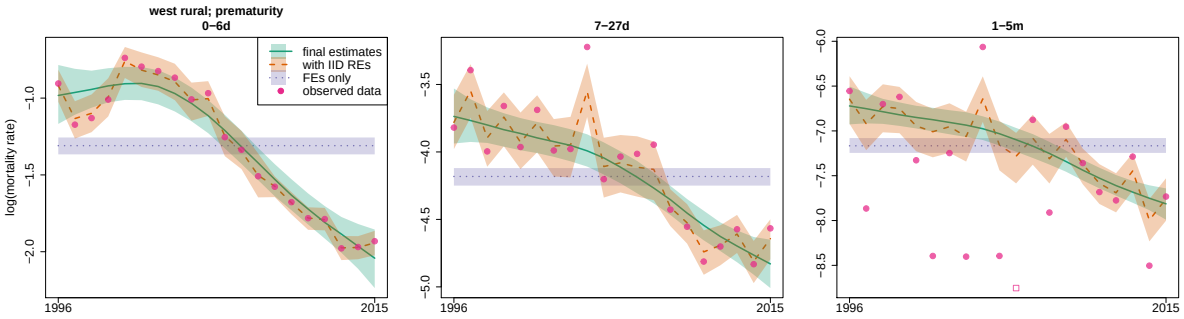

mid urban; other non−communicable
0−6d
7−27d
1−5m
6−11m
12−23m
24−59m
final estimates
with IID REs
FEs only
observed data
log(mortality rate)
1996
2015

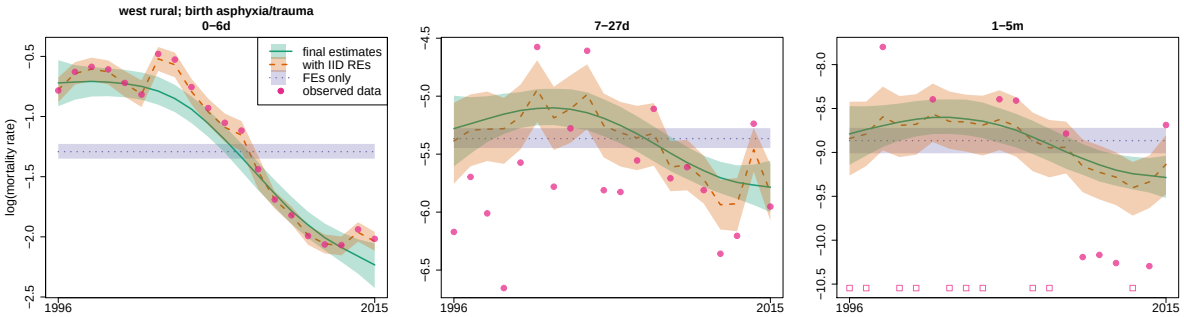

mid urban; injuries
0–6d
7–27d
1–5m
6–11m
12–23m
24–59m
final estimates
with IID REs
FEs only
observed data
log(mortality rate)
1996
2015

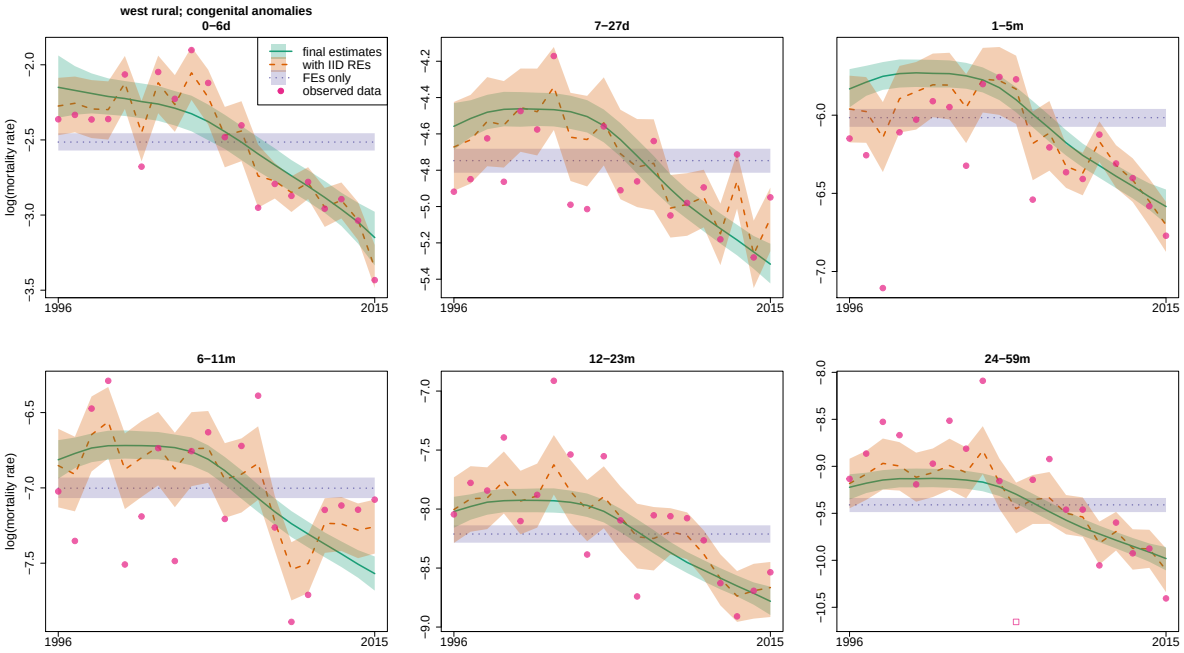

mid urban; other grp 1
0–6d
7–27d
1–5m
6–11m
12–23m
24–59m
final estimates
with IID REs
FEs only
observed data
log(mortality rate)
1996
2015

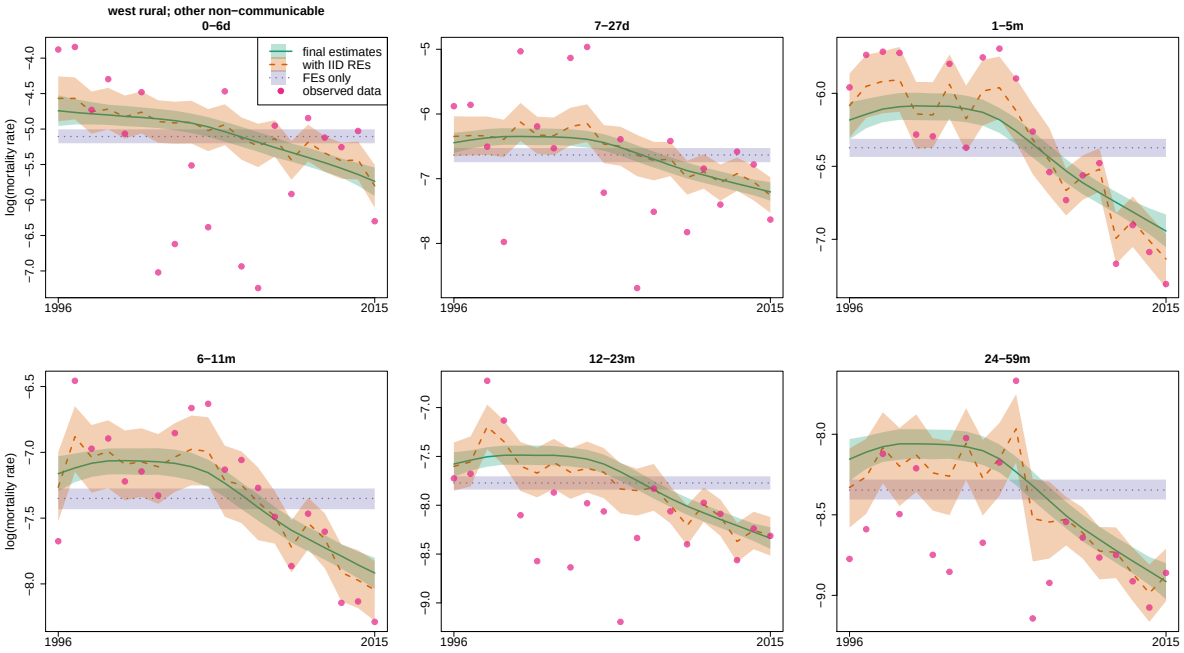

mid urban; diarrhea
0–6d
7–27d
1–5m
6–11m
12–23m
24–59m
final estimates
with IID REs
FEs only
observed data
log(mortality rate)
1996
2015

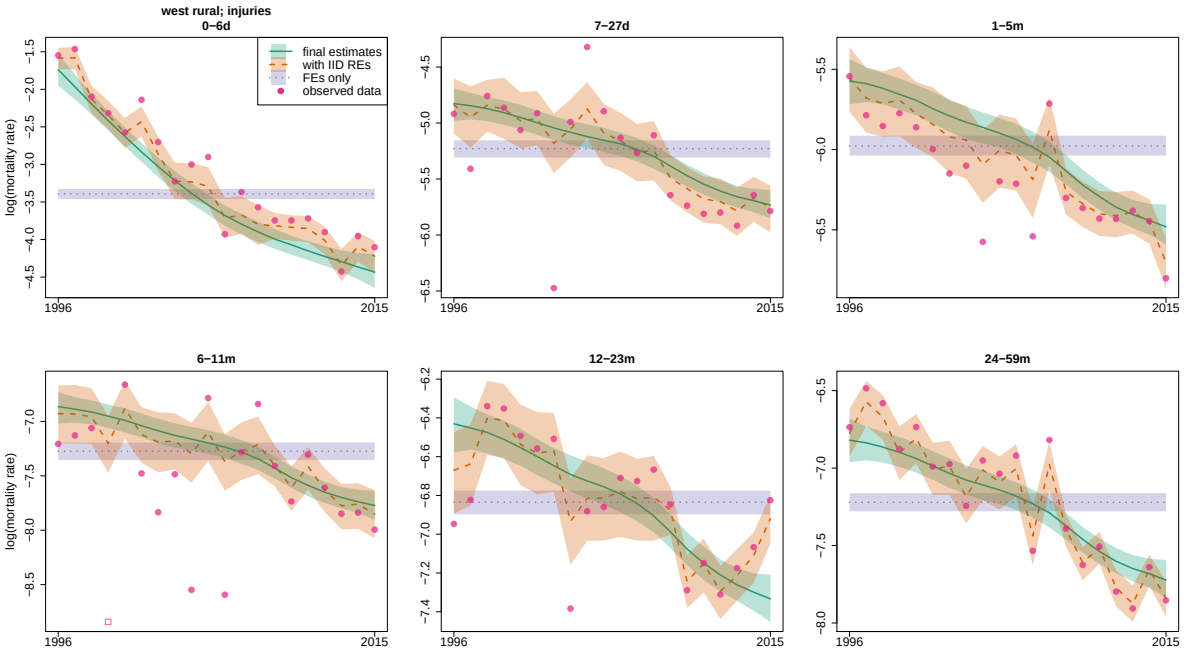

mid urban; acute resp. infections
0–6d
7–27d
1–5m
6–11m
12–23m
24–59m
final estimates
with IID REs
FEs only
observed data
log(mortality rate)
1996
2015

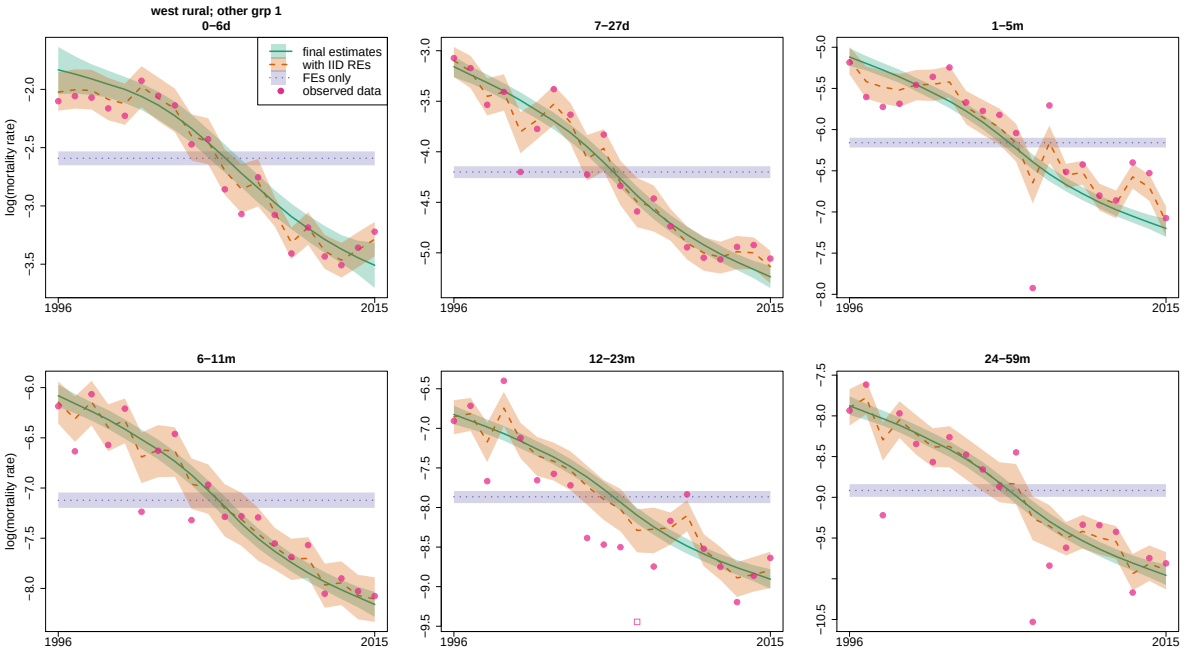

west rural; prematurity
0–6d
7–27d
1–5m
final estimates
with IID REs
FEs only
observed data
log(mortality rate)
−1.0
−1.5
−2.0
−3.5
−4.0
−4.5
−5.0
−6.0
−6.5
−7.0
−7.5
−8.0
−8.5
1996
2015

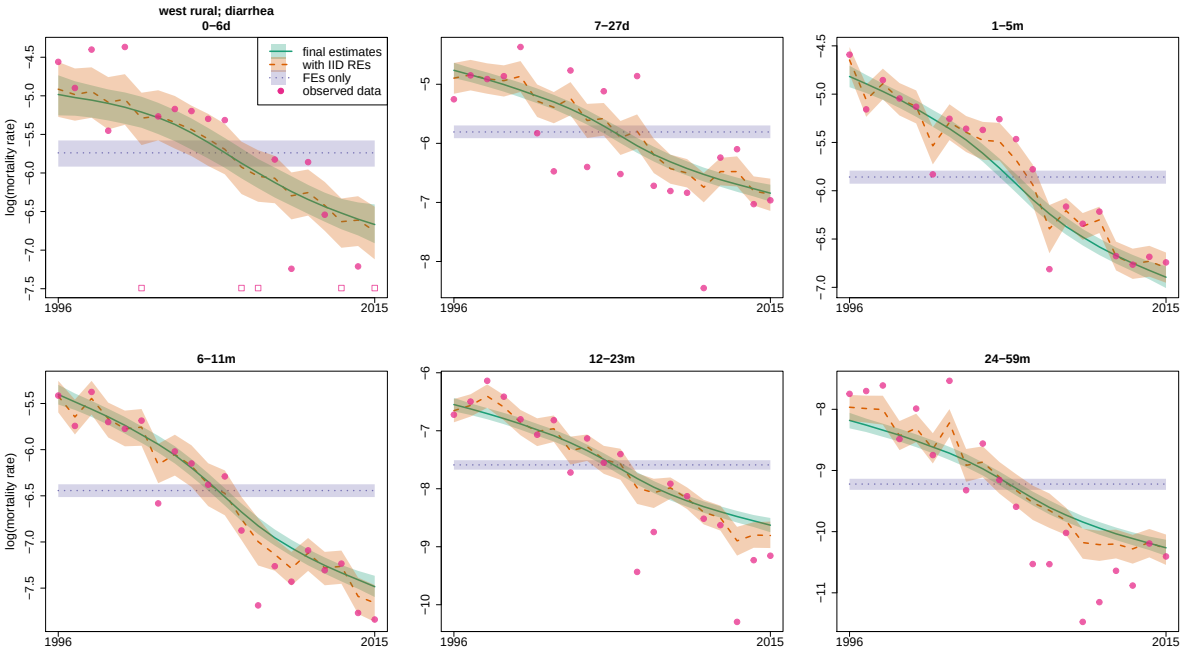

west rural; birth asphyxia/trauma
0–6d
7–27d
1–5m
final estimates
with IID REs
FEs only
observed data
log(mortality rate)
1996
2015

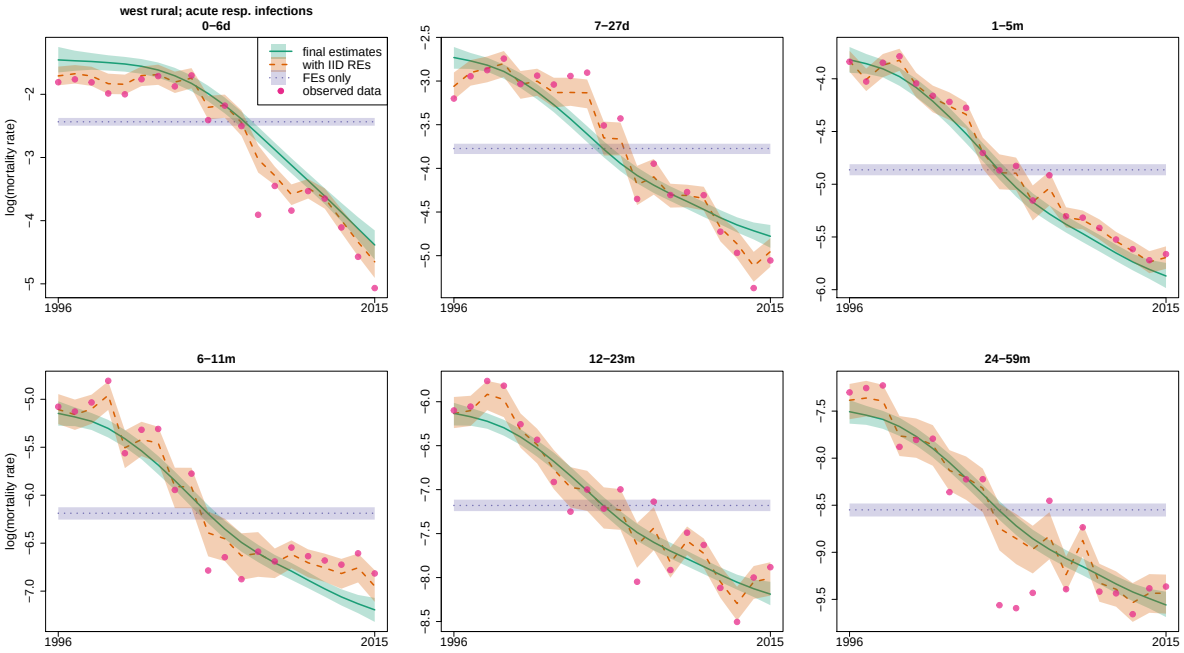

west rural; congenital anomalies
0–6d
7–27d
1–5m
6–11m
12–23m
24–59m
final estimates
with IID REs
FEs only
observed data
log(mortality rate)
1996
2015

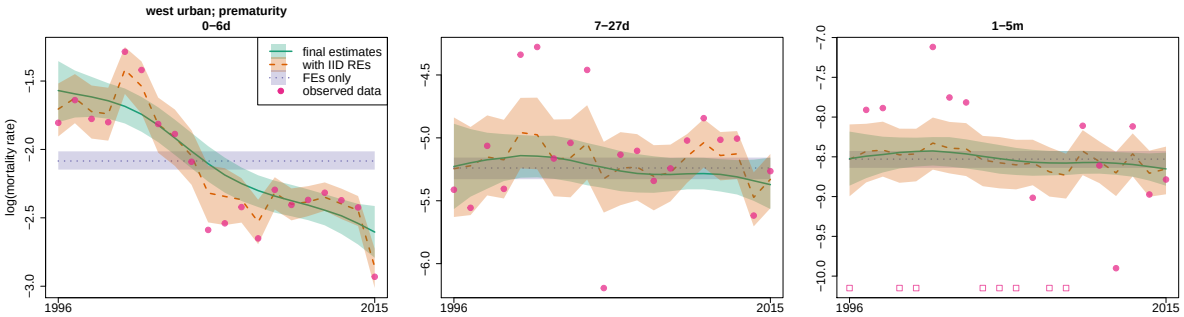

west rural; other non−communicable
0−6d
7−27d
1−5m
6−11m
12−23m
24−59m
final estimates
with IID REs
FEs only
observed data
log(mortality rate)
1996
2015

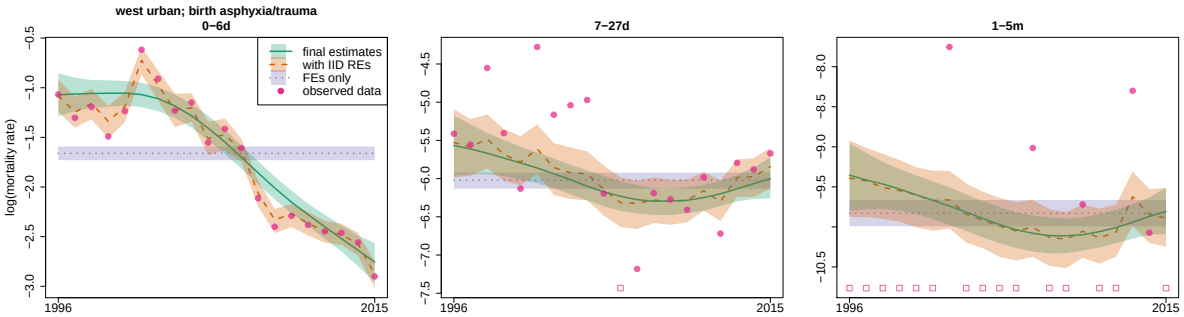

west rural; injuries
0–6d
7–27d
1–5m
6–11m
12–23m
24–59m
final estimates
with IID REs
FEs only
observed data
log(mortality rate)
1996
2015

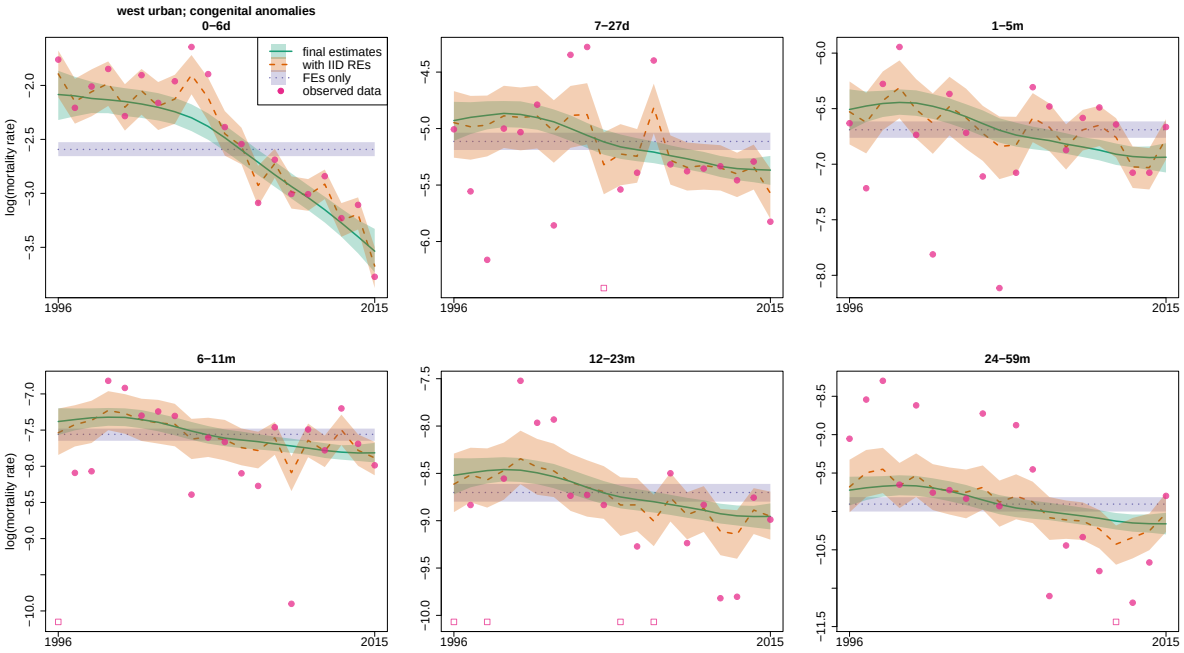

west rural; other grp 1
0–6d
7–27d
1–5m
6–11m
12–23m
24–59m
final estimates
with IID REs
FEs only
observed data
log(mortality rate)
1996
2015

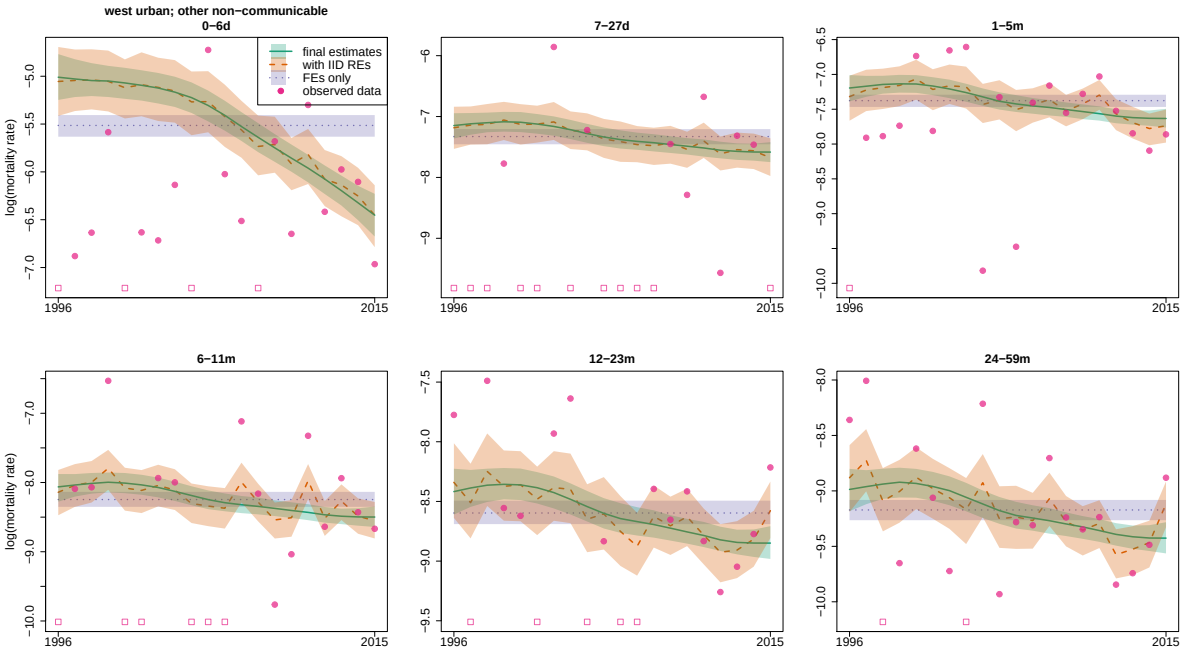

west rural; diarrhea
0–6d
7–27d
1–5m
6–11m
12–23m
24–59m
final estimates
with IID REs
FEs only
observed data
log(mortality rate)
1996
2015

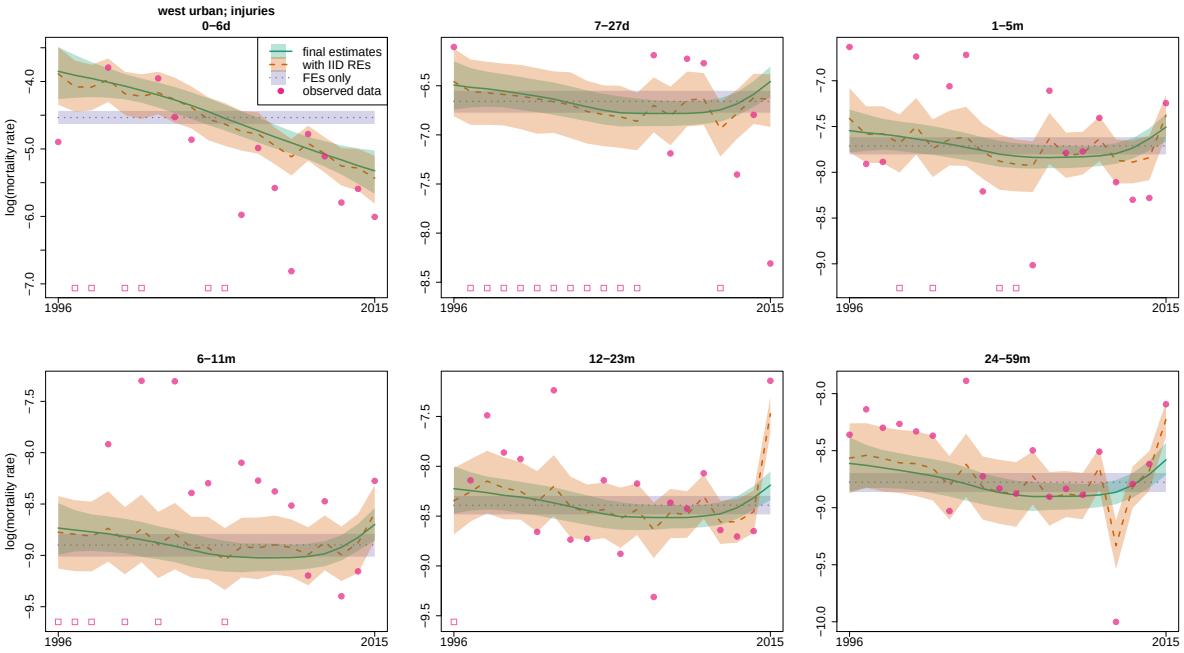

west rural; acute resp. infections
0–6d
7–27d
1–5m
6–11m
12–23m
24–59m
final estimates
with IID REs
FEs only
observed data
log(mortality rate)
1996
2015

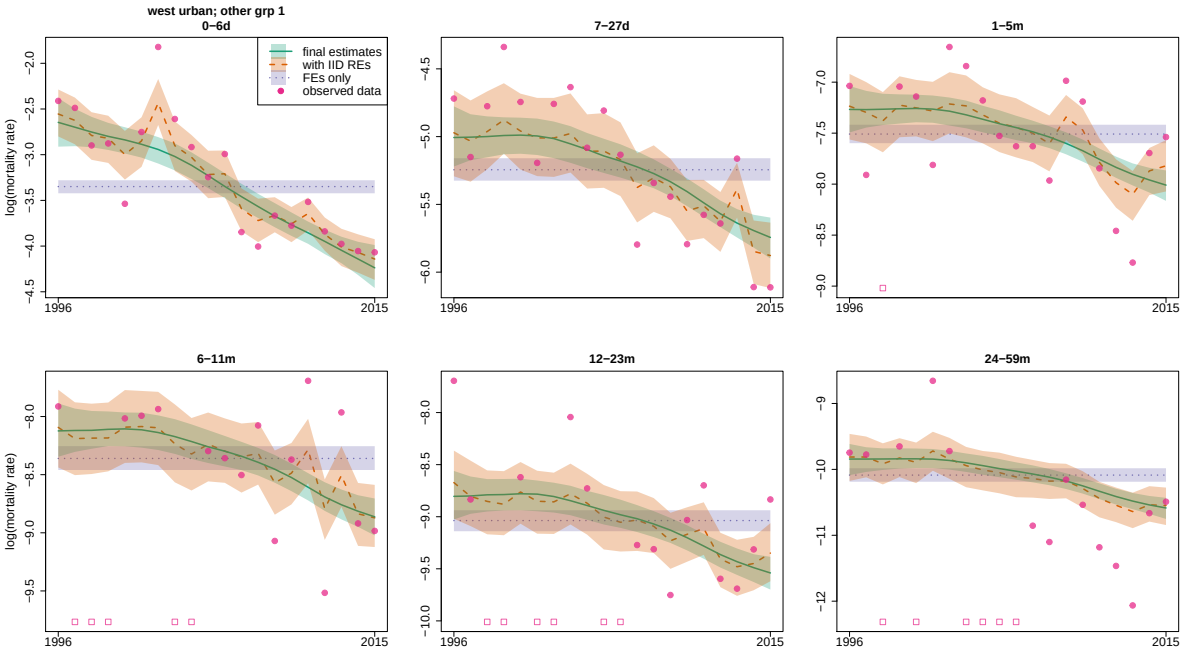

west urban; prematurity
0–6d
7–27d
1–5m
final estimates
with IID REs
FEs only
observed data
log(mortality rate)
1996
2015

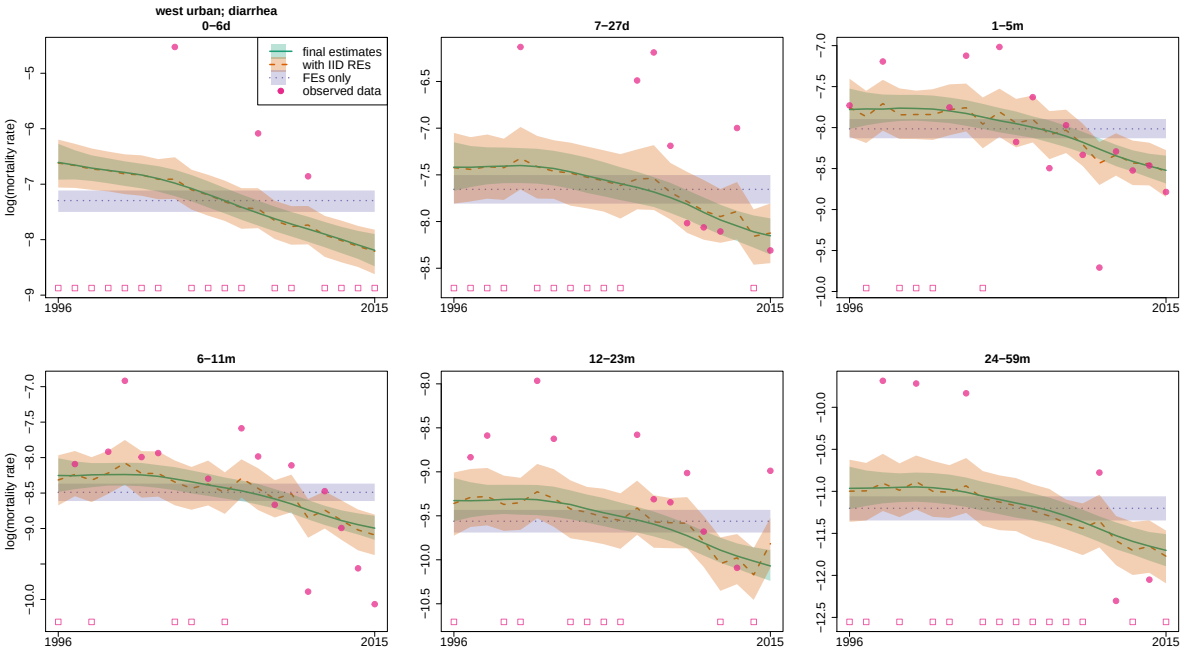

west urban; birth asphyxia/trauma
0–6d
7–27d
1–5m
final estimates
with IID REs
FEs only
observed data
log(mortality rate)
1996
2015

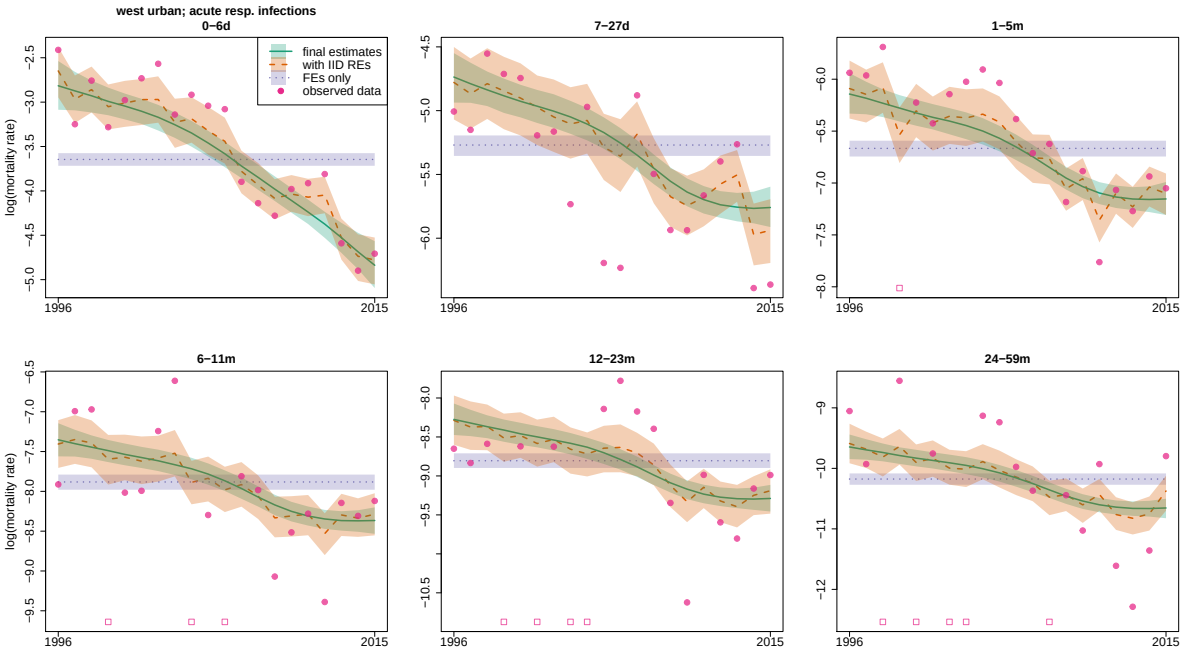

west urban; congenital anomalies
0–6d
7–27d
1–5m
6–11m
12–23m
24–59m
final estimates
with IID REs
FEs only
observed data
log(mortality rate)
1996
2015


west urban; other non−communicable
0−6d
7−27d
1−5m
6−11m
12−23m
24−59m
final estimates
with IID REs
FEs only
observed data
log(mortality rate)
1996
2015


west urban; injuries
0–6d
7–27d
1–5m
6–11m
12–23m
24–59m
final estimates
with IID REs
FEs only
observed data
log(mortality rate)
1996
2015


west urban; other grp 1
0–6d
7–27d
1–5m
6–11m
12–23m
24–59m
final estimates
with IID REs
FEs only
observed data
log(mortality rate)
1996
2015


west urban; diarrhea
0–6d
7–27d
1–5m
6–11m
12–23m
24–59m
final estimates
with IID REs
FEs only
observed data
log(mortality rate)
1996
2015


west urban; acute resp. infections
0–6d
7–27d
1–5m
6–11m
12–23m
24–59m
final estimates
with IID REs
FEs only
observed data
log(mortality rate)
1996
2015

A.12 Plots of cause-specific mortality fractions over time from our final model

Figure A.12 shows estimated CSMFs from the observed data, our final model, and the model from He et al. (2017). Plots are by age group and region. Estimates are aggregated into 0–1 month and 1–59 month age groups because this is the level of modeling from He et al. (2017). We only plot the causes that are comparable between our model and He et al. (2017).

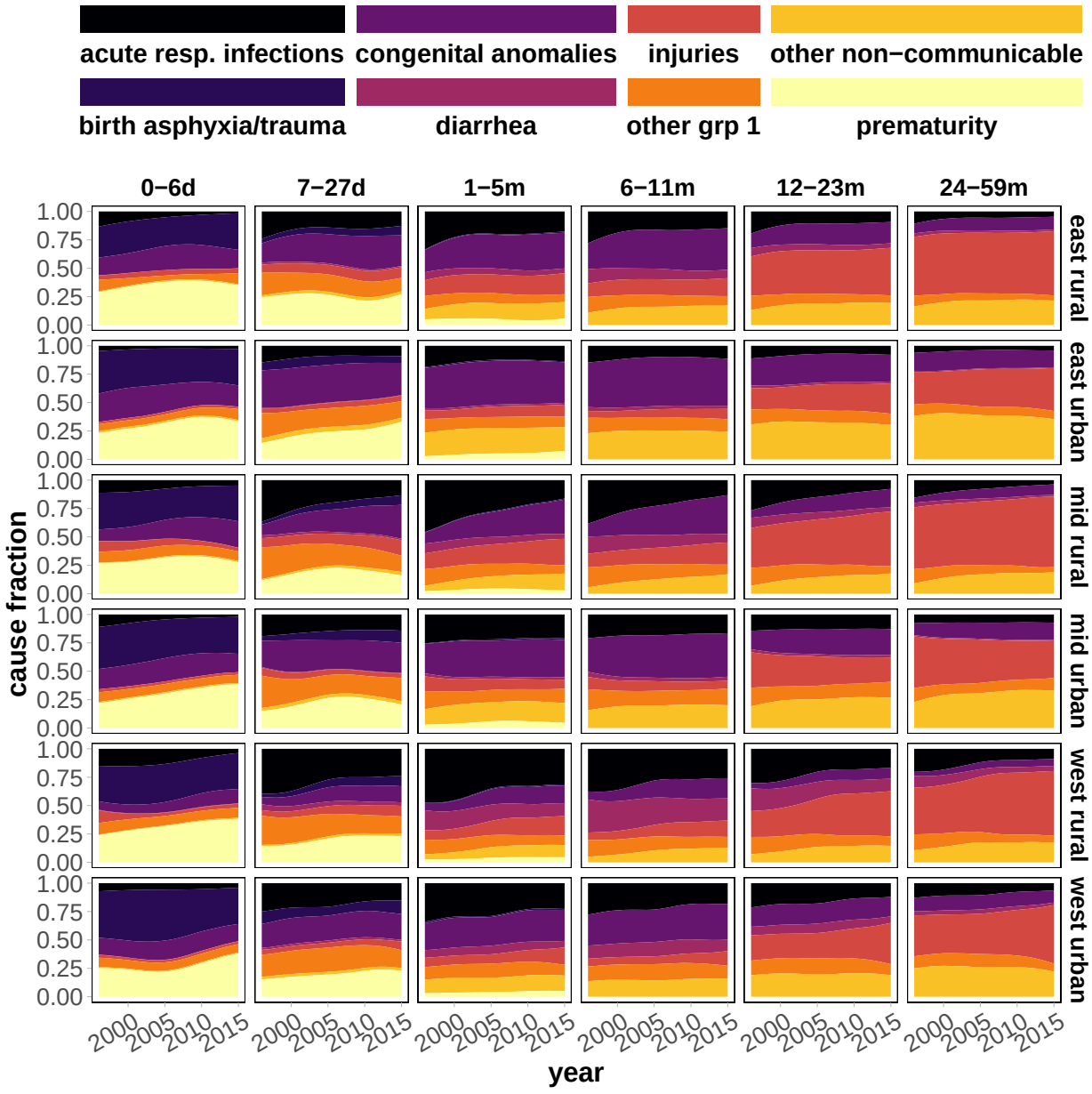


Figure A.12: CSMFs over time by age and region.

A.13 Distributions of residuals from final model fit to MCHSS data by region, age, time, and cause

In this section, we present the distributions of standardized residuals from our final model tabulated by different combinations of region, age, time, and cause. Standardized residuals are calculated as $(y_{r,a,t,c} - N_{r,a,t}\widehat{\lambda}_{r,a,t,c})/(N_{r,a,t}\widehat{\lambda}_{r,a,t,c})^{1/2}$, where $y_{r,a,t,c}$ is the observed number of deaths, $N_{r,a,t}$ is the estimated exposure time, and $\widehat{\lambda}_{r,a,t,c}$ is the estimated mortality rate from our final model for region r, age group a, year t, and cause c. Any patterns in these can help diagnose problems with the fit of the final model.

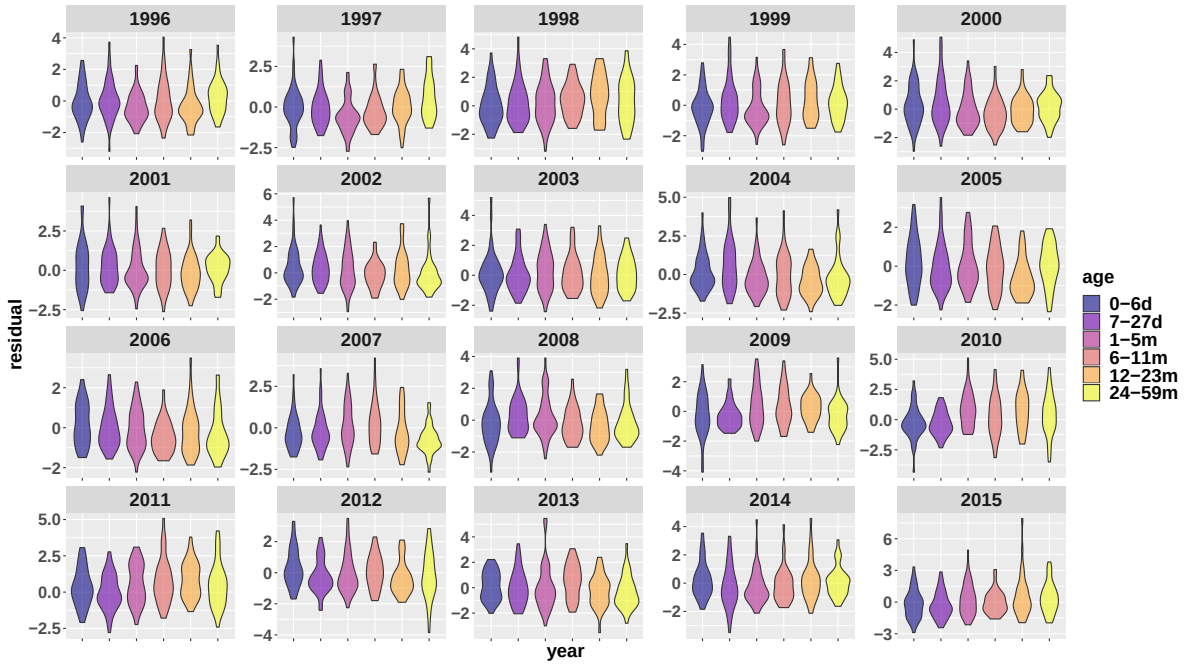


Figure A.13: Distribution of standardized residuals by year and age. Standardized residuals are calculated as $(y_{r,a,t,c} - N_{r,a,t}\widehat{\lambda}_{r,a,t,c})/(N_{r,a,t}\widehat{\lambda}_{r,a,t,c})^{1/2}$.

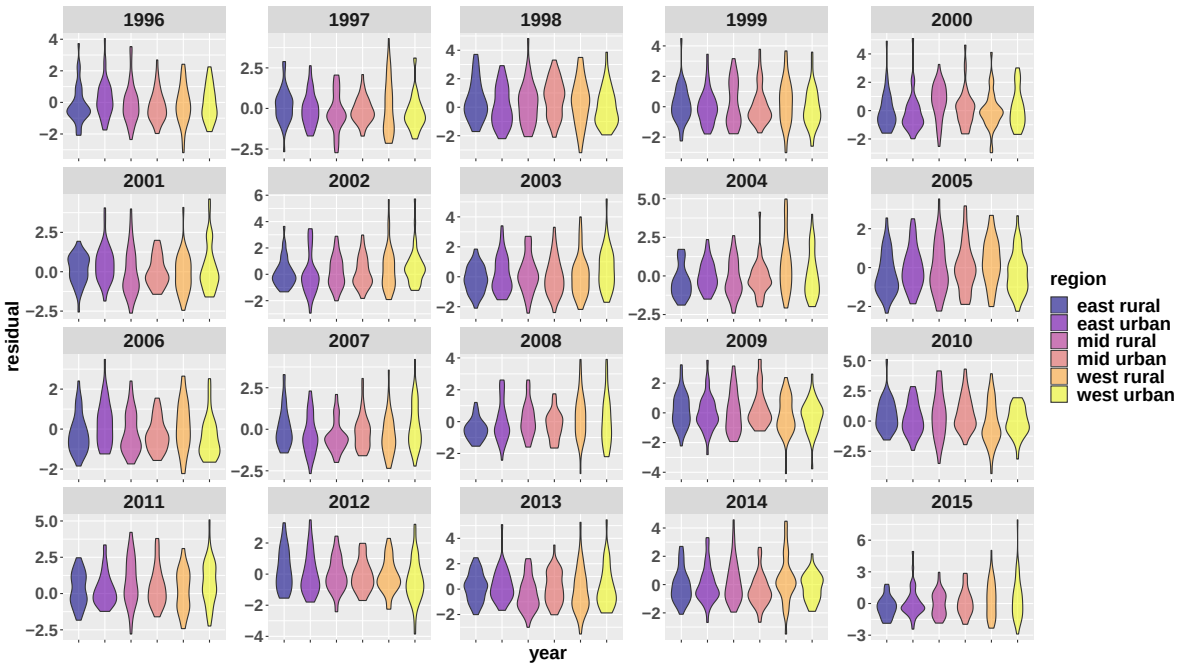


Figure A.14: Distribution of standardized residuals by year and region. Standardized residuals are calculated as $(y_{r,a,t,c} - N_{r,a,t}\widehat{\lambda}_{r,a,t,c})/(N_{r,a,t}\widehat{\lambda}_{r,a,t,c})^{1/2}$.

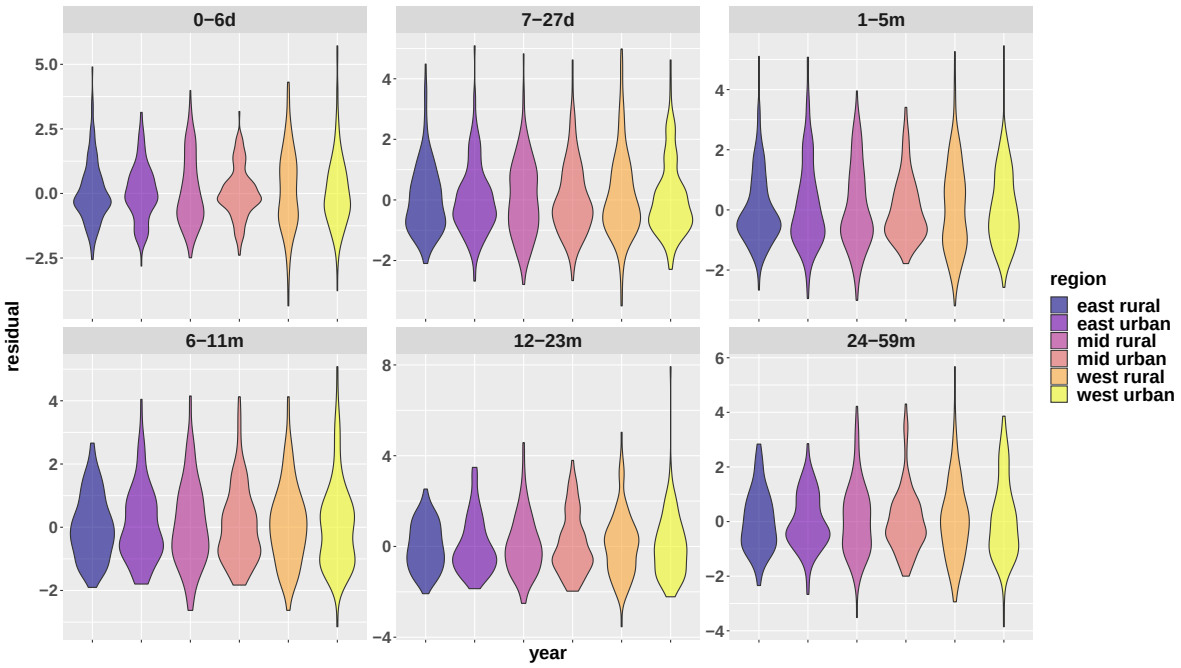


Figure A.15: Distribution of standardized residuals by age and region. Standardized residuals are calculated as $(y_{r,a,t,c} - N_{r,a,t}\widehat{\lambda}_{r,a,t,c})/(N_{r,a,t}\widehat{\lambda}_{r,a,t,c})^{1/2}$.

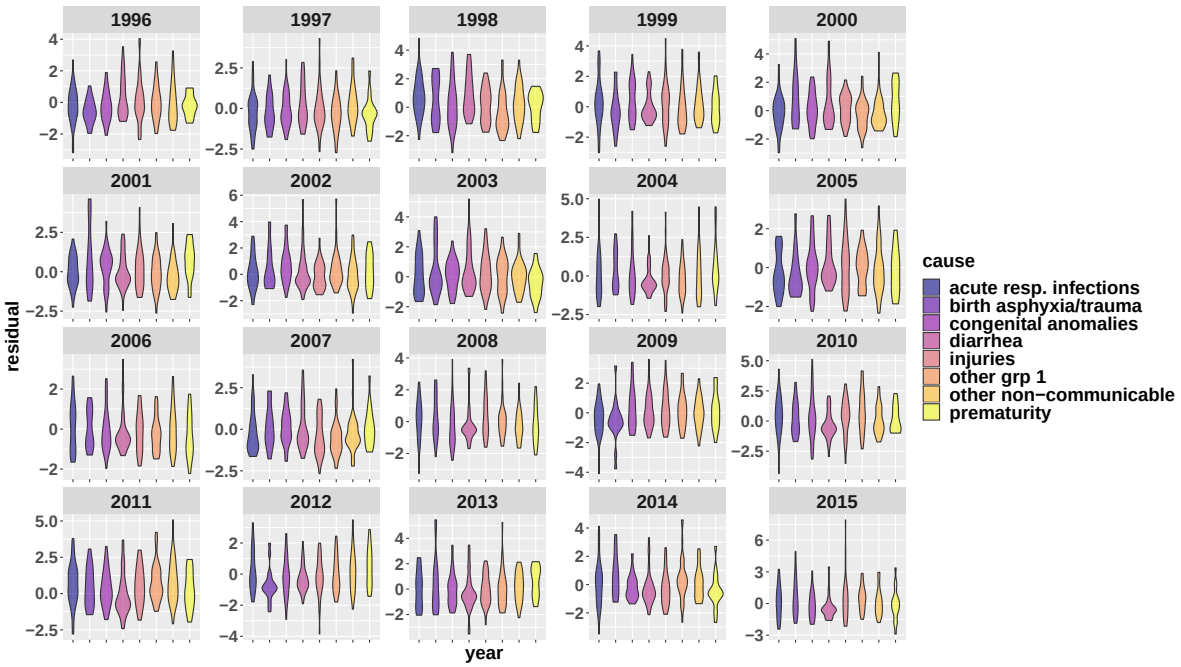


Figure A.16: Distribution of standardized residuals by year and cause. Standardized residuals are calculated as $(y_{r,a,t,c} - N_{r,a,t}\widehat{\lambda}_{r,a,t,c})/(N_{r,a,t}\widehat{\lambda}_{r,a,t,c})^{1/2}$.

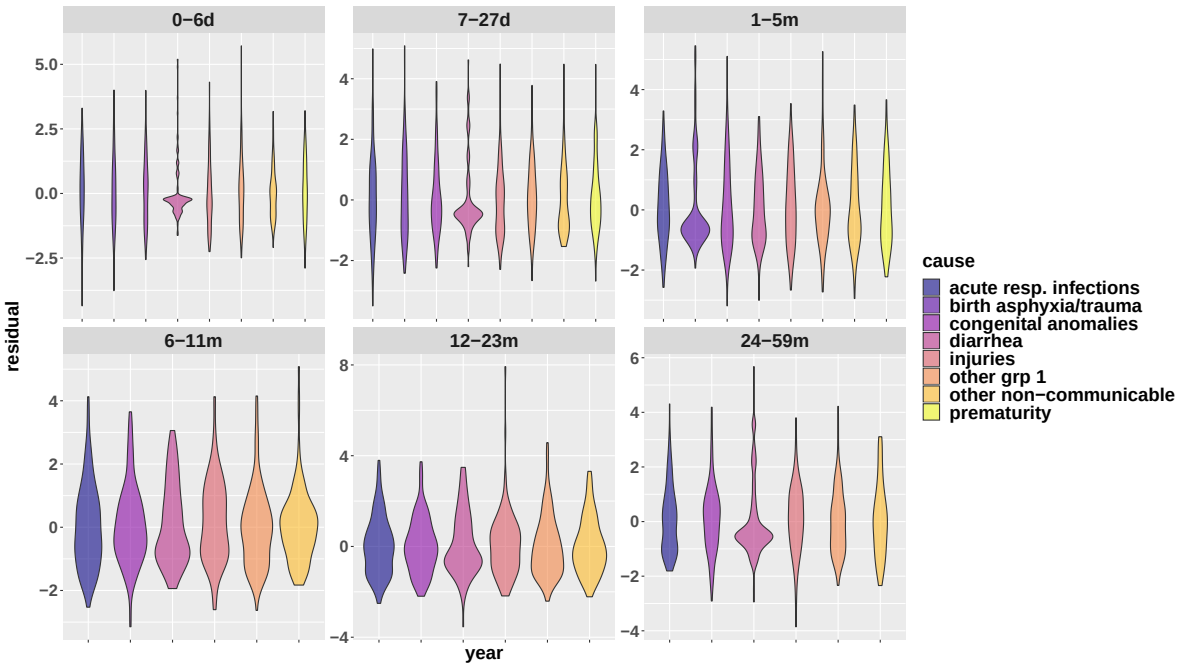


Figure A.17: Distribution of standardized residuals by age and cause. Standardized residuals are calculated as $(y_{r,a,t,c} - N_{r,a,t}\widehat{\lambda}_{r,a,t,c})/(N_{r,a,t}\widehat{\lambda}_{r,a,t,c})^{1/2}$.

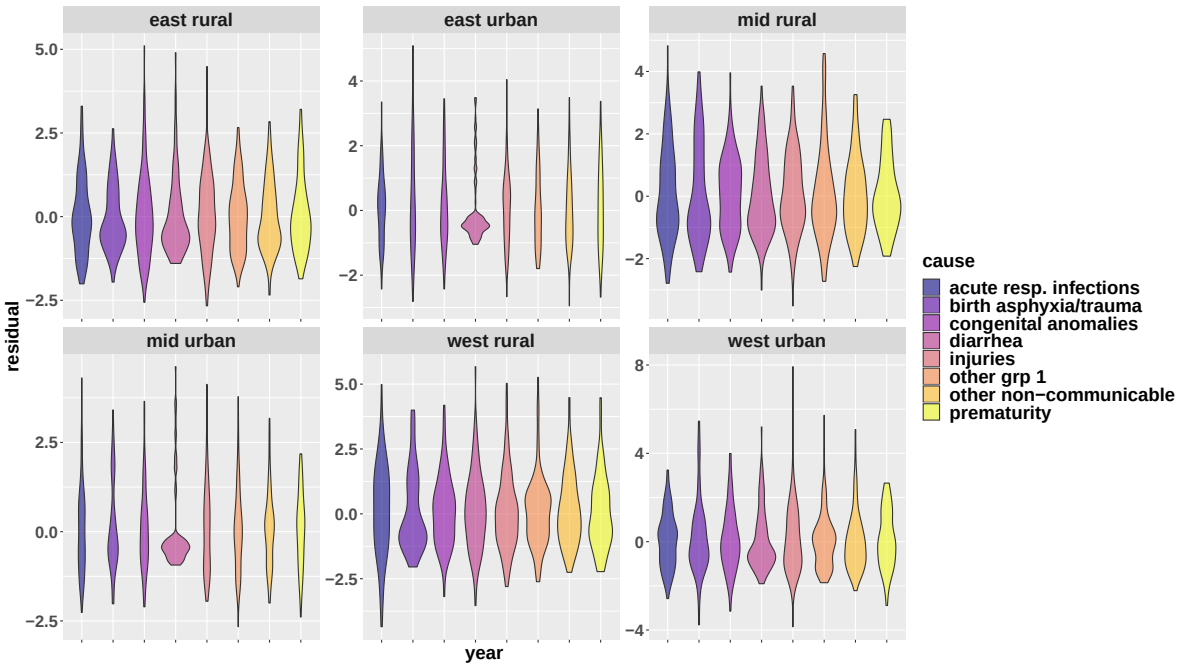


Figure A.18: Distribution of standardized residuals by region and cause. Standardized residuals are calculated as $(y_{r,a,t,c} - N_{r,a,t}\widehat{\lambda}_{r,a,t,c})/(N_{r,a,t}\widehat{\lambda}_{r,a,t,c})^{1/2}$.

Figure A.19: Distribution of standardized residuals over time by age and region. Standardized residuals are calculated as $(y_{r,a,t,c} - N_{r,a,t}\widehat{\lambda}_{r,a,t,c})/(N_{r,a,t}\widehat{\lambda}_{r,a,t,c})^{1/2}$.

Figure A.20: Distribution of standardized residuals over time by age and cause. Standardized residuals are calculated as $(y_{r,a,t,c} - N_{r,a,t}\widehat{\lambda}_{r,a,t,c})/(N_{r,a,t}\widehat{\lambda}_{r,a,t,c})^{1/2}$.

Figure A.21: Distribution of standardized residuals over time by region and cause. Standardized residuals are calculated as $(y_{r,a,t,c} - N_{r,a,t}\widehat{\lambda}_{r,a,t,c})/(N_{r,a,t}\widehat{\lambda}_{r,a,t,c})^{1/2}$.

A.14 Final year holdout comparison of predictions and held-out observations

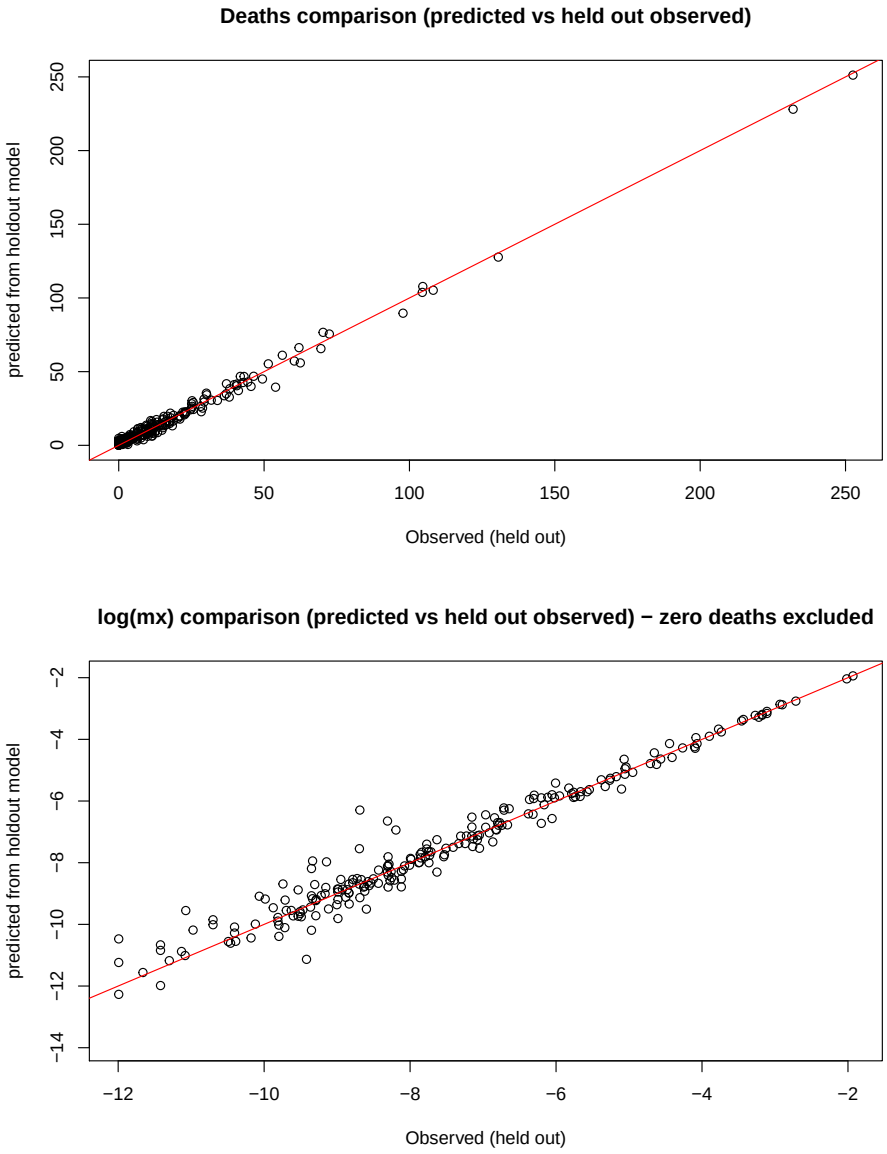


Figure A.22: Scatter plots comparing predictions to held out observations from our final model fit to all data except the final year held out. The top plot shows posterior median predicted deaths, and the bottom plot shows posterior median predicted log mortality rates, excluding observations with zero deaths because they have an undefined log mortality rate.

A.15 Model fit comparison of our final model with two alternate models

Model	DIC	WAIC	$-\sum \log(CPO)$
Final model	**15341.31**	**15379.94**	**8095.364**
No FE interactions	17300.76	16699.79	17672.156
AR1	15361.37	15383.14	8156.577

Table A.2: Comparison of deviance information criteria, Watanabe-Akaike information criterion, and negative sum of log conditional predictive ordinates for our final model, a similar model with no interactions between fixed effects, and a similar model with first order autoregressive processes rather than second order random walks for the temporal trends.

A.16 Simulated data modeling examples of correlations between cause-specific mortality rates

In order to assess whether correlations would be possible to model in the MCHSS data, we performed a simulation study. We simulated data with the same dimensionality as the MCHSS data (6 ages, 6 regions, 8 causes, 20 years). Let $a \in 1, \ldots, A$ index age groups, $r \in 1, \ldots, R$ index regions, $c \in 1, \ldots, C$ index causes, and $t \in 1, \ldots, T$ index years. We generated the data as

$$\begin{aligned} y_{artc} &\sim \text{Poisson}(T_{artc}\lambda_{artc}) \\ \log(\lambda_{artc}) &= \alpha + \sum_{a=2}^{A} \beta_a^A + \sum_{r=2}^{R} \beta_r^R + \sum_{c=2}^{C} \beta_c^C + b_{artc} \\ \mathbf{b}_{art} &= \{b_{art1}, \ldots, b_{artC}\} \\ &\sim \text{Normal}_C(\mathbf{0}, \boldsymbol{\Sigma}). \end{aligned}$$

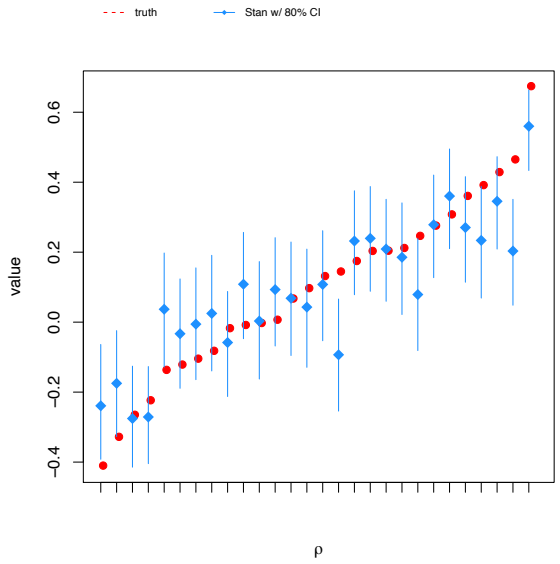
We label the diagonal elements of Σ as $\{\sigma_1^2, \ldots, \sigma_C^2\}$ and the off-diagonal elements as $\rho_{cc'}$ for $c \neq c'$. Thus, the σ_c parameters control the cause-specific overdispersion and the $\rho_{cc'}$ parameters control the correlations between cause-specific mortality rates. We set the α parameter equal to -4, all the β parameters equal to 0.5, all the σ_c parameters equal to 0.1, and the $\rho_{cc'}$ to be between -0.7 and 0.7.

We fit a correctly specified model to this data in Stan with diffuse Normal(0,10) priors on the α and β parameters, half-t(0,10) priors with 3 degrees of freedom on the standard deviations σ_c, and an LKJ(1) prior (Lewandowski, Kurowicka, & Joe, 2009) on the Cholesky decomposition of the correlation matrix. We present the estimated posterior medians of the correlation parameters along with 80% credible intervals in Figure A.23. They are not very well estimated. In general, the lower correlations tend to be estimated to be lower and the higher ones tend to be estimated to be higher.

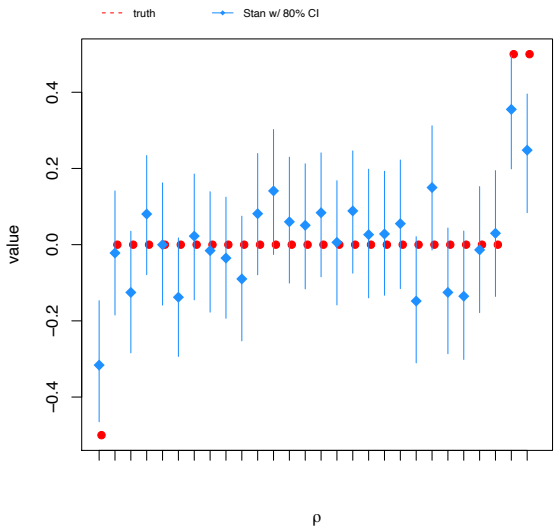
In order to show how this imprecise estimation may be impractical, we generated similar data but we set all the correlation parameters $\rho_{cc'} = 0$ except for two that we set to 0.5 and one that we set to -0.5. This scenario mimics that which may be common in that a few

causes are correlated but the majority are not. We present the estimated posterior medians of the correlation parameters along with 80% credible intervals in Figure A.23. While the positive correlations do have the highest estimates and the negative correlation has the lowest, the differences between these and the estimates for the correlations that are equal to 0 are very minimal. This leads us to conclude that the amount of MCHSS data is not sufficient to estimate correlations reliably.

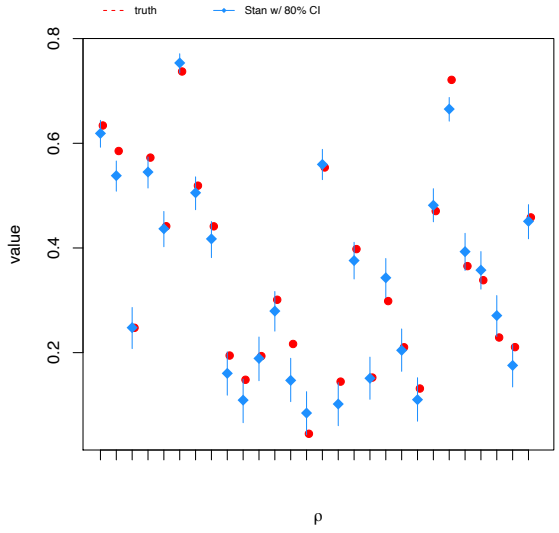
However, given more data, estimating correlation parameters well is possible. To show this, we simulated data similar to the first of the previous simulations except we used 100 regions rather than 6. We present the estimated posterior medians of the correlation parameters along with posterior 80% intervals. The parameters are well-estimated with narrow intervals. Thus, with more data, we would recommend a researcher to model correlations if they believe there are important correlations between causes to consider, for example if there are covariates that impact specific causes for which data are not collected.



(a) MCHSS dimension



(b) MCHSS dimension



(c) 100 regions, 6 ages, 20 years, 8 causes

Figure A.23: Plots of true correlations parameters vs. posterior medians with 80% credible intervals. Scenarios (a) and (b) use data the same dimension as the MCHSS data. Scenario (c) uses 100 regions rather than 6. Scenarios (a) and (c) have a range of true correlation parameters from -0.7 and 0.7. Scenario (b) has all correlation parameters set to 0 except for one set to -0.5 and two set to 0.5.

Appendix B

APPENDIX FOR CHAPTER 3

B.1 Derivation of the design-based covariance estimate for the first stage model

The following calculations are all done within each region, so we drop the r subscripts. Let $\boldsymbol{M} = (M^{HAZ}, M^{WAZ})$ be the overall survey-weighted mean HAZ and WAZ scores, e.g. for HAZ we have $M^{HAZ} = \sum_{i=1}^{n} \left(\frac{w_i y_i^{HAZ}}{\sum_{i=1}^{n} w_i} \right)$. We subtract this from each survey-weighted estimate and then survey-weight these centered estimates, e.g. for HAZ we have

$$\bar{y}_i^{HAZ} = \frac{w_i (y_i^{HAZ} - M^{HAZ})}{\sum_{i=1}^{n} w_i}.$$

These are what the help file means when referring to the "observations".

Next, for each strata, we do the following. First, let $j = 1, \ldots, k_s$ index clusters within a strata with k_s be the number of clusters within strata s, and let $s = 1, \ldots, S$ index strata within a region. Define m_j to be the number of individuals within cluster j. For each strata, we add up the "observations" in each cluster and subtract the strata-specific means, e.g. for HAZ we calculate

$$\tilde{y}_j^{HAZ} = \sum_{i=1}^{m_j} \bar{y}_i^{HAZ},$$

and then subtract the strata-specific means $M_s = \frac{1}{k_s} \sum_{j=1}^{J} \tilde{y}_j^{HAZ}$, such that we have

$$\bar{\tilde{y}}_j^{HAZ} = \tilde{y}_j^{HAZ} - M_s.$$

Then, we take the outer product of these and multiply by a scaling factor of $k_s/(k_s - 1)$ (this is where you'd multiply by the finite population correction, if necessary). In other words, let $\hat{\boldsymbol{X}}_s$ be a $k_s \times 2$ matrix whose first column is the $\bar{\tilde{y}}_j^{HAZ}$ values and second column is the $\bar{\tilde{y}}_j^{WAZ}$ values. We then calculate

$$\tilde{V}_s = \hat{\boldsymbol{X}}_s^T \hat{\boldsymbol{X}}_s.$$

We then element-wise add these strata-specific matrices together and end with our 2x2 covariance matrix estimate.

$$\hat{\boldsymbol{V}}^{DES} = \sum_{s=1}^{S} (\frac{k_s}{(k_s - 1)}) \tilde{V}_s.$$

In detail, we have:

$$\hat{\boldsymbol{V}}^{DES}[1,1] = \sum_{s=1}^{S}(\frac{k_s}{(k_s-1)})\left(\sum_{j=1}^{k_s}\left[\left(\sum_{i=1}^{m_j}\left[\frac{w_i\left(y_i^{HAZ}-\left[\sum_{i=1}^{n}\left(\frac{w_i y_i^{HAZ}}{\sum_{i=1}^{n}w_i}\right)\right]\right)}{\sum_{i=1}^{n}w_i}\right]\right)\right.\right.$$
$$\left.\left.-\left(\frac{1}{k_s}\sum_{j=1}^{J}\left(\sum_{i=1}^{m_j}\left[\frac{w_i\left(y_i^{HAZ}-\sum_{i=1}^{n}\left(\frac{w_i y_i^{HAZ}}{\sum_{i=1}^{n}w_i}\right)\right)}{\sum_{i=1}^{n}w_i}\right]\right)\right)\right]^2\right)$$

$$\hat{\boldsymbol{V}}^{DES}[2,2] = \sum_{s=1}^{S}(\frac{k_s}{(k_s-1)})\left(\sum_{j=1}^{k_s}\left[\left(\sum_{i=1}^{m_j}\left[\frac{w_i\left(y_i^{WAZ}-\left[\sum_{i=1}^{n}\left(\frac{w_i y_i^{WAZ}}{\sum_{i=1}^{n}w_i}\right)\right]\right)}{\sum_{i=1}^{n}w_i}\right]\right)\right.\right.$$
$$\left.\left.-\left(\frac{1}{k_s}\sum_{j=1}^{J}\left(\sum_{i=1}^{m_j}\left[\frac{w_i\left(y_i^{WAZ}-\sum_{i=1}^{n}\left(\frac{w_i y_i^{WAZ}}{\sum_{i=1}^{n}w_i}\right)\right)}{\sum_{i=1}^{n}w_i}\right]\right)\right)\right]^2\right)$$

$$\hat{\boldsymbol{V}}^{DES}[1,2] = \hat{\boldsymbol{V}}^{DES}[2,1] = \sum_{s=1}^{S}(\frac{k_s}{(k_s-1)})\left(\sum_{j=1}^{k_s}\left[\left(\sum_{i=1}^{m_j}\left[\frac{w_i\left(y_i^{HAZ}-\left[\sum_{i=1}^{n}\left(\frac{w_i y_i^{HAZ}}{\sum_{i=1}^{n}w_i}\right)\right]\right)}{\sum_{i=1}^{n}w_i}\right]\right)\right.\right.$$
$$\left.\left.-\left(\frac{1}{k_s}\sum_{j=1}^{J}\left(\sum_{i=1}^{m_j}\left[\frac{w_i\left(y_i^{HAZ}-\sum_{i=1}^{n}\left(\frac{w_i y_i^{HAZ}}{\sum_{i=1}^{n}w_i}\right)\right)}{\sum_{i=1}^{n}w_i}\right]\right)\right)\right]$$
$$\left[\left(\sum_{i=1}^{m_j}\left[\frac{w_i\left(y_i^{WAZ}-\left[\sum_{i=1}^{n}\left(\frac{w_i y_i^{WAZ}}{\sum_{i=1}^{n}w_i}\right)\right]\right)}{\sum_{i=1}^{n}w_i}\right]\right)\right.$$
$$\left.\left.-\left(\frac{1}{k_s}\sum_{j=1}^{J}\left(\sum_{i=1}^{m_j}\left[\frac{w_i\left(y_i^{WAZ}-\sum_{i=1}^{n}\left(\frac{w_i y_i^{WAZ}}{\sum_{i=1}^{n}w_i}\right)\right)}{\sum_{i=1}^{n}w_i}\right]\right)\right)\right]\right)$$

We will calculate these for each region separately and call them $\hat{\boldsymbol{V}}_r^{DES}$.

Due to the CLT, we have

$$(\hat{\bar{y}}_r^{HAZ}, \hat{\bar{y}}_r^{WAZ}) \sim N_2\left((\bar{Y}_r^{HAZ}, \bar{Y}_r^{WAZ}), \hat{\boldsymbol{V}}_r^{DES}\right)$$

These calculations are performed by the `survey` package in R. We conclude by noting what the help file from the `survey` package writes about how the variance is calculated: "The observations for each cluster are added, then centered within each stratum and the outer product is taken of the row vector resulting for each cluster. This is added within strata,

multiplied by a degrees-of-freedom correction and by a finite population correction (if supplied) and added across strata." Note: for our application, we do not use a finite population correction because we assume this is negligible. This could be included if need be.

B.2 Equivalency of nonsymmetric shared component models

Let $r = 1, \ldots, R$ index regions and $c = 1, 2$ index outcomes. Let β_c be fixed effects and z_c be random effects.

Model 1:

$$\mu_{r1} = \beta_1 + z_{r1} + \lambda z_{r2}$$
$$\mu_{r2} = \beta_2 + z_{r2}$$

Model 2:

$$\tilde{\mu}_{r1} = \tilde{\beta}_1 + \tilde{z}_{r1}$$
$$\tilde{\mu}_{r2} = \tilde{\beta}_2 + \tilde{z}_{r2} + \tilde{\lambda}\tilde{z}_{r1}$$

Let $\tilde{\beta}_c = \beta_c$. Now, we want to show that there is some relationship between $(\boldsymbol{z}, \lambda)$ and $(\tilde{\boldsymbol{z}}, \tilde{\lambda})$ that yields $\tilde{\mu}_{rc} = \mu_{rc}$.

Suppose $\lambda\tilde{\lambda} = 1/2$, so $\tilde{\lambda} = 1/2\lambda$. Plugging this in gives

$$\tilde{z}_{r1} = z_{r1} + \lambda z_{r2}$$
$$\tilde{z}_{r1} = \frac{1}{2} z_{r2} - \frac{1}{2\lambda} z_{r1}.$$

Now, substituting into our equations for $\tilde{\mu}_{rc}$, we arrive at

$$\begin{aligned}\tilde{\mu}_{r1} &= \tilde{\beta}_1 + \tilde{z}_{r1} \\ &= \beta_1 + z_{r1} + \lambda z_{r2} \\ &= \mu_{rc}\end{aligned}$$

and

$$
\begin{aligned}
\tilde{\mu}_{r2} &= \tilde{\beta}_2 + \tilde{z}_{r2} + \tilde{\lambda}\tilde{z}_{r1} \\
&= \beta_2 + \frac{1}{2}z_{r2} - \frac{1}{2\lambda}z_{r1} \\
&+ \frac{1}{2\lambda}(z_{r1} + \lambda z_{r2}) \\
&= \beta_2 + \frac{1}{2}z_{r2} - \frac{1}{2\lambda}z_{r1} \\
&+ \frac{1}{2\lambda}z_{r1} + \frac{\lambda}{2\lambda}z_{r2} \\
&= \beta_2 + z_{r2} \\
&= \mu_{r2}.
\end{aligned}
$$

We validated these relationships numerically in the bivariate shared BYM model fit to HAZ and WAZ from the 2014 Kenya DHS, and the relationships held.

B.3 Simulation results for all models in each of the nine scenarios

Table B.1: Area-level simulation results for latent means in Scenario 1

Model	Latent mean	Bias	Rel bias	Var	MSE	80% cov	80% width	95% cov	95% width
O	outcome 1	0.001	0.000	0.012	0.012	0.798	0.264	0.950	0.403
	outcome 2	0.000	0.000	0.009	0.009	0.796	0.230	0.946	0.352
I	outcome 1	-0.002	0.011	0.007	0.009	0.799	0.235	0.950	0.360
	outcome 2	0.002	0.006	0.007	0.008	0.798	0.217	0.948	0.332
II	outcome 1	-0.002	0.011	0.007	0.009	0.797	0.236	0.949	0.362
	outcome 2	0.003	0.006	0.007	0.008	0.798	0.217	0.948	0.332
III	outcome 1	-0.002	0.011	0.007	0.009	0.799	0.235	0.949	0.360
	outcome 2	0.002	0.006	0.007	0.008	0.798	0.217	0.947	0.332
IV	outcome 1	-0.002	0.011	0.007	0.009	0.797	0.236	0.948	0.362
	outcome 2	0.003	0.006	0.007	0.008	0.798	0.217	0.947	0.332
V	outcome 1	-0.002	0.011	0.007	0.009	0.799	0.235	0.949	0.360
	outcome 2	0.002	0.007	0.007	0.008	0.797	0.216	0.946	0.331
VI	outcome 1	-0.002	0.011	0.007	0.009	0.797	0.236	0.949	0.362
	outcome 2	0.003	0.006	0.007	0.008	0.796	0.217	0.947	0.331

Table B.2: Area-level simulation results for latent means in Scenario 2

Model	Latent mean	Bias	Rel bias	Var	MSE	80% cov	80% width	95% cov	95% width
O	outcome 1	0.001	0.000	0.012	0.012	0.798	0.264	0.950	0.403
	outcome 2	0.000	0.000	0.009	0.009	0.796	0.230	0.946	0.352
I	outcome 1	-0.002	0.011	0.007	0.009	0.799	0.235	0.950	0.360
	outcome 2	0.002	0.006	0.007	0.008	0.798	0.217	0.948	0.332
II	outcome 1	-0.002	0.011	0.007	0.009	0.797	0.236	0.949	0.362
	outcome 2	0.003	0.006	0.007	0.008	0.798	0.217	0.948	0.332
III	outcome 1	-0.002	0.011	0.007	0.009	0.799	0.235	0.949	0.360
	outcome 2	0.002	0.006	0.007	0.008	0.798	0.217	0.947	0.332
IV	outcome 1	-0.002	0.011	0.007	0.009	0.797	0.236	0.948	0.362
	outcome 2	0.003	0.006	0.007	0.008	0.798	0.217	0.947	0.332
V	outcome 1	-0.002	0.011	0.007	0.009	0.799	0.235	0.949	0.360
	outcome 2	0.002	0.007	0.007	0.008	0.797	0.216	0.946	0.331
VI	outcome 1	-0.002	0.011	0.007	0.009	0.797	0.236	0.949	0.362
	outcome 2	0.003	0.006	0.007	0.008	0.796	0.217	0.947	0.331

Table B.3: Area-level simulation results for latent means in Scenario 3

Model	Latent mean	Bias	Rel bias	Var	MSE	80% cov	80% width	95% cov	95% width
O	outcome 1	0.001	0.000	0.012	0.012	0.798	0.264	0.950	0.403
	outcome 2	0.000	0.000	0.009	0.009	0.796	0.230	0.946	0.352
I	outcome 1	-0.002	0.011	0.007	0.009	0.799	0.235	0.950	0.360
	outcome 2	0.002	0.006	0.007	0.008	0.798	0.217	0.948	0.332
II	outcome 1	-0.002	0.011	0.007	0.009	0.797	0.236	0.949	0.362
	outcome 2	0.003	0.006	0.007	0.008	0.798	0.217	0.948	0.332
III	outcome 1	-0.002	0.011	0.007	0.009	0.799	0.235	0.949	0.360
	outcome 2	0.002	0.006	0.007	0.008	0.798	0.217	0.947	0.332
IV	outcome 1	-0.002	0.011	0.007	0.009	0.797	0.236	0.948	0.362
	outcome 2	0.003	0.006	0.007	0.008	0.798	0.217	0.947	0.332
V	outcome 1	-0.002	0.011	0.007	0.009	0.799	0.235	0.949	0.360
	outcome 2	0.002	0.007	0.007	0.008	0.797	0.216	0.946	0.331
VI	outcome 1	-0.002	0.011	0.007	0.009	0.797	0.236	0.949	0.362
	outcome 2	0.003	0.006	0.007	0.008	0.796	0.217	0.947	0.331

Table B.4: Area-level simulation results for latent means in Scenario 4

Model	Latent mean	Bias	Rel bias	Var	MSE	80% cov	80% width	95% cov	95% width
O	outcome 1	0.001	0.000	0.012	0.012	0.798	0.264	0.950	0.403
	outcome 2	0.000	0.000	0.009	0.009	0.796	0.230	0.946	0.352
I	outcome 1	-0.002	0.011	0.007	0.009	0.799	0.235	0.950	0.360
	outcome 2	0.002	0.006	0.007	0.008	0.798	0.217	0.948	0.332
II	outcome 1	-0.002	0.011	0.007	0.009	0.797	0.236	0.949	0.362
	outcome 2	0.003	0.006	0.007	0.008	0.798	0.217	0.948	0.332
III	outcome 1	-0.002	0.011	0.007	0.009	0.799	0.235	0.949	0.360
	outcome 2	0.002	0.006	0.007	0.008	0.798	0.217	0.947	0.332
IV	outcome 1	-0.002	0.011	0.007	0.009	0.797	0.236	0.948	0.362
	outcome 2	0.003	0.006	0.007	0.008	0.798	0.217	0.947	0.332
V	outcome 1	-0.002	0.011	0.007	0.009	0.799	0.235	0.949	0.360
	outcome 2	0.002	0.007	0.007	0.008	0.797	0.216	0.946	0.331
VI	outcome 1	-0.002	0.011	0.007	0.009	0.797	0.236	0.949	0.362
	outcome 2	0.003	0.006	0.007	0.008	0.796	0.217	0.947	0.331

Table B.5: Area-level simulation results for latent means in Scenario 5

Model	Latent mean	Bias	Rel bias	Var	MSE	80% cov	80% width	95% cov	95% width
O	outcome 1	0.001	0.000	0.012	0.012	0.798	0.264	0.950	0.403
	outcome 2	0.000	0.000	0.009	0.009	0.796	0.230	0.946	0.352
I	outcome 1	-0.002	0.011	0.007	0.009	0.799	0.235	0.950	0.360
	outcome 2	0.002	0.006	0.007	0.008	0.798	0.217	0.948	0.332
II	outcome 1	-0.002	0.011	0.007	0.009	0.797	0.236	0.949	0.362
	outcome 2	0.003	0.006	0.007	0.008	0.798	0.217	0.948	0.332
III	outcome 1	-0.002	0.011	0.007	0.009	0.799	0.235	0.949	0.360
	outcome 2	0.002	0.006	0.007	0.008	0.798	0.217	0.947	0.332
IV	outcome 1	-0.002	0.011	0.007	0.009	0.797	0.236	0.948	0.362
	outcome 2	0.003	0.006	0.007	0.008	0.798	0.217	0.947	0.332
V	outcome 1	-0.002	0.011	0.007	0.009	0.799	0.235	0.949	0.360
	outcome 2	0.002	0.007	0.007	0.008	0.797	0.216	0.946	0.331
VI	outcome 1	-0.002	0.011	0.007	0.009	0.797	0.236	0.949	0.362
	outcome 2	0.003	0.006	0.007	0.008	0.796	0.217	0.947	0.331

Table B.6: Area-level simulation results for latent means in Scenario 6

Model	Latent mean	Bias	Rel bias	Var	MSE	80% cov	80% width	95% cov	95% width
O	outcome 1	0.001	0.000	0.012	0.012	0.798	0.264	0.950	0.403
	outcome 2	0.000	0.000	0.009	0.009	0.796	0.230	0.946	0.352
I	outcome 1	-0.002	0.011	0.007	0.009	0.799	0.235	0.950	0.360
	outcome 2	0.002	0.006	0.007	0.008	0.798	0.217	0.948	0.332
II	outcome 1	-0.002	0.011	0.007	0.009	0.797	0.236	0.949	0.362
	outcome 2	0.003	0.006	0.007	0.008	0.798	0.217	0.948	0.332
III	outcome 1	-0.002	0.011	0.007	0.009	0.799	0.235	0.949	0.360
	outcome 2	0.002	0.006	0.007	0.008	0.798	0.217	0.947	0.332
IV	outcome 1	-0.002	0.011	0.007	0.009	0.797	0.236	0.948	0.362
	outcome 2	0.003	0.006	0.007	0.008	0.798	0.217	0.947	0.332
V	outcome 1	-0.002	0.011	0.007	0.009	0.799	0.235	0.949	0.360
	outcome 2	0.002	0.007	0.007	0.008	0.797	0.216	0.946	0.331
VI	outcome 1	-0.002	0.011	0.007	0.009	0.797	0.236	0.949	0.362
	outcome 2	0.003	0.006	0.007	0.008	0.796	0.217	0.947	0.331

Table B.7: Area-level simulation results for latent means in Scenario 7

Model	Latent mean	Bias	Rel bias	Var	MSE	80% cov	80% width	95% cov	95% width
O	outcome 1	0.001	0.000	0.012	0.012	0.798	0.264	0.950	0.403
	outcome 2	0.000	0.000	0.009	0.009	0.796	0.230	0.946	0.352
I	outcome 1	-0.002	0.011	0.007	0.009	0.799	0.235	0.950	0.360
	outcome 2	0.002	0.006	0.007	0.008	0.798	0.217	0.948	0.332
II	outcome 1	-0.002	0.011	0.007	0.009	0.797	0.236	0.949	0.362
	outcome 2	0.003	0.006	0.007	0.008	0.798	0.217	0.948	0.332
III	outcome 1	-0.002	0.011	0.007	0.009	0.799	0.235	0.949	0.360
	outcome 2	0.002	0.006	0.007	0.008	0.798	0.217	0.947	0.332
IV	outcome 1	-0.002	0.011	0.007	0.009	0.797	0.236	0.948	0.362
	outcome 2	0.003	0.006	0.007	0.008	0.798	0.217	0.947	0.332
V	outcome 1	-0.002	0.011	0.007	0.009	0.799	0.235	0.949	0.360
	outcome 2	0.002	0.007	0.007	0.008	0.797	0.216	0.946	0.331
VI	outcome 1	-0.002	0.011	0.007	0.009	0.797	0.236	0.949	0.362
	outcome 2	0.003	0.006	0.007	0.008	0.796	0.217	0.947	0.331

Table B.8: Area-level simulation results for latent means in Scenario 8

Model	Latent mean	Bias	Rel bias	Var	MSE	80% cov	80% width	95% cov	95% width
O	outcome 1	0.001	0.000	0.012	0.012	0.798	0.264	0.950	0.403
	outcome 2	0.000	0.000	0.009	0.009	0.796	0.230	0.946	0.352
I	outcome 1	-0.002	0.011	0.007	0.009	0.799	0.235	0.950	0.360
	outcome 2	0.002	0.006	0.007	0.008	0.798	0.217	0.948	0.332
II	outcome 1	-0.002	0.011	0.007	0.009	0.797	0.236	0.949	0.362
	outcome 2	0.003	0.006	0.007	0.008	0.798	0.217	0.948	0.332
III	outcome 1	-0.002	0.011	0.007	0.009	0.799	0.235	0.949	0.360
	outcome 2	0.002	0.006	0.007	0.008	0.798	0.217	0.947	0.332
IV	outcome 1	-0.002	0.011	0.007	0.009	0.797	0.236	0.948	0.362
	outcome 2	0.003	0.006	0.007	0.008	0.798	0.217	0.947	0.332
V	outcome 1	-0.002	0.011	0.007	0.009	0.799	0.235	0.949	0.360
	outcome 2	0.002	0.007	0.007	0.008	0.797	0.216	0.946	0.331
VI	outcome 1	-0.002	0.011	0.007	0.009	0.797	0.236	0.949	0.362
	outcome 2	0.003	0.006	0.007	0.008	0.796	0.217	0.947	0.331

Table B.9: Area-level simulation results for latent means in Scenario 9

Model	Latent mean	Bias	Rel bias	Var	MSE	80% cov	80% width	95% cov	95% width
O	outcome 1	0.001	0.000	0.012	0.012	0.798	0.264	0.950	0.403
	outcome 2	0.000	0.000	0.009	0.009	0.796	0.230	0.946	0.352
I	outcome 1	-0.002	0.011	0.007	0.009	0.799	0.235	0.950	0.360
	outcome 2	0.002	0.006	0.007	0.008	0.798	0.217	0.948	0.332
II	outcome 1	-0.002	0.011	0.007	0.009	0.797	0.236	0.949	0.362
	outcome 2	0.003	0.006	0.007	0.008	0.798	0.217	0.948	0.332
III	outcome 1	-0.002	0.011	0.007	0.009	0.799	0.235	0.949	0.360
	outcome 2	0.002	0.006	0.007	0.008	0.798	0.217	0.947	0.332
IV	outcome 1	-0.002	0.011	0.007	0.009	0.797	0.236	0.948	0.362
	outcome 2	0.003	0.006	0.007	0.008	0.798	0.217	0.947	0.332
V	outcome 1	-0.002	0.011	0.007	0.009	0.799	0.235	0.949	0.360
	outcome 2	0.002	0.007	0.007	0.008	0.797	0.216	0.946	0.331
VI	outcome 1	-0.002	0.011	0.007	0.009	0.797	0.236	0.949	0.362
	outcome 2	0.003	0.006	0.007	0.008	0.796	0.217	0.947	0.331

Appendix C

APPENDIX FOR CHAPTER 4

C.1 Simulation results for all models in each of the seven scenarios

Table C.1: Unit-level simulation results for Scenario 1

Model	Latent mean	Bias	Abs bias	Var	MSE	80% cov	80% width	95% cov	95% width
I	outcome 1	-0.018	0.058	0.008	0.008	0.886	0.221	0.978	0.328
	outcome 2	0.003	0.073	0.006	0.006	0.710	0.192	0.955	0.292
II	outcome 1	-0.014	0.058	0.007	0.008	0.880	0.216	0.967	0.318
	outcome 2	0.005	0.075	0.006	0.006	0.707	0.201	0.934	0.301
III	outcome 1	-0.014	0.060	0.008	0.008	0.880	0.220	0.968	0.332
	outcome 2	0.001	0.073	0.006	0.006	0.675	0.193	0.935	0.284
IV	outcome 1	-0.015	0.059	0.008	0.008	0.914	0.219	0.978	0.327
	outcome 2	0.002	0.074	0.006	0.006	0.691	0.195	0.954	0.295

Table C.2: Unit-level simulation results for Scenario 2

Model	Latent mean	Bias	Abs bias	Var	MSE	80% cov	80% width	95% cov	95% width
I	outcome 1	0.011	0.087	0.008	0.008	0.689	0.219	0.841	0.328
	outcome 2	-0.022	0.068	0.006	0.006	0.708	0.190	0.902	0.288
II	outcome 1	0.014	0.082	0.007	0.007	0.674	0.203	0.837	0.301
	outcome 2	-0.018	0.061	0.006	0.006	0.783	0.190	0.924	0.279
III	outcome 1	0.013	0.091	0.007	0.007	0.620	0.207	0.815	0.316
	outcome 2	-0.019	0.063	0.006	0.006	0.750	0.188	0.902	0.285
IV	outcome 1	0.012	0.086	0.007	0.007	0.665	0.202	0.817	0.303
	outcome 2	-0.020	0.058	0.005	0.006	0.762	0.181	0.871	0.273

Table C.3: Unit-level simulation results for Scenario 3

Model	Latent mean	Bias	Abs bias	Var	MSE	80% cov	80% width	95% cov	95% width
I	outcome 1	-0.001	0.076	0.007	0.007	0.734	0.212	0.893	0.318
	outcome 2	0.009	0.062	0.007	0.007	0.825	0.203	0.978	0.305
II	outcome 1	-0.005	0.074	0.007	0.007	0.749	0.210	0.901	0.315
	outcome 2	0.012	0.063	0.006	0.006	0.772	0.194	0.956	0.295
III	outcome 1	-0.001	0.071	0.007	0.007	0.782	0.209	0.902	0.314
	outcome 2	0.013	0.062	0.006	0.006	0.805	0.197	0.968	0.301
IV	outcome 1	-0.004	0.073	0.007	0.007	0.758	0.207	0.902	0.316
	outcome 2	0.015	0.065	0.006	0.006	0.774	0.197	0.958	0.295

Table C.4: Unit-level simulation results for Scenario 4

Model	Latent mean	Bias	Abs bias	Var	MSE	80% cov	80% width	95% cov	95% width
I	outcome 1	-0.012	0.066	0.006	0.006	0.724	0.189	0.902	0.287
	outcome 2	0.001	0.072	0.006	0.006	0.748	0.190	0.952	0.285
II	outcome 1	-0.016	0.070	0.005	0.006	0.664	0.184	0.868	0.274
	outcome 2	0.002	0.070	0.005	0.005	0.675	0.180	0.893	0.272
III	outcome 1	-0.013	0.060	0.005	0.005	0.761	0.180	0.880	0.273
	outcome 2	-0.001	0.067	0.005	0.005	0.674	0.176	0.902	0.263
IV	outcome 1	-0.014	0.062	0.004	0.005	0.708	0.166	0.836	0.248
	outcome 2	0.005	0.066	0.005	0.005	0.728	0.185	0.902	0.273

Table C.5: Unit-level simulation results for Scenario 5

Model	Latent mean	Bias	Abs bias	Var	MSE	80% cov	80% width	95% cov	95% width
I	outcome 1	0.014	0.083	0.009	0.009	0.702	0.233	0.899	0.354
	outcome 2	0.004	0.073	0.010	0.010	0.858	0.250	0.951	0.370
II	outcome 1	0.013	0.078	0.009	0.010	0.761	0.244	0.924	0.362
	outcome 2	0.001	0.075	0.009	0.009	0.793	0.243	0.971	0.355
III	outcome 1	0.019	0.065	0.008	0.008	0.761	0.222	0.924	0.337
	outcome 2	0.000	0.073	0.008	0.008	0.815	0.230	0.924	0.347
IV	outcome 1	0.015	0.066	0.008	0.008	0.750	0.216	0.902	0.323
	outcome 2	-0.005	0.075	0.007	0.007	0.774	0.206	0.914	0.310

Table C.6: Unit-level simulation results for Scenario 6

Model	Latent mean	Bias	Abs bias	Var	MSE	80% cov	80% width	95% cov	95% width
I	outcome 1	0.000	0.039	0.003	0.003	0.902	0.147	0.980	0.223
	outcome 2	0.003	0.043	0.003	0.003	0.800	0.142	0.935	0.207
II	outcome 1	0.000	0.039	0.004	0.004	0.912	0.151	0.978	0.224
	outcome 2	0.005	0.042	0.003	0.003	0.825	0.139	0.903	0.211
III	outcome 1	0.000	0.038	0.004	0.004	0.859	0.149	0.957	0.221
	outcome 2	0.006	0.042	0.003	0.003	0.761	0.136	0.924	0.205
IV	outcome 1	0.000	0.039	0.003	0.003	0.882	0.148	0.967	0.220
	outcome 2	0.007	0.041	0.003	0.003	0.761	0.140	0.934	0.205

Table C.7: Unit-level simulation results for Scenario 7

Model	Latent mean	Bias	Abs bias	Var	MSE	80% cov	80% width	95% cov	95% width
I	outcome 1	0.016	0.080	0.008	0.009	0.773	0.233	0.892	0.342
	outcome 2	-0.005	0.054	0.004	0.004	0.800	0.167	0.920	0.248
II	outcome 1	0.012	0.076	0.008	0.008	0.793	0.228	0.935	0.342
	outcome 2	-0.003	0.054	0.004	0.004	0.794	0.166	0.956	0.249
III	outcome 1	0.014	0.069	0.008	0.008	0.783	0.225	0.935	0.337
	outcome 2	-0.010	0.047	0.004	0.004	0.784	0.165	0.967	0.246
IV	outcome 1	0.014	0.065	0.007	0.007	0.815	0.205	0.913	0.305
	outcome 2	-0.007	0.047	0.004	0.004	0.815	0.159	0.924	0.237

Table C.8: Simulation results for Scenario 2 of the unit-level modeling simulation

Model	Latent mean	Bias	Abs bias	Var	MSE	80% cov	80% width	95% cov	95% width
I	outcome 1	0.011	0.087	0.008	0.008	0.689	0.219	0.841	0.328
	outcome 2	-0.022	0.068	0.006	0.006	0.708	0.190	0.902	0.288
II	outcome 1	0.014	0.082	0.007	0.007	0.674	0.203	0.837	0.301
	outcome 2	-0.018	0.061	0.006	0.006	0.783	0.190	0.924	0.279
III	outcome 1	0.013	0.091	0.007	0.007	0.620	0.207	0.815	0.316
	outcome 2	-0.019	0.063	0.006	0.006	0.750	0.188	0.902	0.285
IV	outcome 1	0.012	0.086	0.007	0.007	0.665	0.202	0.817	0.303
	outcome 2	-0.020	0.058	0.005	0.006	0.762	0.181	0.871	0.273

Table C.9: Simulation results for Scenario 3 of the unit-level modeling simulation

Model	Latent mean	Bias	Abs bias	Var	MSE	80% cov	80% width	95% cov	95% width
I	outcome 1	-0.001	0.076	0.007	0.007	0.734	0.212	0.893	0.318
	outcome 2	0.009	0.062	0.007	0.007	0.825	0.203	0.978	0.305
II	outcome 1	-0.005	0.074	0.007	0.007	0.749	0.210	0.901	0.315
	outcome 2	0.012	0.063	0.006	0.006	0.772	0.194	0.956	0.295
III	outcome 1	-0.001	0.071	0.007	0.007	0.782	0.209	0.902	0.314
	outcome 2	0.013	0.062	0.006	0.006	0.805	0.197	0.968	0.301
IV	outcome 1	-0.004	0.073	0.007	0.007	0.758	0.207	0.902	0.316
	outcome 2	0.015	0.065	0.006	0.006	0.774	0.197	0.958	0.295

Table C.10: Simulation results for Scenario 4 of the unit-level modeling simulation

Model	Latent mean	Bias	Abs bias	Var	MSE	80% cov	80% width	95% cov	95% width
I	outcome 1	-0.012	0.066	0.006	0.006	0.724	0.189	0.902	0.287
	outcome 2	0.001	0.072	0.006	0.006	0.748	0.190	0.952	0.285
II	outcome 1	-0.016	0.070	0.005	0.006	0.664	0.184	0.868	0.274
	outcome 2	0.002	0.070	0.005	0.005	0.675	0.180	0.893	0.272
III	outcome 1	-0.013	0.060	0.005	0.005	0.761	0.180	0.880	0.273
	outcome 2	-0.001	0.067	0.005	0.005	0.674	0.176	0.902	0.263
IV	outcome 1	-0.014	0.062	0.004	0.005	0.708	0.166	0.836	0.248
	outcome 2	0.005	0.066	0.005	0.005	0.728	0.185	0.902	0.273

Table C.11: Simulation results for Scenario 5 of the unit-level modeling simulation

Model	Latent mean	Bias	Abs bias	Var	MSE	80% cov	80% width	95% cov	95% width
I	outcome 1	0.014	0.083	0.009	0.009	0.702	0.233	0.899	0.354
	outcome 2	0.004	0.073	0.010	0.010	0.858	0.250	0.951	0.370
II	outcome 1	0.013	0.078	0.009	0.010	0.761	0.244	0.924	0.362
	outcome 2	0.001	0.075	0.009	0.009	0.793	0.243	0.971	0.355
III	outcome 1	0.019	0.065	0.008	0.008	0.761	0.222	0.924	0.337
	outcome 2	0.000	0.073	0.008	0.008	0.815	0.230	0.924	0.347
IV	outcome 1	0.015	0.066	0.008	0.008	0.750	0.216	0.902	0.323
	outcome 2	-0.005	0.075	0.007	0.007	0.774	0.206	0.914	0.310

Table C.12: Simulation results for Scenario 6 of the unit-level modeling simulation

Model	Latent mean	Bias	Abs bias	Var	MSE	80% cov	80% width	95% cov	95% width
I	outcome 1	0.000	0.039	0.003	0.003	0.902	0.147	0.980	0.223
	outcome 2	0.003	0.043	0.003	0.003	0.800	0.142	0.935	0.207
II	outcome 1	0.000	0.039	0.004	0.004	0.912	0.151	0.978	0.224
	outcome 2	0.005	0.042	0.003	0.003	0.825	0.139	0.903	0.211
III	outcome 1	0.000	0.038	0.004	0.004	0.859	0.149	0.957	0.221
	outcome 2	0.006	0.042	0.003	0.003	0.761	0.136	0.924	0.205
IV	outcome 1	0.000	0.039	0.003	0.003	0.882	0.148	0.967	0.220
	outcome 2	0.007	0.041	0.003	0.003	0.761	0.140	0.934	0.205

Table C.13: Simulation results for Scenario 7 of the unit-level modeling simulation

Model	Latent mean	Bias	Abs bias	Var	MSE	80% cov	80% width	95% cov	95% width
I	outcome 1	0.016	0.080	0.008	0.009	0.773	0.233	0.892	0.342
	outcome 2	-0.005	0.054	0.004	0.004	0.800	0.167	0.920	0.248
II	outcome 1	0.012	0.076	0.008	0.008	0.793	0.228	0.935	0.342
	outcome 2	-0.003	0.054	0.004	0.004	0.794	0.166	0.956	0.249
III	outcome 1	0.014	0.069	0.008	0.008	0.783	0.225	0.935	0.337
	outcome 2	-0.010	0.047	0.004	0.004	0.784	0.165	0.967	0.246
IV	outcome 1	0.014	0.065	0.007	0.007	0.815	0.205	0.913	0.305
	outcome 2	-0.007	0.047	0.004	0.004	0.815	0.159	0.924	0.237

C.2 Equivalence of maximum likelihood estimates from a multinomial model and a multinomial-Poisson surrogate model (i.e. the "Poisson Trick"

Let $k = 1, \ldots, K$ index clusters and $c = 1 \ldots, C$ index multinomial categories. For cluster k, define N_k as the total number of individuals, Y_{kc} as the number of individuals that are in category c (in vector form $\boldsymbol{Y}_k = [Y_{k1}, \ldots, Y_{kC}]'$), with p_{kc} the associated probability of being in category c (in vector form $\boldsymbol{p}_k = [p_{k1}, \ldots, p_{kC}]'$).

First, our cluster-level likelihood is

$$\boldsymbol{Y}_k | N_k, \boldsymbol{p}_k \sim \text{Multinomial}(N_k, \boldsymbol{p}_k). \tag{C.1}$$

Suppose $p_{kc} = \frac{\zeta_{kc}}{\zeta_{k+}}$ where $\zeta_{k+} = \sum_{c=1}^{C} \zeta_{kc}$ (note that this holds in the baseline category multinomial logistic model by letting $\zeta_{kc} = \exp(\mu_{kc})$ and setting $\mu_{kC} = 0$). Now, in the multinomial likelihood in Equation (4.1), we treated N_k as fixed. Instead, suppose $N_k \sim \delta_k \zeta_{k+}$ independently for each cluster. Thus, we now have a multinomial-Poisson mixture with joint probability function

$$\begin{aligned} P(\boldsymbol{Y}_k = \boldsymbol{y}_k \text{ and } N_k = n_k) &= P(N_k = n_k) P(\boldsymbol{Y}_k = \boldsymbol{y}_k | N_k = n_k) \\ &= \exp(-\delta_k \zeta_{k+}) \frac{(\delta_k \zeta_{k+})^{n_k}}{n_k!} \times \frac{n_k!}{\prod_{c=1}^{C} y_{kc}!} \prod_{c=1}^{C} \left(\frac{\zeta_{kc}}{\zeta_{k+}} \right)^{y_{kc}} \\ &= \prod_{c=1}^{C} \left[\frac{\exp(-\delta_k \zeta_{kc})(\delta_k \zeta_{kc})^{y_{kc}}}{y_{kc}!} \right] \text{ iff } N_k = n_k \end{aligned}$$

We can integrate over all possible values of N_k which gives the marginal probability

$$P(\boldsymbol{Y}_k = \boldsymbol{y}_k) = \prod \left[\frac{\exp(-\delta_k \zeta_{kc})(\delta_k \zeta_{kc})^{y_{kc}}}{y_{kc}!} \right]$$

which is the product of independent Poisson distributions, i.e. $Y_{kc} \sim \text{Poisson}(\delta_k \zeta_{kc})$ independently for eack k and c. Combining all observations gives the loglikelihood

$$\sum_{k=1}^{K} \ell^P(\delta_k \zeta_{k+}; N_k) + \sum_{k=1}^{K} \ell^M(\boldsymbol{\zeta}_k; \boldsymbol{Y}_k | N_k) = \sum_{k=1}^{K} \sum_{c=1}^{C} \ell^P(\delta_k \zeta_{kc}; Y_{kc})$$

with ℓ^P the Poisson and ℓ^M the multinomial loglikelihood functions.

Now, we wish to fit a model for $\sum_{k=1}^{K} \ell^M(\boldsymbol{\zeta}_k; \boldsymbol{Y}_k | N_k)$. However, we can instead fit a model for $\sum_{k=1}^{K} \sum_{c=1}^{C} \ell^P(\delta_k \zeta_{kc}; Y_{kc})$. To see this, first note that the multinomial loglikelihood function can be written as

$$\sum_{k=1}^{K} \ell^M(\boldsymbol{\zeta}_k; \boldsymbol{Y}_k | N_k) = \sum_{k=1}^{K} \log(n_k!) - \sum_{k=1}^{K} \sum_{c=1}^{C} \log(y_{kc}!) + \sum_{k=1}^{K} \sum_{c=1}^{C} y_{kc} \log(\zeta_{kc}) - \sum_{k=1}^{K} \log(\zeta_{k+})$$

and the Poisson-multinomial loglikelihood function can be written as

$$\sum_{k=1}^{K} \sum_{c=1}^{C} \ell^P(\delta_k \zeta_{kc}; Y_{kc}) = -\sum_{k=1}^{K} \delta_k \zeta_{k+} + \sum_{k=1}^{K} n_k \log(\delta_k) + \sum_{k=1}^{K} \sum_{c=1}^{C} y_{kc} \log(\zeta_{k+}) - \sum_{k=1}^{K} \log(n_k!).$$

Now, the MLE of δ_k from the Poisson-multinomial loglikelihood is $\hat{\delta}_k = n_k / \zeta_{k+}$, and if we plug this into the Poisson-multinomial loglikelihood function, we get the same loglikelihood as the multinomial model up to an additive constant. Thus, likelihood-based inference will be identical between the multinomial model for the vector of counts and independent Poisson models on each of the individual counts.

REFERENCES

Abdullah, S., Adazu, K., Masanja, H., Diallo, D., Hodgson, A., Ilboudo-Sanogo, E., et al.others. (2007). Patterns of age-specific mortality in children in endemic areas of sub-saharan africa. *The American Journal of Tropical Medicine and Hygiene*, *77*(6 Supp), 99–105.

AbouZahr, C., De Savigny, D., Mikkelsen, L., Setel, P. W., Lozano, R., & Lopez, A. D. (2015). Towards universal civil registration and vital statistics systems: The time is now. *The Lancet*, *386*(10001), 1407–1418.

AbouZahr, C., De Savigny, D., Mikkelsen, L., Setel, P. W., Lozano, R., Nichols, E., … Lopez, A. D. (2015). Civil registration and vital statistics: Progress in the data revolution for counting and accountability. *The Lancet*, *386*(10001), 1373–1385.

Ahmadipanahmehrabadi, V., Hassanzadeh, A., & Mahaki, B. (2019). Bivariate spatio-temporal shared component modeling: Mapping of relative death risk due to colorectal and stomach cancers in iran provinces. *International Journal of Preventive Medicine*, *10*(9).

Alkema, L., & New, J. R. (2014). Global estimation of child mortality using a Bayesian B-spline bias-reduction model. *The Annals of Applied Statistics*, 2122–2149.

Aponte, J. J., Schellenberg, D., Egan, A., Breckenridge, A., Carneiro, I., Critchley, J., et al.others. (2009). Efficacy and safety of intermittent preventive treatment with sulfadoxine-pyrimethamine for malaria in African infants: A pooled analysis of six randomised, placebo-controlled trials. *The Lancet*, *374*(9700), 1533–1542.

Arima, S., Bell, W. R., Datta, G. S., Franco, C., & Liseo, B. (2017). Multivariate fay–herriot bayesian estimation of small area means under functional measurement error. *Journal of the Royal Statistical Society: Series A (Statistics in Society)*,

180(4), 1191–1209.

Banerjee, S., Carlin, B. P., & Gelfand, A. E. (2014). *Hierarchical modeling and analysis for spatial data*. CRC Press.

Battese, G. E., Harter, R. M., & Fuller, W. A. (1988). An error-components model for prediction of county crop areas using survey and satellite data. *Journal of the American Statistical Association*, *83*(401), 28–36.

Bchir, A., Bhutta, Z., Binka, F., Black, R., Bradshaw, D., Garnett, G., et al.others. (2006). Better health statistics are possible. *The Lancet*, *367*(9506), 190–193.

Bell, W. R., Datta, G. S., & Ghosh, M. (2013). Benchmarking small area estimators. *Biometrika*, *100*(1), 189–202.

Benavent, R., & Morales, D. (2016). Multivariate fay–herriot models for small area estimation. *Computational Statistics & Data Analysis*, *94*, 372–390.

Bennett, J., & Wakefield, J. (2001). Errors-in-variables in joint population pharmacokinetic/pharmacodynamic modeling. *Biometrics*, *57*(3), 803–812.

Berg, E. J., & Fuller, W. A. (2014). Small area prediction of proportions with applications to the canadian labour force survey. *Journal of Survey Statistics and Methodology*, *2*(3), 227–256.

Besag, J., York, J., & Mollié, A. (1991). Bayesian image restoration, with two applications in spatial statistics. *Annals of the Institute of Statistical Mathematics*, *43*(1), 1–20.

Binder, D. A. (1983). On the variances of asymptotically normal estimators from complex surveys. *International Statistical Review/Revue Internationale de Statistique*, 279–292.

Blangiardo, M., Boulieri, A., Diggle, P., Piel, F. B., Shaddick, G., & Elliott, P. (2020). Advances in spatiotemporal models for non-communicable disease surveillance. *International Journal of Epidemiology*, *49*(Supplement_1), i26–i37.

Boerma, J. T. (2013). Public health information needs in districts. *BMC Health Services Research*, *13*(2), S12.

Boerma, J. T., & Stansfield, S. K. (2007). Health statistics now: Are we making the right investments? *The Lancet*, *369*(9563), 779–786.

Breslow, N. E., & Clayton, D. G. (1993). Approximate inference in generalized linear mixed models. *Journal of the American Statistical Association*, *88*(421), 9–25.

Burstein, R., Henry, N. J., Collison, M. L., Marczak, L. B., Sligar, A., Watson, S., et al.others. (2019). Mapping 123 million neonatal, infant and child deaths between 2000 and 2017. *Nature*, *574*(7778), 353–358.

Carlin, B. P., & Banerjee, S. (2003). Hierarchical multivariate CAR models for spatio-temporally correlated survival data. *Bayesian Statistics*, *7*(7), 45–63.

Carlin, B. P., & Ma, H. (2007). Bayesian multivariate areal wombling for multiple disease boundary analysis. *Bayesian Analysis*, *2*(2), 281–302.

Centers for Disease Control and Prevention. (1996). *CDC standard deviation-derived growth reference curves derived from NCHS/CDC reference population: NCHS growth curves for children, birth to 18 years.* Atlanta, GA: Centers for Disease Control; Prevention.

Chi, E. M., & Reinsel, G. C. (1989). Models for longitudinal data with random effects and AR (1) errors. *Journal of the American Statistical Association*, *84*(406), 452–459.

Clark, S. J., Setel, P., & Li, Z. R. (2019). Verbal autopsy in civil registration and vital statistics: The symptom-cause information archive. *arXiv Preprint arXiv:1910.00405*.

Corsi, D. J., Neuman, M., Finlay, J. E., & Subramanian, S. (2012). Demographic and health surveys: A profile. *International Journal of Epidemiology*, *41*(6), 1602–1613.

Cunningham, S. A., Shaikh, N. I., Nhacolo, A., Raghunathan, P. L., Kotloff, K., Naser, A. M., et al.others. (2019). Health and demographic surveillance systems within the child health and mortality prevention surveillance network. *Clinical Infectious Diseases*, *69*(Supplement_4), S274–S279.

Datta, A., Banerjee, S., Finley, A. O., Hamm, N. A., & Schaap, M. (2016). Nonseparable dynamic nearest neighbor gaussian process models for large

spatio-temporal data with an application to particulate matter analysis. *The Annals of Applied Statistics*, *10*(3), 1286.

Datta, A., Banerjee, S., Hodges, J. S., Gao, L., et al. (2019). Spatial disease mapping using directed acyclic graph auto-regressive (DAGAR) models. *Bayesian Analysis*, *14*(4), 1221–1244.

Datta, G. S. (2009). Model-based approach to small area estimation. *Handbook of Statistics*, *29*, 251–288.

Datta, G. S., Day, B., & Maiti, T. (1998). Multivariate bayesian small area estimation: An application to survey and satellite data. *Sankhya: The Indian Journal of Statistics, Series A*, 344–362.

Datta, G. S., Fay, R. E., & Ghosh, M. (1991). Hierarchical and empirical bayes analysis in small-area estimation. In *Proceedings of the annual research conference* (pp. 63–78).

Datta, G. S., Ghosh, M., Steorts, R., & Maples, J. (2011). Bayesian benchmarking with applications to small area estimation. *Test*, *20*(3), 574–588.

Datta, G. S., Kubokawa, T., Molina, I., & Rao, J. N. (2011). Estimation of mean squared error of model-based small area estimators. *Test*, *20*(2), 367–388.

Desai, N., Aleksandrowicz, L., Miasnikof, P., Lu, Y., Leitao, J., Byass, P., et al.others. (2014). Performance of four computer-coded verbal autopsy methods for cause of death assignment compared with physician coding on 24,000 deaths in low-and middle-income countries. *BMC Medicine*, *12*(1), 20.

Diggle, P. J., & Giorgi, E. (2019). *Model-based geostatistics for global public health: Methods and applications*. CRC Press.

Dong, T. Q., & Wakefield, J. (2021). Modeling and presentation of vaccination coverage estimates using data from household surveys. *Vaccine*, *39*(18), 2584–2594.

Earnest, A., Beard, J. R., Morgan, G., Lincoln, D., Summerhayes, R., Donoghue, D., … Mengersen, K. (2010). Small area estimation of sparse disease counts using shared component models-application to birth defect registry data in new south wales,

australia. *Health & Place*, *16*(4), 684–693.

Esteban, M. D., Herrador, M., Hobza, T., & Morales, D. (2011). An area-level model with fixed or random domain effects in small area estimation problems. In *Modern mathematical tools and techniques in capturing complexity* (pp. 303–314). Springer.

Esteban, M. D., Lombardía, M. J., López-Vizcaíno, E., Morales, D., & Pérez, A. (2022). Small area estimation of expenditure means and ratios under a unit-level bivariate linear mixed model. *Journal of Applied Statistics*, *49*(1), 143–168.

Esteban, M. D., Morales, D., & Pérez, A. (2016). Area-level spatio-temporal small area estimation models. *Analysis of Poverty Data by Small Area Estimation*, 205–226.

Fagbamigbe, A., Kandala, N.-B., & Uthman, A. (2020). Demystifying the factors associated with rural–urban gaps in severe acute malnutrition among under-five children in low-and middle-income countries: A decomposition analysis. *Scientific Reports*, *10*(1), 1–15.

Fay, R. E. (1987). Application of multivariate regression to small domain estimation. *Small Area Statistics*, 91–102.

Fay, R. E., & Herriot, R. A. (1979). Estimates of income for small places: An application of james-stein procedures to census data. *Journal of the American Statistical Association*, *74*(366a), 269–277.

Franco, C., & Bell, W. R. (2021). Using american community survey data to improve estimates from smaller surveys through bivariate small area estimation models. *Journal of Survey Statistics and Methodology,*.

Friberg, I. K., Kinney, M. V., Lawn, J. E., Kerber, K. J., Odubanjo, M. O., Bergh, A.-M., et al.others. (2010). Sub-saharan africa's mothers, newborns, and children: How many lives could be saved with targeted health interventions? *PLoS Medicine*, *7*(6), e1000295.

Fuglstad, G.-A., Li, Z. R., & Wakefield, J. (2021). The two cultures for prevalence mapping: Small area estimation and spatial statistics. *arXiv Preprint arXiv:2110.09576*.

Gelfand, A. E., & Vounatsou, P. (2003). Proper multivariate conditional autoregressive models for spatial data analysis. *Biostatistics*, *4*(1), 11–15.

Gelman, A., Hwang, J., & Vehtari, A. (2014). Understanding predictive information criteria for bayesian models. *Statistics and Computing*, *24*(6), 997–1016.

Ghosh, M., Natarajan, K., Stroud, T., & Carlin, B. P. (1998). Generalized linear models for small-area estimation. *Journal of the American Statistical Association*, *93*(441), 273–282.

Ghosh, M., & Steorts, R. C. (2013). Two-stage benchmarking as applied to small area estimation. *Test*, *22*(4), 670–687.

Ghosh, M., Zhang, L., & Mukherjee, B. (2006). Equivalence of posteriors in the bayesian analysis of the multinomial-poisson transformation. *Metron - International Journal of Statistics*, *64*(1), 19–28.

Glass, R. I., Guttmacher, A. E., & Black, R. E. (2012). Ending preventable child death in a generation. *Journal of the American Medical Association*, *308*(2), 141–142.

Godwin, J., & Wakefield, J. (2021). Space-time modeling of child mortality at the admin-2 level in a low and middle income countries context. *Statistics in Medicine*, *40*(7), 1593–1638.

González-Manteiga, W., Lombarda, M., Molina, I., Morales, D., & Santamaría, L. (2010). Small area estimation under fay–herriot models with non-parametric estimation of heteroscedasticity. *Statistical Modelling*, *10*(2), 215–239.

González-Manteiga, W., Lombardía, M. J., Molina, I., Morales, D., & Santamaría, L. (2008). Analytic and bootstrap approximations of prediction errors under a multivariate fay–herriot model. *Computational Statistics & Data Analysis*, *52*(12), 5242–5252.

Guha, S., & Chandra, H. (2021). Measuring disaggregate level food insecurity via multivariate small area modelling: Evidence from rural districts of Uttar Pradesh, India. *Food Security*, *13*(3), 597–615.

Hájek, J. (1971). Discussion of "An essay on the logical foundations of survey sampling,

Part i," by D. Basu. *Foundations of Statistical Inference*, 326.

Hartzel, J., Agresti, A., & Caffo, B. (2001). Multinomial logit random effects models. *Statistical Modelling*, *1*(2), 81–102.

He, C., Liu, L., Chu, Y., Perin, J., Dai, L., Li, X., et al.others. (2017). National and subnational all-cause and cause-specific child mortality in China, 1996–2015: A systematic analysis with implications for the Sustainable Development Goals. *The Lancet Global Health*, *5*(2), e186–e197.

Heaton, M. J., Datta, A., Finley, A. O., Furrer, R., Guinness, J., Guhaniyogi, R., et al.others. (2019). A case study competition among methods for analyzing large spatial data. *Journal of Agricultural, Biological and Environmental Statistics*, *24*(3), 398–425.

Held, L., Natário, I., Fenton, S. E., Rue, H., & Becker, N. (2005). Towards joint disease mapping. *Statistical Methods in Medical Research*, *14*(1), 61–82.

Held, L., Schrödle, B., & Rue, H. (2010). Posterior and cross-validatory predictive checks: A comparison of MCMC and INLA. In *Statistical modelling and regression structures* (pp. 91–110). Springer.

Herrador, M., Esteban, M., Hobza, T., & Morales, D. (2011). A fay–herriot model with different random effect variances. *Communications in Statistics—Theory and Methods*, *40*(5), 785–797.

Hobza, T., Morales, D., & Santamaría, L. (2018). Small area estimation of poverty proportions under unit-level temporal binomial-logit mixed models. *Test*, *27*(2), 270–294.

Holford, T. R. (1976). Life tables with concomitant information. *Biometrics*, 587–597.

Holford, T. R. (1980). The analysis of rates and of survivorship using log-linear models. *Biometrics*, 299–305.

Horvitz, D. G., & Thompson, D. J. (1952). A generalization of sampling without replacement from a finite universe. *Journal of the American Statistical Association*, *47*(260), 663–685.

Huang, E. T., & Bell, W. R. (2006). Using the t-distribution in small area estimation: An application to SAIPE state poverty models. In *Proceedings of the survey research methods section* (pp. 3142–3149). Citeseer.

ICF International. (2012). *DDemographic and health survey sampling and household listing manual*. Calverton, Maryland, USA: ICF International.

Ilinca, S., Di Giorgio, L., Salari, P., & Chuma, J. (2019). Socio-economic inequality and inequity in use of health care services in kenya: Evidence from the fourth kenya household health expenditure and utilization survey. *International Journal for Equity in Health*, *18*(1), 1–13.

Ito, T., & Kubokawa, T. (2021). Empirical best linear unbiased predictors in multivariate nested-error regression models. *Communications in Statistics-Theory and Methods*, *50*(10), 2224–2249.

Jha, P. (2012). Counting the dead is one of the world's best investments to reduce premature mortality. *Hypothesis*, *10*(1), e3.

Jiang, J., Nguyen, T., & Rao, J. S. (2011). Best predictive small area estimation. *Journal of the American Statistical Association*, *106*(494), 732–745.

Jin, X., Banerjee, S., & Carlin, B. P. (2007). Order-free co-regionalized areal data models with application to multiple-disease mapping. *Journal of the Royal Statistical Society: Series B (Statistical Methodology)*, *69*(5), 817–838.

Jin, X., Carlin, B. P., & Banerjee, S. (2005). Generalized hierarchical multivariate CAR models for areal data. *Biometrics*, *61*(4), 950–961.

Joseph, N. K., Macharia, P. M., Ouma, P. O., Mumo, J., Jalang'o, R., Wagacha, P. W., et al.others. (2020). Spatial access inequities and childhood immunisation uptake in kenya. *BMC Public Health*, *20*(1), 1–12.

Keenan, J. D., Bailey, R. L., West, S. K., Arzika, A. M., Hart, J., Weaver, J., et al.others. (2018). Azithromycin to reduce childhood mortality in sub-saharan africa. *New England Journal of Medicine*, *378*(17), 1583–1592.

Keller, J. P., Olives, C., Kim, S.-Y., Sheppard, L., Sampson, P. D., Szpiro, A. A., …

Kaufman, J. D. (2015). A unified spatiotemporal modeling approach for predicting concentrations of multiple air pollutants in the multi-ethnic study of atherosclerosis and air pollution. *Environmental Health Perspectives*, *123*(4), 301–309.

Kenya National Bureau of Statistics. (2010). *The 2009 kenya population and housing census* (Vol. 1). Kenya National Bureau of Statistics.

Kenya National Bureau of Statistics. (2015). *Kenya demographic and health survey 2014*. Rockville, Maryland, USA: Kenya National Bureau of Statistics.

Khan, S., & Hancioglu, A. (2019). Multiple indicator cluster surveys: Delivering robust data on children and women across the globe. *Studies in Family Planning*, *50*(3), 279–286.

Knorr-Held, L., & Best, N. G. (2001). A shared component model for detecting joint and selective clustering of two diseases. *Journal of the Royal Statistical Society: Series A (Statistics in Society)*, *164*(1), 73–85.

Kubokawa, T. (2012). On measuring uncertainty of small area estimators with higher order accuracy. *Journal of the Japan Statistical Society*, *41*(2), 93–119.

Laird, N., & Olivier, D. (1981). Covariance analysis of censored survival data using log-linear analysis techniques. *Journal of the American Statistical Association*, *76*(374), 231–240.

Lawson, A., Schritz, A., Villarroel, L., & Aguayo, G. A. (2020). Multi-scale multivariate models for small area health survey data: A chilean example. *International Journal of Environmental Research and Public Health*, *17*(5), 1682.

Lee, J. Y., Green, P. J., & Ryan, L. M. (2017). On the "poisson trick" and its extensions for fitting multinomial regression models. *arXiv Preprint arXiv:1707.08538*.

Lehtonen, R., & Veijanen, A. (2009). Design-based methods of estimation for domains and small areas. In *Handbook of statistics* (Vol. 29, pp. 219–249). Elsevier.

Leroux, B. G., Lei, X., & Breslow, N. (2000). Estimation of disease rates in small areas: A new mixed model for spatial dependence. In *Statistical models in epidemiology, the environment, and clinical trials* (pp. 179–191). Springer.

Lewandowski, D., Kurowicka, D., & Joe, H. (2009). Generating random correlation matrices based on vines and extended onion method. *Journal of Multivariate Analysis*, *100*(9), 1989–2001.

Li, F., & Zaslavsky, A. M. (2010). Using a short screening scale for small-area estimation of mental illness prevalence for schools. *Journal of the American Statistical Association*, *105*(492), 1323–1332.

Li, Z. R., Hsiao, Y., Godwin, J., Martin, B. D., Wakefield, J., Clark, S. J., … group, its technical advisory. (2019). Changes in the spatial distribution of the under-five mortality rate: Small-area analysis of 122 DHS surveys in 262 subregions of 35 countries in africa. *PloS One*, *14*(1), e0210645.

Lindgren, F., Rue, H., & Lindstrom, J. (2011). An explicit link between gaussian fields and gaussian markov random fields: The stochastic partial differential equation approach. *Journal of the Royal Statistical Society B*, *73*(4), 423–498.

Liu, L., Oza, S., Hogan, D., Chu, Y., Perin, J., Zhu, J., … Black, R. E. (2016). Global, regional, and national causes of under-5 mortality in 2000–15: An updated systematic analysis with implications for the Sustainable Development Goals. *The Lancet*, *388*(10063), 3027–3035.

Liu, S., Wu, X., Lopez, A. D., Wang, L., Cai, Y., Page, A., et al.others. (2016). An integrated national mortality surveillance system for death registration and mortality surveillance, china. *Bulletin of the World Health Organization*, *94*(1), 46–57.

Lohr, S. L., & Prasad, N. N. (2003). Small area estimation with auxiliary survey data. *Canadian Journal of Statistics*, *31*(4), 383–396.

Lumley, T. (2004). Analysis of complex survey samples. *Journal of Statistical Software*, *9*(1), 1–19.

Lumley, T. (2020). Weights in statistics. Retrieved from https://notstatschat.rbind.io/2020/08/04/weights-in-statistics/

Lumley, T. (2021). svyVGAM: Design-based inference in vector generalised linear models.

MacGibbon, B., & Tomberlin, T. J. (1989). Small area estimates of proportions via empirical bayes techniques. *Survey Methodology*, *15*, 237–252.

MacNab, Y. C. (2010). On bayesian shared component disease mapping and ecological regression with errors in covariates. *Statistics in Medicine*, *29*(11), 1239–1249.

MacNab, Y. C. (2016). Linear models of coregionalization for multivariate lattice data: A general framework for coregionalized multivariate CAR models. *Statistics in Medicine*, *35*(21), 3827–3850.

Mahaki, B., Mehrabi, Y., Kavousi, A., & Schmid, V. J. (2018). Joint spatio-temporal shared component model with an application in iran cancer data. *Asian Pacific Journal of Cancer Prevention: APJCP*, *19*(6), 1553.

Mahapatra, P. (2010). An overview of the sample registration system in india. In *Prince mahidol award conference & global health information forum* (pp. 27–30).

Maher, D., Biraro, S., Hosegood, V., Isingo, R., Lutalo, T., Mushati, P., et al.others. (2010). Translating global health research aims into action: The example of the ALPHA network. *Tropical Medicine & International Health*, *15*(3), 321–328.

Manda, S., Haushona, N., & Bergquist, R. (2020). A scoping review of spatial analysis approaches using health survey data in sub-saharan africa. *International Journal of Environmental Research and Public Health*, *17*(9), 3070.

Marhuenda, Y., Molina, I., & Morales, D. (2013). Small area estimation with spatio-temporal fay–herriot models. *Computational Statistics & Data Analysis*, *58*, 308–325.

McCormick, T. H., Li, Z. R., Calvert, C., Crampin, A. C., Kahn, K., & Clark, S. J. (2016). Probabilistic cause-of-death assignment using verbal autopsies. *Journal of the American Statistical Association*, *111*(515), 1036–1049.

Mercer, L. D., Wakefield, J., Pantazis, A., Lutambi, A. M., Masanja, H., & Clark, S. (2015). Space-time smoothing of complex survey data: Small area estimation for child mortality. *The Annals of Applied Statistics*, *9*(4), 1889–1905.

Mikkelsen, L., Phillips, D. E., AbouZahr, C., Setel, P. W., De Savigny, D., Lozano, R.,

& Lopez, A. D. (2015). A global assessment of civil registration and vital statistics systems: Monitoring data quality and progress. *The Lancet*, *386*(10001), 1395–1406.

Molina, I., Saei, A., & Jose Lombardia, M. (2007). Small area estimates of labour force participation under a multinomial logit mixed model. *Journal of the Royal Statistical Society: Series A (Statistics in Society)*, *170*(4), 975–1000.

Mulatya, D. M., & Ochieng, C. (2020). Disease burden and risk factors of diarrhoea in children under five years: Evidence from kenya's demographic health survey 2014. *International Journal of Infectious Diseases*, *93*, 359–366.

Murray, C. J., Lozano, R., Flaxman, A. D., Serina, P., Phillips, D., Stewart, A., et al.others. (2014). Using verbal autopsy to measure causes of death: The comparative performance of existing methods. *BMC Medicine*, *12*(1), 5.

Neelon, B., Gelfand, A. E., & Miranda, M. L. (2014). A multivariate spatial mixture model for areal data: Examining regional differences in standardized test scores. *Journal of the Royal Statistical Society: Series C (Applied Statistics)*, *63*(5), 737–761.

Ngaruye, I., Nzabanita, J., Rosen, D. von, & Singull, M. (2017). Small area estimation under a multivariate linear model for repeated measures data. *Communications in Statistics-Theory and Methods*, *46*(21), 10835–10850.

Nkengasong, J., Gudo, E., Macicame, I., Maunze, X., Amouzou, A., Banke, K., … Jani, I. (2020). Improving birth and death data for african decision making. *The Lancet Global Health*, *8*(1), e35–e36.

O'Brien, K. L., Wolfson, L. J., Watt, J. P., Henkle, E., Deloria-Knoll, M., McCall, N., et al.others. (2009). Burden of disease caused by streptococcus pneumoniae in children younger than 5 years: Global estimates. *The Lancet*, *374*(9693), 893–902.

Paige, J., Fuglstad, G.-A., Riebler, A., & Wakefield, J. (2022). Design-and model-based approaches to small-area estimation in a low-and middle-income country context: Comparisons and recommendations. *Journal of Survey Statistics and Methodology*, *10*(1), 50–80.

Palmí-Perales, F., Gómez-Rubio, V., & Martinez-Beneito, M. A. (2019). Bayesian

multivariate spatial models for lattice data with INLA. *arXiv Preprint arXiv:1909.10804*.

Paradinas, I., Conesa, D., López-Quílez, A., & Bellido, J. M. (2017). Spatio-temporal model structures with shared components for semi-continuous species distribution modelling. *Spatial Statistics*, *22*, 434–450.

Parker, P. A., Holan, S. H., & Janicki, R. (2022). Computationally efficient bayesian unit-level models for non-gaussian data under informative sampling with application to estimation of health insurance coverage. *The Annals of Applied Statistics*, *16*(2), 887–904.

Parker, P. A., Janicki, R., & Holan, S. H. (2019). Unit level modeling of survey data for small area estimation under informative sampling: A comprehensive overview with extensions. *arXiv Preprint arXiv:1908.10488*.

Pedersen, J., & Liu, J. (2012). Child mortality estimation: Appropriate time periods for child mortality estimates from full birth histories.

Penny, M. A., Verity, R., Bever, C. A., Sauboin, C., Galactionova, K., Flasche, S., et al.others. (2016). Public health impact and cost-effectiveness of the RTS,S/AS01 malaria vaccine: A systematic comparison of predictions from four mathematical models. *The Lancet*, *387*(10016), 367–375.

Pfeffermann, D. (2013). New important developments in small area estimation. *Statistical Science*, *28*(1), 40–68.

Pfeffermann, D., Sikov, A., & Tiller, R. (2014). Single-and two-stage cross-sectional and time series benchmarking procedures for small area estimation. *Test*, *23*(4), 631–666.

Phillips, D. E., AbouZahr, C., Lopez, A. D., Mikkelsen, L., De Savigny, D., Lozano, R., … Setel, P. W. (2015). Are well functioning civil registration and vital statistics systems associated with better health outcomes? *The Lancet*, *386*(10001), 1386–1394.

Plummer, M. (2015). Cuts in bayesian graphical models. *Statistics and Computing*, *25*(1), 37–43.

Porter, A. T., Wikle, C. K., & Holan, S. H. (2015). Small area estimation via

multivariate fay–herriot models with latent spatial dependence. *Australian & New Zealand Journal of Statistics*, *57*(1), 15–29.

Prentice, R. L., Kalbfleisch, J. D., Peterson Jr, A. V., Flournoy, N., Farewell, V. T., & Breslow, N. E. (1978). The analysis of failure times in the presence of competing risks. *Biometrics*, 541–554.

Preston, H., Samuel, Heuveline, P., & Guillot, M. (2001). *Demography: Measuring and modeling population processes*. Oxford: Blackwell.

R Core Team. (2013). R: A language and environment for statistical computing. *R Foundation for Statistical Computing*.

Rao, C., Soemantri, S., Djaja, S., Adair, T., Wiryawan, Y., Pangaribuan, L., et al.others. (2010). Mortality in central java: Results from the indonesian mortality registration system strengthening project. *BMC Research Notes*, *3*(1), 325.

Rao, J. N., & Molina, I. (2015). *Small area estimation*. John Wiley & Sons.

Riebler, A., Sørbye, S. H., Simpson, D., & Rue, H. (2016). An intuitive bayesian spatial model for disease mapping that accounts for scaling. *Statistical Methods in Medical Research*, *25*(4), 1145–1165.

Roberts, D. R., Bahn, V., Ciuti, S., Boyce, M. S., Elith, J., Guillera-Arroita, G., et al.others. (2017). Cross-validation strategies for data with temporal, spatial, hierarchical, or phylogenetic structure. *Ecography*, *40*(8), 913–929.

Rue, H., & Held, L. (2005). *Gaussian markov random fields: Theory and applications*. CRC Press.

Rue, H., Martino, S., & Chopin, N. (2009). Approximate bayesian inference for latent gaussian models by using integrated nested laplace approximations. *Journal of the Royal Statistical Society B*, *71*(2), 319–392.

Rumiati, A. T., Kuswanto, H., et al. (2019). Spatio-temporal fay-herriot models in small area estimation to obtain factors that affecting poverty in polewali mandar district. In *IOP conference series: Materials science and engineering* (Vol. 546, p. 052076). IOP Publishing.

Saegusa, T., Sugasawa, S., & Lahiri, P. (2020). Parametric bootstrap confidence intervals for the multivariate fay-herriot model. *arXiv Preprint arXiv:2006.14820*.

Saei, A., & Taylor, A. (2012). Labour force status estimates under a bivariate random components model. *Journal of the Indian Society of Agricultural Statistics*, *66*(1), 187–201.

Sankoh, O., & Byass, P. (2012). The INDEPTH network: Filling vital gaps in global epidemiology. Oxford University Press.

Särndal, C.-E., Swensson, B., & Wretman, J. (2003). *Model assisted survey sampling*. Springer Science & Business Media.

Schmidt, A. M., & Gelfand, A. E. (2003). A bayesian coregionalization approach for multivariate pollutant data. *Journal of Geophysical Research: Atmospheres*, *108*(D24).

Schumacher, A. E., McCormick, T. H., Wakefield, J., Chu, Y., Perin, J., Villavicencio, F., … Liu, L. (2022). A flexible bayesian framework to estimate age-and cause-specific child mortality over time from sample registration data. *The Annals of Applied Statistics*, *16*(1), 124–143.

Scott, A., & Smith, T. M. (1969). Estimation in multi-stage surveys. *Journal of the American Statistical Association*, *64*(327), 830–840.

Simpson, D., Rue, H., Riebler, A., Martins, T. G., Sørbye, S. H., et al. (2017). Penalising model component complexity: A principled, practical approach to constructing priors. *Statistical Science*, *32*(1), 1–28.

Skinner, C., & Wakefield, J. (2017). Introduction to the design and analysis of complex survey data. *Statistical Science*, *32*(2), 165–175.

Slud, E. V., & Maiti, T. (2011). Small-area estimation based on survey data from a left-censored fay–herriot model. *Journal of Statistical Planning and Inference*, *141*(11), 3520–3535.

Smith, M. S., & Khaled, M. A. (2012). Estimation of copula models with discrete margins via bayesian data augmentation. *Journal of the American Statistical*

Association, *107*(497), 290–303.

Snow, R. W., Omumbo, J. A., Lowe, B., Molyneux, C. S., Obiero, J.-O., Palmer, A., et al.others. (1997). Relation between severe malaria morbidity in children and level of plasmodium falciparum transmission in africa. *The Lancet*, *349*(9066), 1650–1654.

Soleman, N., Chandramohan, D., & Shibuya, K. (2006). Verbal autopsy: Current practices and challenges. *Bulletin of the World Health Organization*, *84*, 239–245.

Speckman, P. L., & Sun, D. (2003). Fully bayesian spline smoothing and intrinsic autoregressive priors. *Biometrika*, *90*(2), 289–302.

Stein, M. L. (1999). *Interpolation of spatial data: Some theory for kriging*. Springer Science & Business Media.

Sun, H., Berg, E., & Zhu, Z. (2021). Bivariate small-area estimation for binary and gaussian variables based on a conditionally specified model. *Biometrics*.

Ulijaszek, S. J., & Kerr, D. A. (1999). Anthropometric measurement error and the assessment of nutritional status. *British Journal of Nutrition*, *82*(3), 165–177.

United Nations. (2015). Transforming our world: The 2030 Agenda for Sustainable Development. Resolution adopted by the General Assembly on 25 September 2015. Retrieved from https://www.un.org/sustainabledevelopment/sustainable-development-goals/

United Nations Inter-agency Group for Child Mortality Estimation. (2020). *Levels & trends in child mortality 2020*. United Nations Children's Fund.

United Nations Population Division. (2019). *World Population Prospects*. Dept of International Economic; Social Affairs; United Nations. Retrieved from https://esa.un.org/unpd/wpp

Utazi, C. E., Wagai, J., Pannell, O., Cutts, F. T., Rhoda, D. A., Ferrari, M. J., et al.others. (2020). Geospatial variation in measles vaccine coverage through routine and campaign strategies in nigeria: Analysis of recent household surveys. *Vaccine*, *38*(14), 3062–3071.

Vos, T., Lim, S. S., Abbafati, C., Abbas, K. M., Abbasi, M., Abbasifard, M., et

al.others. (2020). Global burden of 369 diseases and injuries in 204 countries and territories, 1990–2019: A systematic analysis for the global burden of disease study 2019. *The Lancet*, *396*(10258), 1204–1222.

Wakefield, J. (2008). Ecologic studies revisited. *Annual Review of Public Health*, *29*, 75–90.

Wakefield, J. (2020). Prevalence mapping. *Wiley StatsRef: Statistics Reference Online*, 1–7.

Wakefield, J., Fuglstad, G.-A., Riebler, A., Godwin, J., Wilson, K., & Clark, S. J. (2019). Estimating under-five mortality in space and time in a developing world context. *Statistical Methods in Medical Research*, *28*(9), 2614–2634.

Wakefield, J., Okonek, T., & Pedersen, J. (2020). Small area estimation for disease prevalence mapping. *International Statistical Review*, *88*(2), 398–418.

Walker, C. L. F., Rudan, I., Liu, L., Nair, H., Theodoratou, E., Bhutta, Z. A., … Black, R. E. (2013). Global burden of childhood pneumonia and diarrhoea. *The Lancet*, *381*(9875), 1405–1416.

Walker, N., Bryce, J., & Black, R. E. (2007). Interpreting health statistics for policymaking: The story behind the headlines. *The Lancet*, *369*(9565), 956–963.

Wang, H., Abbas, K. M., Abbasifard, M., Abbasi-Kangevari, M., Abbastabar, H., Abd-Allah, F., et al.others. (2020). Global age-sex-specific fertility, mortality, healthy life expectancy (HALE), and population estimates in 204 countries and territories, 1950–2019: A comprehensive demographic analysis for the global burden of disease study 2019. *The Lancet*, *396*(10258), 1160–1203.

Wang, X., Berg, E., Zhu, Z., Sun, D., & Demuth, G. (2018). Small area estimation of proportions with constraint for national resources inventory survey. *Journal of Agricultural, Biological and Environmental Statistics*, *23*(4), 509–528.

Watjou, K., Faes, C., Lawson, A., Kirby, R., Aregay, M., Carroll, R., & Vandendijck, Y. (2017). Spatial small area smoothing models for handling survey data with nonresponse. *Statistics in Medicine*, *36*(23), 3708–3745.

Wheldon, M. C., Raftery, A. E., Clark, S. J., & Gerland, P. (2013). Reconstructing past populations with uncertainty from fragmentary data. *Journal of the American Statistical Association*, *108*(501), 96–110.

WHO Collaborative Study Team on the Role of Breastfeeding on the Prevention of Infant Mortality. (2001). Effect of breastfeeding on infant and child mortality due to infectious diseases in less developed countries: A pooled analysis. *The Lancet*, *355*, 451–455.

WorldPop. (2022, July). Retrieved from www.worldpop.org

Wu, Y., Li, Z. R., Mayala, B. K., Wang, H., Gao, P., Paige, J., … Wakefield, J. (2021). *Spatial modeling for subnational administrative level 2 small-area estimation*. DHS Spatial Analysis Reports No. 21. Rockville, Maryland, USA: ICF.

Yang, G., Hu, J., Rao, K. Q., Ma, J., Rao, C., & Lopez, A. D. (2005). Mortality registration and surveillance in china: History, current situation and challenges. *Population Health Metrics*, *3*(1), 3.

Ybarra, L. M., & Lohr, S. L. (2008). Small area estimation when auxiliary information is measured with error. *Biometrika*, *95*(4), 919–931.

You, D., Hug, L., Ejdemyr, S., Idele, P., Hogan, D., Mathers, C., et al.others. (2015). Global, regional, and national levels and trends in under-5 mortality between 1990 and 2015, with scenario-based projections to 2030: A systematic analysis by the UN Inter-agency Group for Child Mortality Estimation. *The Lancet*, *386*(10010), 2275–2286.

You, Y., & Zhou, Q. M. (2011). Hierarchical bayes small area estimation under a spatial model with application to health survey data. *Survey Methodology*, *37*(1), 25–37.

Zhang, L.-C., & Chambers, R. L. (2004). Small area estimates for cross-classifications. *Journal of the Royal Statistical Society: Series B (Statistical Methodology)*, *66*(2), 479–496.