Bayesian epidemic models to infer spatio-temporal HIV incidence with applications to Malawi

---

A Thesis

Presented to the

Department of Mathematics

Faculty of Natural Sciences

Imperial College London

In Partial Fulfillment

of the Requirements for the Degree

Doctor of Philosophy

Timothy M Wolock

10 March 2022

Dr Seth Flaxman (Supervisor)

Dr Jeffrey W Eaton (Co-supervisor)

Copyright

Statement of Originality

All of the work presented here was work that I conducted myself, with the guidance of my supervisors, unless otherwise referenced. I wrote every chapter in this document, including Chapter 7, which reproduces a previously published paper. In all cases where I describe others' work, I provide appropriate references.

Abstract

The rate of new human immunodeficiency virus (HIV) infections has decreased since its global peak in the mid-1990s. As the epidemic continues to recede, the efficacy of treatment and prevention programmes will depend on how well they target the right people in the right places. Few existing models of HIV burden offer spatio-temporally resolved estimates of incidence.

I propose a novel epidemic model of HIV that bridges the gap between spatially resolved models of prevalence and epidemiologically sound compartmental models of incidence. It relies less heavily on fixed, exogenous data than previous models and fits directly to data from household surveys, antenatal care facilities, and HIV treatment programmes. Here, I present the details of the model, a broad set of specification tests, and descriptive results from an application to Malawi, as well as a comparison of methods for smoothing and interpolating sexual partner age data in preparation for adding age structure.

The model comparisons identified a set of specification decisions that consistently led to better out-of-sample fit to district-level data from Malawi. I selected a single specification that fit well to most data and reproduced the programme data nearly perfectly. It estimated that increases in ART coverage resulted in decreases in HIV incidence in Malawi but that spatial heterogeneity in incidence was high.

The proposed model offers several improvements on previous models of HIV incidence, and cross-validated model specification experiments provide relative confidence that the specification used here is appropriate for the epidemic in Malawi. The results from the selected specification underscore the continued success of the Malawian HIV treatment programme and highlight the possibility that, even in a high-prevalence setting, the epidemic is becoming increasingly concentrated.

Table of Contents

List of Tables

List of Figures

Dedication

This work is dedicated to my grandmother, Dr Isabel Wolock, who I am still trying to grow up to be.

Acknowledgements

Throughout my PhD, I have been helped and supported by so many individuals and groups. First, I want to thank my supervisors, Dr Seth Flaxman and Dr Jeff Eaton, for their knowledge, advice, and patience and for encouraging me to take the leap to move to London. I am also grateful for the support of my colleagues the Imperial College HIV Inference Research Group, who are too many in number to list here.

I am also grateful to my funding source, the Imperial College President's PhD Scholarship.

I thank the Malawi Department of HIV & AIDS for their assistance and for graciously granting us access to their data. I also want to thank the participants in and investigators of the 2004, 2010, and 2015-2016 Malawi Demographic and Health Surveys and the 2016 Malawi Population HIV Impact Assessment survey, which have provided so many valuable insights.

I want to thank my family and friends for their support and (mostly) unending patience. I thank my parents for believing in me unconditionally. Finally, I want to express my deepest gratitude to my partner, Alanna, who has been a font of wisdom, a well of stability, and a sympathetic ear for ten years.

Table of abbreviations

Term	Abbreviation
acquired immune deficiency syndrome	AIDS
African Health Research Institute	AHRI
antenatal care	ANC
antiretroviral therapy	ART
autoregressive integrated moving average	ARIMA
credible interval	CI
Demographic and Health Surveys	DHS
Estimation and Projection Package	EPP
expected log posterior density	ELPD
exponential-family random graph model	ERGM
graphical processing unit	GPU
high-performance computing	HPC
human immunodeficiency virus	HIV
incidence rate ratio	IRR
Joint United Nations Programme on HIV and AIDS	UNAIDS
Malawi Department of HIV and AIDS	DHA
Malawi Population-Based Health Impact Assessment	MPHIA
Markov Chain Monte Carlo	MCMC
maximum a posteriori	MAP
Médecins Sans Frontières	MSF
No U-Turn Sampler	NUTS

(continued)

Term	Abbreviation
ordinary differential equation	ODE
people living with HIV	PLHIV
Population-Based Health Impact Assessment	PHIA
primary sampling unit	PSU
probability density function	PDF
quantile-quantile	QQ
root mean squared error	RMSE
sexually transmitted infection	STI
sub-Saharan Africa	SSA
Susceptible-Infectious-Recovered	SIR
Template Model Builder	TMB

Chapter 1

Introduction

With the advent of widely available, highly effective treatment, the health burden of the global human immunodeficiency virus (HIV) epidemic has improved dramatically over the past several decades. New HIV infections have decreased from a global peak of 3 million annually in 1997 to 1.5 million annually in 2020, while annual deaths due to HIV infection have decreased from a peak of 1.9 million in 2004 to 680,000 in 2020 (UNAIDS, 2021). Although HIV elimination is thought to be a target for the far distant future, measuring progress towards *epidemic transition*, in which low levels of acquired immunodeficiency syndrome (AIDS) mortality are maintained and new infections are reduced to zero, is critical for guiding policymaking at both the global and local levels (Ghys et al., 2018; Over, 2012).

The rate of new HIV infection in a population, or HIV *incidence*, is arguably the most important metric for measuring the trajectory of the epidemic (Ghys et al., 2018; Hallett, 2011). Relative to other potential metrics, incidence provides a clearer picture of short- and long-term treatment needs and the future trajectory of the epidemic. Observing increasing incidence in a population with low treatment coverage suggests that there will soon be a large treatment gap in that population. For example, observing high incidence in a population with a low burden of current infection suggests that the population could soon experience a treatment gap.

However, population-level HIV incidence is difficult to measure directly (Brookmeyer, 2010; Hallett, 2011). A person living with HIV could survive for ten years or more without treatment and 30 years or longer with treatment (Collaborative Group on AIDS Incubation

and HIV Survival, 2000; Todd et al., 2007), so cross-sectional measurements of the share of people currently living with HIV (or *prevalence*) are bound to represent individuals who could have been infected many years in the past. If population-representative HIV testing data are widely available over space and time, we can leverage assumptions about survival with and without treatment to backcalculate incidence from prevalence (Bacchetti, Segal, & Jewell, 1993; Brookmeyer & Gail, 1988; Sakarovitch et al., 2007).

In this application, however, I focus on incidence estimation in high-burden settings where data are typically sparse and insufficient for backcalculation of general population incidence. The epidemic still disproportionately affects sub-Saharan Africa (SSA), with approximately 67% of people living with HIV (PLHIV) globally living in an SSA country in 2020 (UNAIDS, 2021). In high-income countries detailed case surveillance data describing characteristics for all new HIV diagnoses have been the standard for tracking incidence. In lower-income countries and those with large HIV burdens, reliable data on key HIV indicators are typically sparse. These challenges have led to the development of a variety of modelling strategies for inferring temporal trends in HIV incidence in data-sparse settings, which are detailed in Section 1.5.

The majority of models of HIV incidence have focused on estimation of national-level changes over time, but a growing body of literature argues that, as declines in incidence continue, new infections will become increasingly concentrated in particular areas and populations (Anderson et al., 2014; Cuadros et al., 2022; McGillen et al., 2016; Meyer-Rath et al., 2018). As a result, prevention strategies must be sensitive to the fundamental spatial variation in the epidemic. These authors argue that the success of global efforts to fight HIV will increasingly depend on whether treatment and prevention programmes are implemented in the right locations at the right times.

Although being in a particular location will not directly cause an individual to be infected with HIV, infection arises through complex socio-biological processes that depend on individual- and societal-level factors, many of which vary systematically over space (Boerma & Weir, 2005; Lewis et al., 2007). Place-of-residence and place-of-work are often effective proxies for risk factors that might be difficult or impossible to measure. More directly, nearly all new HIV infections in sub-Saharan Africa (SSA) are caused by sexual contact with people living with HIV (Kharsany & Karim, 2016).

Multiple studies have demonstrated that HIV prevalence varies dramatically over space both across and within countries (Cuadros et al., 2018; Dwyer-Lindgren et al., 2019), but similar attention has not been paid to incidence. If prevalence varies over space, incidence likely does as well, but the spatial distribution of incidence is not well measured.

Existing models of HIV burden vary in many ways, but they share features that make them inappropriate for spatio-temporal inference of HIV incidence. Broadly speaking, the models that infer temporal trends in incidence scale poorly to smaller areas (both in terms of computation and validity of assumptions), while those that include explicit spatial components fail to connect prevalence data to incidence.

1.1 Thesis aims

The goal of my thesis is to fill this gap by developing a spatially structured epidemic model that infers HIV incidence over space and time simultaneously. The proposed method connects epidemic modelling tools used to estimate national-level HIV incidence to small-area estimation methods commonly incorporated into spatial models of HIV burden.

Without directly observed data on incidence, validating the accuracy of this model is difficult; we cannot evaluate how well the model predicts true incidence if incidence is not observed. Instead, I propose a cross-validation strategy that estimates how well the model predicts recent data on prevalence and treatment provision. If each set of observable indicators corresponds to a unique incidence series, then this strategy does provide cross-validation of incidence. On other hand, if multiple incidence series lead to similarly well-fitting predictions of prevalence and treatment, then the cross-validation strategy will only be able to identify a *set* of incidence series that could have generated the data. Given that incidence is just one of the unknown quantities that affects prevalence, the second scenario seems more likely *a priori*.

Although other models of population-level HIV incidence share these problems, we can validate them relative to each other. Any model of incidence with sufficiently granular spatio-temporal resolution can be aggregated to align with corresponding national-level and cross-sectional models. Concordance in estimated incidence across varying methodologies suggests either that the methods are generally accurate or that they are inaccurate in similar

ways. By measuring out-of-sample fit to data on observable indicators and comparing to external estimates, I provide some amount empirical support for the model, despite the difficulty in direct validation.

Because it is impossible to measure how accurately this model predicts true incidence, I cannot define an acceptable margin of error in terms of prediction accuracy. Instead, I define sufficient precision in terms of the posterior probabilities of certain percent changes in incidence across districts. Specifically, I call the model sufficiently precise if, for every study region, it estimates that the posterior probability of incidence having decreased by 50% or more between 2010 and 2021 was less than 20% or greater than 80%. In other words, the model must be relatively certain that incidence was (or was not) halved over this period. This strategy favours models that predict changes in incidence precisely, even if we are unable to validate the levels of incidence. I test for a 50% decrease because it represents substantial but achievable progress, although it is a considerably less ambitious reduction than the 75% reduction proposed in the 2016 United Nations Political Declaration on Ending AIDS (*Political declaration on HIV and AIDS*, 2016). I define probabilities of less than 20% or greater than 80% as reasonable certainty essentially arbitrarily; 80% aligns with conventional power calculations. Any of these thresholds could be defined differently depending on the intended use of the model. For example, if the user is concerned with elimination, then they should target high posterior probabilities of low levels of incidence.

In the remainder of this chapter, I will provide background on data streams measuring HIV in sub-Saharan Africa, outline several relevant models of HIV incidence and other key indicators, and describe how these data sources and models have influenced my own work.

1.2 Epidemiological indicators

I have defined prevalence and incidence loosely above, but more precise definitions of those and several other key epidemiological indicators will be useful. “Prevalence” is the number of PLHIV in a given population divided by the size of that population at a single, instantaneous point in time (more precisely called “point prevalence”). Similarly, “ART coverage” is the number of PLHIV receiving antiretroviral therapy (ART) in a population

divided by the number of PLHIV in that population at a given point in time. Note that prevalence and ART coverage must be between zero and one. Finally, I define "incidence" as the rate of new infections per susceptible person-year. Whereas prevalence and ART coverage are defined in terms of people, incidence is defined in terms of person-time, which accounts for the differential impact of each new infection as the pool of susceptibles is depleted.[1] Note that incidence rates can be greater than one, although such a high level of new HIV infection is highly unlikely.

1.3 Data in sub-Saharan Africa

The modelling strategies used to estimate HIV burden in sub-Saharan Africa have been developed to take advantage of the specific configurations of data available in that region, so I will provide a brief outline of population-level data sources commonly available in SSA before describing previous work. Section 2.5 describes in much greater detail how these data can be incorporated into a Bayesian epidemic model.

Most sub-Saharan African countries lack reliable historical time series of the two indicators most frequently used to estimate HIV in higher income settings: cause-specific death counts and case surveillance. In higher income countries, new HIV diagnoses are reported to central health authorities, which combine them with reported counts of age- and sex-specific deaths due to acquired immune deficiency syndrome (AIDS) to estimate HIV burden (and in particular, incidence) quite precisely (Centers for Disease Control and Prevention, 2021; Mahiane et al., 2019). Many lower and middle-income countries lack the reporting infrastructure necessary to provide these series, so we must leverage other data sources.

1.3.1 Household surveys

The most important sources of population-level HIV burden data in SSA have been large, nationally representative household surveys like the Demographic and Health Surveys

[1] If I defined incidence in terms of susceptible people at the start of a period, it would more appropriately be called a "risk," as opposed to a "rate." That said, because new HIV infection is a rare event, incidence risks and rates will be nearly identical in most populations.

(DHS) and the Population-Based Health Impact Assessment (PHIA) surveys (ICAP at Columbia University & PEPFAR, 2019; "The DHS program," 2021). DHS surveys are conducted approximately once every five years in much of sub-Saharan Africa, while PHIA surveys are conducted in pre-planned waves in specific focus countries. Both survey series test individuals for HIV, providing reliable estimates of prevalence, and the PHIA surveys additionally test for ART adherence, recent infection, and viral load.

Although the specific sampling methodologies differ across series, they typically use two-stage cluster-stratified sampling schemes that need to be considered when integrating the data into a model. As an illustration, DHS surveys are sampled as follows:

1. The nation is stratified into some number of subnational administrative units.
2. The complete list of potential primary sampling units (PSUs) and their populations is compiled, usually from the list of enumeration areas of the most recent census.
3. PSUs are randomly selected proportionate to their population sizes within each subnational unit.
4. Depending on the resulting split between urban and rural PSUs, more urban PSUs might be added to the sample to account for their relative importance compared to the share of the population they represent.
5. Using maps of each PSU, some number of households are selected heuristically by a central survey planning office. For example, if the desired number of households in a given PSU is 10% of the total number, the survey planner might select every tenth house according to their order on a path through the PSU.
6. Interviewers then visit each selected household and interview all adults aged 15-49 years.

The surveys use a reference population (typically, the most recent census) and the observed sample to calculate each individual and household's probability of having been sampled. Taking the inverse of a respondent's probability of selection gives the number of individuals that respondent represents, a quantity referred to as *survey weight*. These weights are constructed so that representative estimates can be produced at national and subnational levels stratified by urban/rural status. For example, the sum of all individual-level survey weights within a subnational unit will match the population of that subnational unit in the reference population.

The variance of estimated counts derived from these surveys are calculated in a way that takes into account the complexity of their sampling schemes. To incorporate these design-based variances into statistical models, we can find the *effective sample size* of the estimate. Let Y be an estimated survey count, τ be the design-based variance associated with Y, and $p = Y/N$ be the estimated proportion associated with Y. We can define the "design effect" of the survey as

$$D_{\text{eff}} = \frac{Np(1-p)}{\tau}. \tag{1.1}$$

Here, $Np(1-p)$ is the variance of Y had Y been sampled via simple random sampling. This ratio measures the efficiency of the sampling scheme that generated τ relative to simple random sampling. If we set $D_{\text{eff}} = 1.0$ and fix p and τ to their observed values, we can solve for the sample size under simple random sampling to estimate Y with the same efficiency as the survey:

$$N_{\text{eff}} = \frac{\tau}{p(1-p)}. \tag{1.2}$$

I refer to N_{eff} as the *effective sample size* of Y and pN_{eff} as the *effective count*. This is a convenient approximation that allows us incorporate survey estimates and their variances into a natural framework for count data.

Household surveys offer cross-sectional estimates of HIV prevalence, treatment coverage, and infection recency by age and sex. They are representative at the national and first administrative unit levels, making them an invaluable resource for HIV burden estimation. However, due to budgetary constraints and changing priorities, we cannot be sure that the DHS will continue to conduct HIV testing or that there will be future rounds of the PHIA surveys. Therefore, we must keep in mind that any modelling strategy that relies too heavily on cross-sectional estimates from these surveys might run out of data that can reasonably be called "recent" in the next few years.

Further, household surveys are conducted only infrequently and are therefore poorly suited to estimating temporal trends. For example, including the most recent PHIA, household surveys comprise a total of five data points for Malawi over the entire course of the epidemic.

1.3.2 Antenatal facilities

The second most important data source for HIV incidence estimation in sub-Saharan Africa is reported HIV test results from antenatal care (ANC) facilities. The vast majority of pregnant women in SSA receive antenatal care, including HIV testing, at government-run clinics, making reported ANC test results a relatively complete survey of pregnant women at any given time. For example, Ataguba (2018) estimated that 98% of pregnant women in Uganda attended at ANC services at least once per pregnancy.

ANC testing data have been used extensively by the Joint United Nations Programme on HIV and AIDS (UNAIDS) to estimate HIV burden. Historically, the UNAIDS estimation process has used data from *sentinel surveillance* systems, in which a small number of sites tested anonymised blood samples left over from anaemia testing for HIV and reported the results to UNAIDS (Joint United Nations Programme on HIV/AIDS et al., 2004). However, in 2015, the World Health Organization and UNAIDS recommended that countries switch to routine testing data for burden estimation (World Health Organization & UNAIDS, 2015). The routine testing data still come from the original sentinel surveillance sites, which were typically selected because they were thought to have been in particularly high-prevalence areas, calling into question how well their attendees represent the broader population of pregnant women.

We also cannot assume that pregnant women, who are necessarily sexually active, experience identical risk of HIV infection as the rest of the adult population. Further, as survival continues to improve, the age distribution of current PLHIV in SSA will become older, meaning pregnant women are likely to become an increasingly poor proxy for the general population of PLHIV. In other words, the probability a given PLHIV is in the set of currently pregnant women could be decreasing (Eaton et al., 2014).

1.3.3 ART programme data

Many medical facilities across SSA are now equipped with sophisticated electronic medical record systems that can report treatment indicators to central health authorities with minimal human intervention, giving us a relatively complete picture of treatment at high

spatio-temporal resolution. We might, for example, have monthly reports of the number of patients treated at every facility in a country.

When we consider a country as a whole, we can assume that the ART patient count for a given period is extremely close to the true number of PLHIV living in that country who received treatment during that period and is, therefore, the appropriate numerator for calculating ART coverage (the number of PLHIV on treatment divided by the number of PLHIV total). We only need to assume that a negligibly small share of people on ART receive their treatment outside of reporting facilities and that a similarly small share of patients at those facilities reside outside of the country.

However, as described by Eaton et al. (2021), the second assumption fails when we move from the national level to the subnational level. Because the patient count series are *facility*-based and subnational borders are porous, we cannot assume that all PLHIV receiving treatment in a particular area also live in that area. We might, for example, expect urban areas with more health facilities to attract PLHIV from nearby rural areas, or, in areas where HIV is still stigmatised, PLHIV to seek treatment far from their homes. Any number of possible dynamics could mean that the facility-based patient counts and residency-based surveys are not directly comparable.

1.3.4 Direct measurement of HIV incidence

Notably absent from the population-level data sources listed above is direct measurement of HIV incidence. Cohort studies and the control arms of clinical trials do offer reliable incidence estimates, but their carefully designed sampling schemes do not necessarily correspond well to the general population. The PHIA surveys and some routine testing programmes include recency assays, which measure the avidity of an individual's antibody response to HIV antigens and classifies whether the individual was infected recently (usually, in the previous six months or year). Kassanjee, McWalter, & Welte (2014) provide a method to relate population-level recency to incidence, but the use of these assays to ascertain population-level incidence is still imperfect. Household surveys do not conduct enough recency assays to provide a meaningfully precise estimate of HIV incidence at subnational levels, and the routine testing data are subject the re-testing bias described by Maheu-Giroux et al. (2019).

It is also worth addressing why we cannot simply treat the difference between prevalence at consecutive points in time as incidence. First, the change in the true number of PLHIV from one time point to the next is a function of not only incidence, but also mortality (both AIDS-related and non AIDS-related) and migration. We do not observe mortality directly in most SSA settings, so the change in prevalence over time depends on at least two unobserved quantities. Second, it could take years for a new HIV infection to be detected for the first time, if it is ever detected, meaning changes in prevalence today reflect changes in long-past incidence.

Population cohort studies

General population cohort studies provide valuable, granular information about changes in fertility, mortality, and health, among many other indicators in small, carefully selected areas. They track and test individuals in fixed populations at regular intervals over time and can identify new infections based on each individual's testing history (Gareta et al., 2021; Gregson et al., 2017; Reniers et al., 2016). Cohort studies have been crucial in estimating the biological parameters that govern HIV disease progression. For example, because they track individuals' vital events over time, they can be used to estimate how survival among PLHIV varies by treatment status (Reniers et al., 2014), CD4 count (Trickey et al., 2017), and demographic characteristics (Risher et al., 2021). Further, regular HIV testing allows for reliable estimation of incidence within the study population (Risher et al., 2021). However, because these study populations are carefully selected sets of villages and cities, estimates that depend on local context (such as incidence) might not generalise to regional or national populations.

1.4 HIV data in Malawi

In this work, I have focused on applications to Malawi, a low-income country in Southern Africa. Over the past 40 years, Malawi has, like much of sub-Saharan Africa, experienced a severe HIV epidemic. Annual national-level incidence among adults aged 15-49 peaked at 22 new infections per 1,000 people in 1993, and adult prevalence still ranks among the highest in the world at 8% (UNAIDS, 2020). I have incorporated several datasets of the

types described above into the model I propose in Chapter 2.

As of 2020, four household surveys had conducted HIV testing in Malawi: the 2004, 2010, and 2015-16 Demographic and Household Surveys (DHS), and the 2015-2016 Malawi Population-based HIV Impact Assessment (MPHIA) survey (Ministry of Health, Malawi, 2018; National Statistical Office/Malawi & ICF, 2017; National Statistical Office - NSO/Malawi & ICF Macro, 2011; National Statistical Office - NSO/Malawi & ORC Macro, 2005). A second MPHIA survey was being conducted contemporaneously with this analysis, so I did not include those results here. From the three DHS surveys, I extracted district- and sex-specific HIV assay results, and from MPHIA I extracted district- and sex-specific HIV, ART, and recency assay results. For both survey series, I used data for all 28 districts of Malawi and restricted to individuals aged 15 to 49 years at the time of the survey. Note that because these surveys sample households, the estimates we derived from them implicitly referred to residents of each district.

To measure the number of people receiving ART by district and time, I used data provided by the Malawi Department of HIV and AIDS (DHA) on the number of ART patients in each district by quarter. Médecins Sans Frontières began operating treatment clinics in the Chiradzulu District before the national ART scale-up, so I supplemented programmatic patient counts with manually extracted patient counts from a published case study (Médecins Sans Frontières, Malawi & Ministry of Health and Population, Chiradzulu District, Malawi, 2004).

Finally, I supplemented the ANC sentinel surveillance data used in the UNAIDS estimation process with data provided by the DHA. The UNAIDS process uses data from two facilities per district, which are categorised as either sentinel surveillance or routine testing as described in Section 1.3.2. For years earlier than 2011, I used the sentinel surveillance data when it was available, omitting data from 2010 due to reported difficulties in laboratory testing. From 2011 onwards, I found the identifier for every facility used by UNAIDS in the DHA's reporting system and produced quarterly aggregates to replace the routine testing data used in the UNAIDS process. I verified that I was able to reproduce the annual counts reported to UNAIDS. More than 700 facilities included in the reporting system are not incorporated into the UNAIDS process, so I constructed a quarterly aggregate series for each district, aggregating over the counts of tests and positive tests for every facility not

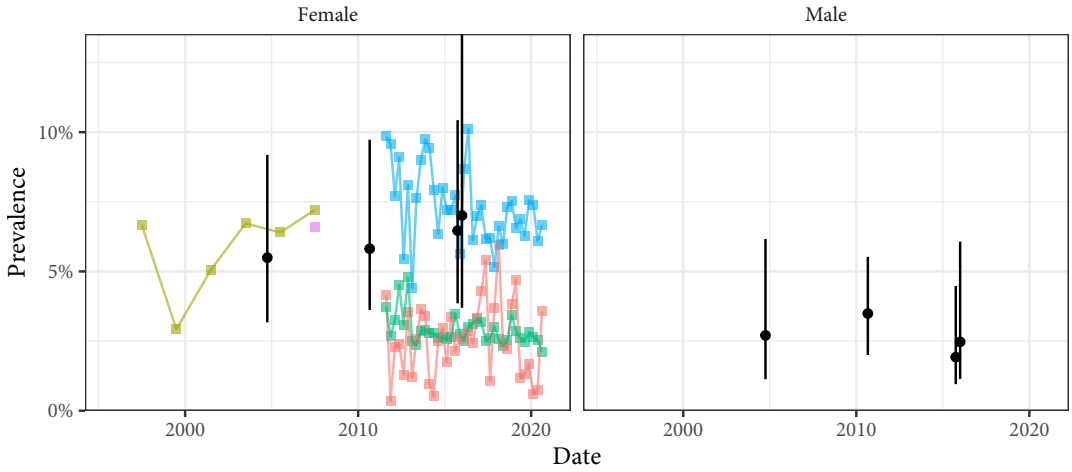

Figure 1.1: HIV prevalence data among adults aged 15-49 years in the Kasungu District of Malawi. Ciricular points and ranges are household survey estimates, and connected square points are ANC facility data. The colour of each series of connected points indicates the facility from which the data originate.

already included in the process.

To provide a clearer image of the available data sources and how they relate to each other, Figure 1.1 presents the prevalence data used in the Kasungu District, which is a medium-sized district of approximately 800,000 people in the centre of the country. The household survey and ANC facility all suggest that there was little change in prevalence in Kasungu between 2005 and 2016. Although the two recent household surveys estimated higher prevalence than the 2010 DHS among women, all three sets of routine ANC data indicate that prevalence among pregnant women is decreasing slowly. I will discuss how these two datasets can be related to each other in greater detail in Section 2.5.2, but this example illustrates how ANC facility data can complement household survey data.

Figure 1.2 presents the number of PLHIV attending ART services in Kasungu. The number of patients in Kasungu increased rapidly between the start of the programme in 2005 and 2020. As of the end of 2020, approximately 13,000 adults aged 15-49 were receiving treatment in Kasungu, compared to only 4,000 at the beginning of 2010.

The descriptive results in Kasungu are broadly similar to descriptive results in other districts. The average change in prevalence among women across the 27 districts that were sampled in both the 2010 and 2015 DHS surveys, was only -1.8 percentage points. The increases in patients were also consistent across district, with the number of people attending ART services increasing by an average of 315% between 2010 and 2020 across all 28 districts.

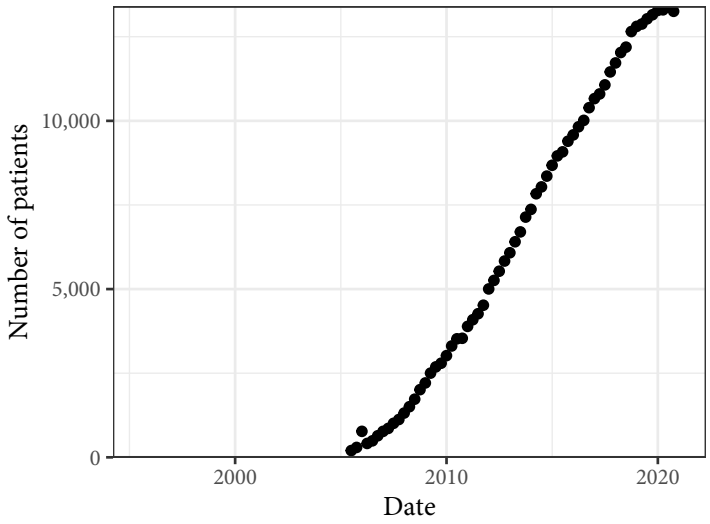

Figure 1.2: Adults aged 15-49 years attending ART services in the Kasungu District of Malawi.

These data suggest that temporal trends in prevalence and treatment were similar across the districts of Malawi but there was substantial spatial heterogeneity in absolute levels. Estimated prevalence among women from the 2015-16 MPHIA survey ranged from a high of 23% in Nsanje to a low of 3% in Dowa, while ART coverage ranged from 92% in Balaka to 59% in Dowa. There was strong systematic variation in prevalence, with HIV prevalence decreasing from the north of the country to the south. Estimated prevalences in 2015-2016 in the Northern, Central, and Southern regions were 7%, 6%, and 16%, respectively.

1.5 Literature review

With this context in mind, we can examine the various inferential models that have been used since the early 1990s to estimate HIV burden in data sparse settings. First, I provide an overview of compartmental models of infectious disease. Then, I describe several previous models of HIV incidence and a number recent spatial models of HIV burden. Finally, I highlight more general work modelling infectious disease over space that could inform approaches to HIV modelling.

I specify that we are interested in "inferential" models because this work focuses on models that are fit to data via some systematic inferential procedure. Many widely used models for HIV burden estimation are, for a variety of reason, only forward projections. Because I am

focused on inferring HIV incidence, I have restricted this review to inferential models.

The Spectrum model is a notable exception because it is used in close coordination with models described below (Stover et al., 2017). Spectrum is a complex compartmental model of HIV and demographic dynamics that includes detailed disease progression, mother-to-child transmission dynamics, and other important features. Spectrum tracks adults by single-year age, sex, HIV positivity, treatment status, and duration on treatment and can been used to project not only the HIV epidemic but also basic demographic trends. Because of its complexity, it is not currently fit to data and is instead used in concert with less complex inferential models.

1.5.1 Compartmental models of infectious disease

Many of the models used to relate observed prevalence to unobserved incidence are *compartmental*, meaning they track the varying sizes of sets of population groups over time. The canonical compartmental model of infectious disease is the "Susceptible-Infectious-Recovered" (SIR) model, which measures the varying sizes of the susceptible, infectious, and recovered groups in a closed population over time (Kermack & McKendrick, 1927).

Compartmental epidemic models are well suited to inferring incidence because they use the fundamental principles of infectious disease project internally consistent epidemic trajectories. Specifically, in a compartmental epidemic model, the count or rate of new infections at a given time depends on the current count or rate of infectious individuals.

The classical SIR model encodes this principle into the following system of ordinary differential equations (ODEs):

$$\begin{aligned} \frac{\partial S}{\partial t} &= -\beta S \frac{I}{N} \\ \frac{\partial I}{\partial t} &= \beta S \frac{I}{N} - \gamma I \\ \frac{\partial R}{\partial t} &= \gamma I, \end{aligned} \tag{1.3}$$

where N is the (constant) population size, β is the expected number of contacts between susceptible and infectious individuals per unit of person-time multiplied by the probability of infection given contact, and γ is the rate of recovery among infectious individuals

(Kermack & McKendrick, 1927). Although γ is canonically referred to as the recovery rate, we can interpret it more generally as the rate at which infectious individuals cease to be infectious via any mechanism (e.g. recovery, death, or perfect treatment).

The formulation presented in Equation (1.3) is often referred to as a *density-dependent* SIR model because the rate of new infections among susceptibles depends on the share of people who are currently infected. For contrast, in a *frequency-dependent* SIR model, we have

$$\frac{\partial S}{\partial t} = -\beta SI, \tag{1.4}$$

and similarly for the partial derivative of $I(t)$. The practical distinction is that, in a frequency-dependent model, the number of potentially infectious contacts scales with the population, while in a density-dependent model, it does not. A frequency-dependent model would be more appropriate for modelling a pathogen that spreads easily through the air, and a density-dependent model might be more appropriate for a sexually transmitted infection. Typically, models of HIV use the density-dependent formulation because an individual's number of sexual partnerships is unlikely to scale proportionately with population.

In either case, the core mechanism of the SIR model is the product of S and I, which represents all pairings of susceptible and infectious individuals. If we assume that all pairs of individuals in the population are equally likely to form at any given time, an assumption referred to as "homogeneous mixing," then the SIR model results in the correct expected incidence.

Compartmental epidemic models differ from each other along two important dimensions: continuity and stochasticity. The model defined in Equation (1.3) varies continuously over time, but discrete-time compartmental models are not uncommon. In a discrete-time SIR model, we have:

$$
\begin{aligned}
S_{t+1} &= S_t - \beta S_t \frac{I_t}{N} \\
I_{t+1} &= I_t + \beta S_t \frac{I_t}{N} - \gamma I_t \\
R_{t+1} &= R_t + \gamma I_t.
\end{aligned}
\tag{1.5}
$$

We note that, depending on the units of person-time used to measure β and γ, the discrete-time SIR model is identical to a continuous-time SIR model integrated with the forward Euler method.

Finally, we can categorise compartmental epidemic models of infectious disease based on whether they are stochastic or deterministic. In a deterministic discrete-time SIR model, the number of new infections in period t is exactly $\beta S_t I_t / N$, while in a stochastic discrete-time SIR model, we assume that it is a random variable of the form $\text{Binom}(S_t, \beta I_t / N)$ or similar. Demographic stochasticity is particularly important to consider when we do not know in advance if a pathogen will reach extinction without producing a meaningful epidemic. For the sake of consistency with previous work and to limit computational cost, I will focus largely on deterministic models.

1.5.2 Models of HIV incidence

Accounting for the lack of direct measurement of HIV incidence described in Section 1.3.4 is the primary purpose of most models of HIV incidence. They necessarily answer the question: how do we relate sparse observations of prevalence or new diagnoses to incidence?

Back-calculation models

Early HIV incidence estimation methods were based on back-calculating new infections from confirmed AIDS case notifications (Bacchetti, Segal, & Jewell, 1993; Brookmeyer & Gail, 1988; Sakarovitch et al., 2007). By making certain assumptions about the distribution of infections over past time periods, $I(s)$, and about the distribution of progression from infection to AIDS given time since infection, $F(t)$, these methods were able to identify the incidence series that were most likely to have generated observed cumulative case counts.

Brookmeyer & Gail (1988) defined the probability an initially susceptible individual who was infected before a given time, T_j, would be diagnosed in time period j as

$$p_j = \int_{T_0}^{T_j} I(s)(F(T_j - s) - F(T_{j-1} - s))ds, \tag{1.6}$$

where $I(s)$ results from an independent and identically distributed Poisson process and $F(t)$ is estimated from analyses of HIV infections via blood transfusion (Brookmeyer & Gail, 1988). They used this model to estimate the minimum size of the AIDS epidemic in the United States up to 1988 (the year of publication).

Ades & Medley (1994) presented a similar model that allowed for variation across age and probability of detection and was fit to seroprevalence data, not AIDS notifications. Using their notation, the odds an individual in age a at time t was HIV-positive was

$$\pi(a,t) = \frac{\int_0^a \{[\exp(-\int_0^{a-z} \theta(u, t-a+u)du)]\theta(a-z, t-z)\phi(a,t,z)\}dz}{\exp(-\int_0^a \theta(u, t-a+u)du)}, \tag{1.7}$$

where $\theta(a,t)$ is the incidence rate among people aged a at time t and $\phi(a,t,z)$ is the probability an individual infected z years ago is detected at age a and time t (Ades & Medley, 1994). They applied this method to seroprevalence data collected from newborns in New York City between 1988 and 1992 to estimate HIV incidence among women of childbearing age. Similarly, Sakarovitch et al. (2007) applied it to seroprevalence data from pregnant women at ANC clinics in Côte d'Ivoire. More recently, Risher et al. (2021) built a similar model into a Bayesian framework to estimate HIV incidence using cohort data from across sub-Saharan Africa.

Back-calculation models are well suited to closed populations with plentiful testing data (*i.e.* populations in which the change in the HIV testing positivity proportion correlates well with true incidence), but we typically do not have such regular testing series in sub-Saharan African settings. We might have suitable data among pregnant women (Sakarovitch et al., 2007) or in a cohort study (Risher et al., 2021), but the relationship between these populations and the general population is not obvious.

The Estimation and Projection Package

The Estimation and Projection Package (EPP) model is a compartmental model used by almost every country in SSA to estimate adult HIV incidence and to produce the incidence and prevalence series necessary to run the more-detailed Spectrum model (Brown et al., 2014; Eaton et al., 2019; Stover et al., 2017). EPP estimates adult HIV incidence from prevalence data collected by large, nationally representative surveys and ANC sentinel surveillance sites.

EPP adapts the basic structure of the SIR model to the particular epidemiological facts of HIV by incorporating vital events, removing the recovered compartment, adding treatment and mortality, and dividing each infected stage into substages that are predictive of mortality.

EPP defines substages using ranges of counts of CD4 T cells per cubic millilitre of blood. CD4 T cells are immune cells that help regulate the body's response to infection and are the cell that HIV most commonly infects. We can quantify how advanced an individual's infection has become by measuring the concentration of these cells in a sample of the individual's blood; a lower CD4 count indicates more advanced disease progression and correlates closely with AIDS mortality (Yiannoutsos et al., 2012).

EPP infers a series of parameters that predict the time-varying HIV transmission rate, typically called $r(t)$ (equivalent to β from the traditional SIR model), that can take one of several functional forms. Given a set of parameters and constant assumptions about treatment, disease progression, and population change derived from other studies, EPP finds $r(t)$ and predicts HIV incidence and corresponding prevalence, $\rho(t)$, from the early 1970s to the present day. Critically, EPP only models a single region at a time.

After projecting $\rho(t)$, EPP synthesises nationally representative household survey data and ANC data by assuming that the surveys offer unbiased measurement of true population-level HIV prevalence and that prevalence at each ANC facility is a random deviation from that of the general population:

$$\Phi^{-1}(\rho^s(t)) = \Phi^{-1}(\rho(t)) + \delta_s + \epsilon_{s,t} \tag{1.8}$$

Here, Φ^{-1} is the inverse of the standard Gaussian density function (the probit transfor-

mation), $\rho^s(t)$ is prevalence among pregnant women at site s, δ_s is a site-specific random effect, and $\epsilon_{s,t}$ is a Gaussian error term (Bao, 2012).

The EPP model has performed well in the past and has influenced HIV modelling and policy-setting substantially, but it is not well-suited to subnational modelling. Households surveys are typically powered to provide precise estimates of HIV prevalence at the national level, not the subnational level. We could improve the precision of our subnational estimates by sharing information across regions (as is typical in spatial statistics and hierarchical modelling more generally), but that is not possible in EPP.

Beyond concerns about subnational data availability, EPP can only model a single population at time and must therefore assume that every new infection in that population is endogenous. This assumption is not realistic at the subnational level and could bias estimated incidence.

Further, EPP requires fixed assumptions about a number of model inputs that might, themselves, be uncertain. In particular, EPP takes a time series of ART patient counts as fixed input. To find the number of PLHIV initiating treatment at each timestep, it subtracts the estimated number of patients in the previous timestep from the corresponding point in the patient count time series; in other words, it finds the number of people it needs to move onto treatment in order to match the ART time series exactly. In doing so, it treats the count series as absolute truth and introduces a numerical problem. If, given the current parameter set, EPP predicts fewer PLHIV *total* than the ART time series expects to see *on treatment*, it cannot possibly allocate enough people to treatment to match the expected count. It must either invent new PLHIV and put them on treatment immediately or add as many PLHIV as it can and carry the difference into the future. Both of these approaches create undesirable discrepancies between the current model state and the dataset it explicitly assumes is true.

Thembisa

Thembisa is a highly detailed compartmental model of the South African HIV epidemic designed specifically for program evaluation (Johnson & Dorrington, 2019). It stratifies the South African population by demographic and sexual risk behaviour characteristics

and by exposure to HIV prevention programmes, resulting in a data-intensive model that can only infer a limited subset of parameters. After years of complementing model fitting with careful manual tuning, Thembisa predicts the South African epidemic accurately, but its generalisability to other countries is unknown, and, like EPP, it does not account for spatial dynamics.

Optima

The Optima model was designed to help facilitate efficient resource allocation in HIV programme planning (Kerr et al., 2015). Like Thembisa, it is a highly detailed compartmental model, but unlike Thembisa, its compartments are defined by the user and can be based on behavioural, demographic, or any other characteristics. Because of this level of detail, a single evaluation with only 10 population groups can take up to two seconds, making it ill-suited to statistical inference. It is calibrated to data via Markov Chain Monte Carlo (MCMC), but the details of the calibration process are only vaguely described.

Local Burden of Disease incidence estimation

Sartorius et al. (2021) used the estimates from Dwyer-Lindgren et al. (2019), which will be discussed shortly, to infer HIV incidence over space by fitting a modified version of EPP to synthetic prevalence data generated from a geostatistical model fit. Fitting one model to estimates from another requires the original model to fit well and introduces the complex problem of uncertainty propagation. If the original model underestimates posterior uncertainty, then fitting to its predicted posterior prevalence, as Sartorius et al. (2021) seem to have done, could underestimate the level uncertainty in incidence that the sparse, original data imply. The authors also fit EPP independently in each region, ignoring the fundamental spatial structure of the epidemic.

Finally, Sartorius et al. (2021) only had access to national-level ART patient count data, a gap that could lead them to understate spatial inequity in HIV incidence. To fill this gap, they conducted a literature review that found estimates of subnational ART coverage in 29 of 44 modelled countries. In countries with no subnational data, they assumed that ART coverage was constant over space. In countries with subnational data from the literature

review, they used a spatio-temporal statistical model to interpolate missing time points.

This approach is problematic because disparities in treatment availability are bound to be a key driver of spatial variation in incidence. Assuming ART coverage is constant over space will therefore almost certainly result in HIV incidence estimates that suppress critical subnational disparities. Further, for the reasons detailed in Section 1.3.3, conflating ART coverage among residents with ART patient counts can lead to biased estimates of prevalence.

1.5.3 Spatial models of HIV burden

A separate vein of work in HIV modelling has focused on the simpler problem of estimating HIV prevalence over space and time. We can categorise spatial models of HIV by their geographic units, denoted r. Recalling the two-stage sampling strategy described in Section 1.3.1, one class of models sets r to be individual survey clusters (point data methods) and the other sets r to be the regions in which the clusters were collected (areal data methods).

Point data methods

Point data methods generally fall into two categories: smoothing and geostatistical models. In both cases, each survey cluster is assumed to be an approximately five kilometre by five kilometre "pixel" that is small enough to be treated as point data. Models are fit to whichever pixels contain survey clusters, and predictions are made on the complete grid of five-by-five pixels.

Larmarange et al. (2011) proposed using both kriging and kernel density estimation methods to interpolate HIV prevalence over space, estimating HIV prevalence at s as a weighted average of prevalence at all observed points:

$$\hat{\rho}_r = \sum_i^N w_r(i)\rho_i, \tag{1.9}$$

where $w_r(i)$ is the weight of i with respect to r (a function of distance) as determined by a fitted semivariogram. More recent applications of kriging to HIV smooth covariates first and use the surfaces of covariates to predict prevalence:

$$\text{logit } \rho_r = \beta \cdot \mathbf{X}_r, \tag{1.10}$$

where **X** is a set of "kriged" covariate surfaces (Cuadros et al., 2017). These models have had success because of their speed and simplicity, but a 2016 model comparison study found that geostatistical methods consistently produced better out-of-sample fit (Subnational Estimates Working Group of the HIV Modelling Consortium, 2016).

The geostatistical models used to model HIV prevalence follow the methodology developed by the Malaria Atlas Project for malaria burden estimation (Gething et al., 2011). Bhatt et al. (2019) (the HIVE model) and Dwyer-Lindgren et al. (2019) (the Local Burden of Disease Model) are the two most prominent examples of this practice applied to HIV. In general, they assume that

$$\begin{aligned} Y_{r,t} &\sim \text{Binom}(N_{r,t}, \rho_{r,t}) \\ \text{logit } \rho_{r,t} &= \beta \cdot \mathbf{X}_{r,t} + Z_{r,t}, \\ \mathbf{Z} &\sim \text{N}(\vec{0}, \Sigma), \end{aligned} \tag{1.11}$$

where, $Y_{r,t}$ is the effective number of positive tests, $N_{r,t}$ is the effective sample size, **X** is a matrix of covariates that cover the entire grid we want to predict and Σ is a variance-covariance matrix for a Gaussian process, **Z**. The estimation of this covariance matrix is extremely intensive, particularly in cases in which space and time interact. To reduce the computational burden, we can assume that the covariance in space and time are separable and reformulate the model as

$$\mathbf{Z} \sim \text{N}(\vec{0}, \Sigma_s \otimes \Sigma_t). \tag{1.12}$$

This reformulation reduces the computational complexity of the model substantially while still allowing differential variation over space and time (Flaxman et al., 2015). This type of model can be fit using the integrated nested Laplace approximation via the **R-INLA** package (Lindgren, Rue, & Lindström, 2011; Martins et al., 2013; Rue, Martino, & Chopin, 2009).

Geostatistical models produce extremely granular estimates of HIV prevalence over space

with robustly quantified uncertainty, but, even after using the Kronecker product to estimate two covariance matrices, they are exceptionally intensive. Dwyer-Lindgren et al. (2019) were forced to model the African continent in four independent regions resulting in implausible border effects in their estimates (Dwyer-Lindgren et al., 2019). Additionally, the utility of such granular estimates has been called into question (Meyer-Rath et al., 2018). Policy is made at the administrative unit-level, so estimates from geostatistical models must be aggregated before they can be interpreted. In view of these concerns, the value of such granular estimates relative to their computational cost is questionable.

Areal data methods

Areal data methods operate on larger, pre-defined administrative units (Fuglstad, Li, & Wakefield, 2021). In these cases, r indexes a fixed region that is too large to treat as point data. Like geostatistical models, areal methods assume that logit $\rho_{r,t}$ is the weighted sum of covariates and some spatio-temporal term, but the spatial component must be constructed using different methods.

One of the simplest areal models uses shrinkage priors, assuming that each region is an independent sample from a shared national distribution:

$$\begin{aligned} Y_r &\sim \text{Binom}(N_r, \rho_r) \\ \text{logit } \rho_r &= \beta \cdot \mathbf{X}_r + v_r, \\ v_r &\sim \text{N}(v_0, \sigma_v). \end{aligned} \tag{1.13}$$

Here, v_r is a region-specific random effect and v_0 is a cross-region mean. This model assumes that each v_r is an independent and identically distributed (i.i.d.) sample from $\text{N}(v_0, \sigma_v)$ and therefore makes no explicit assumptions about spatial structure.

The i.i.d. assumption in the shrinkage model is convenient but often unrealistic, so authors have explored autoregressive models for HIV mapping (Gutreuter et al., 2019). In general, autoregressive models add structure to the prior on region-specific random effects, but there are numerous variations on the exact form. Gutreuter et al. (2019) use a simultaneously autoregressive model

$$v \sim \mathrm{N}(\vec{0}, \Sigma), \tag{1.14}$$

where Σ is a variance-covariance matrix:

$$\Sigma = \sigma_v^2 \left[(\mathbf{I} - \rho \mathbf{W})(\mathbf{I} - \rho \mathbf{W}^\top) \right]^{-1}. \tag{1.15}$$

Here, $\mathbf{I}$ is an identity matrix with the same dimensions as Σ, ρ is an estimated correlation parameter, and $\mathbf{W}$ is an $m \times m$ matrix such that for region j adjacent to r

$$\mathbf{W}_{r,j} = \frac{1}{k_r}, \tag{1.16}$$

where k_r is the number of regions directly adjacent to r (Gutreuter et al., 2019). This model allows for a degree of correlation between v_r and v_j for all j adjacent to r. Gutreuter et al. (2019) found that the simultaneously autoregressive model improved fit well to prevalence data from South Africa relative to a model with i.i.d. random effects.

Naomi Most recently, Eaton et al. (2021) proposed Naomi, a model for district-level estimating HIV burden in SSA that uses small-area regression methods to synthesise survey estimates of HIV prevalence, ART coverage, and recency with ANC data and ART patient counts. Of particular note for this work, Naomi connects small-area models of ART coverage and HIV prevalence and incidence to ART patient counts with a model of ART attendance that allows individuals to seek treatment outside of their regions-of-residence.

For simplicity, I omit the age, sex, and time indices from Eaton et al. (2021) when describing their model. Naomi predicts prevalence and ART coverage in region s using a reparametrisation of the Besag-York-Mollié model presented by Riebler et al. (2016), often referred to as the BYM2 model. Simplifying to space alone, we have

$$\begin{aligned} \operatorname{logit} \rho_r &= v_r \\ v &\sim \mathrm{BYM2}(\sigma_\rho, \phi_\rho), \end{aligned} \tag{1.17}$$

where σ and ρ are hyperparameters for the BYM2 model. They use a similar model for

ART coverage.

Given prevalence, ART coverage, and the number of residences ρ_r, α_r, and N_r, respectively, Eaton et al. (2021) estimate the number of residents of r receiving treatment as

$$\hat{A}_r = N_r \rho_r \alpha_r. \tag{1.18}$$

Because ρ_r and α_r are both fit to household survey data, they specifically apply to residents of r, while the facility-based programmatic data are in reference to people who receive treatment in r.

To account for the possibility that PLHIV seek treatment outside of their home regions, Eaton et al. (2021) proposed a model of ART attendance that maps estimated residents on ART in any region r to adjacent regions. Let $\{j : j \sim r \wedge j \neq r\}$ be the set of regions such that j and r share a border, exclusive of r. I will use $j \sim r \setminus r$ as shorthand for this set. Naomi finds the log-odds that an individual residing in r will seek treatment in j instead of r as

$$\begin{aligned} \log \gamma_{r,j} &= \delta_0 + \delta_j \\ \delta_0 &= -4 \\ \delta_j &\sim \mathrm{N}(0, \sigma_\delta), \end{aligned} \tag{1.19}$$

where $\delta_0 = -4$ corresponds to a prior of 93% of individuals in a region with four neighbours seeking treatment in their home region. As in multinomial regression, the authors fix $\gamma_{r,r} = 1.0$ to maintain identifiability and use the softmax function to find the the share of ART patients residing in r who seek treatment in all $k \in j \sim r \setminus r$:

$$\pi_{r,j} = \frac{\gamma_{r,j}}{1 + \sum_{k \in j \sim r \setminus r} \gamma_{r,j}}. \tag{1.20}$$

They then allocate all ART patients to a region-of-treatment according to their region-of-residence, assuming that $\pi_{r,j} = \pi_{j,r} = 0$ if r and j are not adjacent:

$$\hat{P}_r = \sum_{j=1}^{N} \pi_{j,r} \hat{A}_j. \tag{1.21}$$

They then use $\hat{P}_r$ as the mean for their observation model of facility-based ART patient

count reports.

This ART attendance model serves an important role in reconciling the differing sampling frames of the household survey data and the facility-based patient counts. If a region's facilities attract a disproportionate number of patients, then, given a certain level of survey-estimated prevalence, implied ART coverage might be artificially inflated.

Anecdotally, we can see the effects of this in Malawi. Médecins Sans Frontières (MSF) began operating treatment clinics in the high-prevalence Chiradzulu district several years before Malawi's broader national treatment programme was scaled up, leading PLHIV in nearby regions (including the large city of Blantyre) to seek treatment in Chiradzulu (Médecins Sans Frontières, Malawi & Ministry of Health and Population, Chiradzulu District, Malawi, 2004). If we were to calculate ART coverage in Chiradzulu and nearby districts by comparing PLHIV estimates from a household survey to ART patient count data, we would see high ART coverage in Chiradzulu and low coverage in nearby areas.

The primary difficulty in the Naomi ART attendance model is that we do not directly observe the flows between districts. For a given region r, we only observe $\sum_{j=1}^{N} \pi_{j,r}\hat{A}_j$, as opposed to every pairwise $\pi_{j,r}\hat{A}_j$. The combination of the informative prior on δ_j and the household survey estimates of ART coverage from PHIA identifies the model in many cases, but it will always be impossible to validate.[2]

In this way, Naomi improves substantially on other models. It is, however, fundamentally cross-sectional. To produce estimates over time, it takes a maximum of two, h-year steps with a compartmental model, essentially linearising the epidemic. For example, if we have data in 2016 and want estimates in 2019 and 2022, Naomi produces estimates in 2016, calculates the derivative of a compartmental model model at 2016, and takes three-year steps with a simple single-step projection method. If we are interested in estimating trends in HIV incidence, then this method is likely to be too coarse and prone to integration error.

[2] Identifiability is difficult to assess formally here, but we can use intuition to understand how it might work. Because the PHIA surveys estimate household-based ART coverage, they imply a certain number of residents on treatment. If this number of residents on treatment conflicts with facility-based ART patient counts, then, very loosely speaking, the model will either send patients to or pull patients from adjacent regions. The prior on δ_j should build into the model a "reluctance" to move individuals from their home regions.

Other spatial models of infectious disease

The epidemiological and demographic peculiarities of HIV have limited the spatial scope of inferential models of incidence, but we can look to work in other areas of infectious disease modelling for wisdom. In particular, the Epidemic/Endemic and TSIR models presented by Held, Höhle, & Hofmann (2005) and Wakefield, Dong, & Minin (2017), respectively, offer attractive frameworks for spatial modelling of infectious disease. The two models were developed separately, but they arrive at nearly identical formulations.

Let $Y_{r,t}$ be the number of disease cases observed in region r at time t via some relatively complete surveillance system. Following Section 1.5.1 of Wakefield, the TSIR model is

$$\begin{aligned} Y_{r,t} &\sim \text{NegBinom}(\mu_{r,t}, \phi) \\ \mu_{r,t} &= \left[e^{\lambda_t^{AR}} y_{r,t-1} + e^{\lambda^{NE}} N_r^{\tau_1} \sum_{j=1}^{n} w_{r,j} y_{j,t-1}^{\tau_2} \right]^{\alpha} + N_r e^{\lambda^{EN} + b_r^{EN}}, \end{aligned} \tag{1.22}$$

where $\mu_{i,t}$ can be broken down into infections originating from within r, infections originating from outside of i, and infections originating from outside of the system entirely (Wakefield, Dong, & Minin, 2017). Structurally, the Epidemic/Endemic model is nearly identical.

We are particularly interested in the "epidemic" term in which the spatial weight $w_{r,j}$ is defined as

$$w_{r,j} = \frac{d_{r,j}^{-\theta_1/(1-\theta_1)}}{\sum_{k \neq r} d_{r,k}^{-\theta_1/(1-\theta_1)}}. \tag{1.23}$$

The TSIR model defines $d_{r,j}$ to be the distance between r and j, while Held, Höhle, & Hofmann (2005) define it to be the adjacency-based graph distance between r and j. Both the TSIR model and the Epidemic/Endemic model rely on assumptions that make them poorly suited to HIV (for example, that we can define a time interval after which an initially infectious individual will almost certainly no longer be infectious), but we can use their model of spatial risk as an example.

1.5.4 Discussion

This chapter has provided a brief overview of the data sources available for inferring HIV incidence, outlined the basics of compartmental epidemic modelling, and highlighted a number of previous models of HIV incidence and other key indicators. Based on this review, I have made several observations about the data and modelling strategies used in HIV burden inference.

First, there is an untapped wealth of data that could help inform spatially resolved estimates of key HIV indicators in most sub-Saharan African countries. Contrary to some common wisdom, population-level HIV epidemiology is rapidly becoming a data-rich field. The Malawi Department of HIV & AIDS, for example, has collected monthly reports on HIV testing and treatment provision from nearly 800 ANC sites since the beginning of 2011. For comparison, the UNAIDS estimation process in Malawi incorporates annual data from 55 facilities, which themselves were originally selected because are situated in high-prevalence areas.

Similarly, the methods used to incorporate general population ART patient count data, which are available at a similar resolution to the ANC data, could be improved. The solutions presented by models like EPP induce complex numerical challenges and necessarily treat the patient counts as fixed, not as noisy data. There are particular challenges involved in relating reported patient counts to general-population estimates and household survey data, but the Naomi model provides a blueprint for how to proceed.

Setting concerns about data aside, the set of existing inferential models of population-level HIV indicators fall broadly into two distinct categories: epidemic models that infer incidence in closed populations and statistical models that interpolate prevalence over space and time. Over the past decade, EPP has repeatedly proved to be a reliable tool for inferring HIV incidence from national-level prevalence data, but its assumptions are only realistic in large, closed populations. On the other hand, Dwyer-Lindgren et al. (2019), Gutreuter et al. (2019), and Eaton et al. (2021) all present methods that produce accurate maps of important indicators over space and, in some cases, time. The only models that actually bridge the gap between these two worlds are Sartorius et al. (2021) and Eaton et al. (2021). Both models take advantage of the fundamental principles of infectious disease

epidemiology described in Section 1.5.1 to relate prevalence to incidence, but both could be improved in a variety of ways.

In my PhD, I bridge the gap between epidemic models and spatially structured statistical models in a way that offers epidemiologically plausible estimates of HIV incidence over space and time. Taking ideas from Naomi and EPP specifically, I use small-area methods to estimate the initial state of a compartmental epidemic model of HIV, which then produces internally consistent trajectories of incidence, prevalence, and ART coverage that I use to fit to a number of different datasets.

This model presents several methodological innovations. First, using a small-area statistical methods to infer the initial state of the epidemic model allows computation to begin the year data become available. Second, I propose stochastic time series models for two important underlying dynamics of the epidemic model: regions-specific HIV transmission rates and ART initiation probabilities. Further, by directly modelling ART initiation and incorporating the ART attendance model from Eaton et al. (2021), I am able to avoid treating ART patient count data as noiseless and fixed. Finally, I build this model into a relatively fast, approximate inference framework that allows the model to fit in a reasonable amount of time.

1.6 Document outline

The remainder of this thesis provides an account of the model of HIV incidence I have developed and demonstrates its use in Malawi. Chapter 2 provides a detailed description of the spatio-temporal model of HIV incidence. Chapter 3 reports methods and results for an extensive set of specification tests conducted on the incidence model. In Chapter 4, I estimate the computational and epidemiological impact of explicitly modelling transmission across small areas. Chapter 5, provides estimates of key HIV epidemic indicators for the 28 districts of Malawi. Chapter 6 examines the effect of increasing forecasting horizons on current and future estimates of incidence. Chapter 7 reproduces a previously published article developing a modelling strategy for self-reported partner age distributions (Wolock et al., 2021). Finally, Chapter 8 provides a unifying discussion and proposes several directions for future work.

1.7 Software

Where applicable in the text, I cite the specific statistical software used to produce estimates. All analyses were conducted using R (R Core Team, 2013). All plots were produced using the **ggplot2** R library (Wickham, 2016). I constructed hexagonal tile maps of Malawi using the **geogrid** R library (Bailey, 2018). This document was prepared using the `reedthesis` template from the **thesisdown** R library (Ismay & Solomon, 2021).

Chapter 2

A Bayesian Spatio-Temporal Epidemic Model of HIV

As described in Section 1.5, existing models of HIV burden are not well-suited to infer incidence over space and time simultaneously. Spatially resolved models have focused on estimating prevalence, while models of incidence usually operate on whole countries. However, these models provide foundations for further development. Using concepts from the EPP and Naomi, I have developed a spatio-temporal epidemic model of HIV that synthesises data from household surveys, ANC facilities, and treatment programmes to infer sex-specific adult HIV incidence over relatively small geographic areas.

In this chapter, I identify the data sources my model uses, provide a detailed description of the epidemic model used to simulate epidemic indicators from predictions of underlying dynamics, describe the observation model used to relate predicted epidemic indicators to data, and, finally, outline the computational strategy used to perform inference.

2.1 Inference goal

The goal of this work is to infer quarterly sex-specific HIV incidence among adults aged 15-49 years for some set of subnational administrative units beginning before ART scale-up (approximately 2005 in most SSA countries) and ending at the present day. I restrict to adults aged 15-49 because that is the age range of both the DHS and EPP. Ministries of

health allocate their budgets to administrative units (often districts, which comprise the second administrative or "admin 2" level), so that is my unit of analysis. Let $\lambda_{r,g}(t)$ denote HIV incidence among people of sex g in region r at time t.

For the reasons detailed in Section 1.3.4, HIV incidence is difficult to observe directly, so this work uses a compartmental epidemic model to infer epidemic dynamics using data on observable indicators. I will refer to the model used to simulate prevalence, ART coverage, and incidence from parameters as the *process model* and to the parameters for this model, denoted θ_P, as the *process parameters*. Process parameters are used to simulate epidemic indicators, which are related to data via an *observation model*. Like the process model, the observation model requires a set of parameters, θ_O, such as the parameters that relate general population prevalence to ANC facility prevalence.

I have built this model using a Bayesian framework. The ability to set priors allows for the incorporation of estimates and knowledge from decades of research on HIV and can help identify parameters that might otherwise be non-identifiable. The Bayesian framework is also an intuitively appealing paradigm for the datasets used in this work. We can assume that our data are samples from true, underlying distributions, which are our inference targets.

Given a set of data, $\mathcal{D}$, collated over space, time, data source, and sex, the goal is standard Bayesian inference, that is, to obtain the posterior distribution of the two parameter sets:

$$\mathrm{P}(\theta_P, \theta_O \mid \mathcal{D}). \tag{2.1}$$

Bayes' theorem gives us that

$$\mathrm{P}(\theta_P, \theta_O \mid \mathcal{D}) = \frac{\mathrm{P}(\mathcal{D} \mid \theta_P, \theta_O)\mathrm{P}(\theta_P, \theta_O)}{\mathrm{P}(\mathcal{D})}. \tag{2.2}$$

Because $\mathrm{P}(\mathcal{D})$ does not depend on parameters, we know that, with respect to both parameter sets,

$$\mathrm{P}(\theta_P, \theta_O \mid \mathcal{D}) \propto \mathrm{P}(\mathcal{D} \mid \theta_P, \theta_O)\mathrm{P}(\theta_P, \theta_O). \tag{2.3}$$

Indicator	Data Source	Numerator	Denominator
Prevalence	Household surveys	# of positive HIV tests	# of HIV tests
ANC prevalence	Sentinel ANC clinics	# of positive HIV tests	# of HIV tests
ART coverage	Household surveys	# of positive ART tests	# of ART tests
ART patients	Programmatic data	# of ART patients	-
Recency status	Household surveys	# of positive recency assays	# of recency assays

Table 2.1: Taxonomy of population-level HIV data sources included in my model

I further assume that θ_P and θ_O are independent, giving us

$$\mathrm{P}(\theta_P, \theta_O \mid \mathcal{D}) \propto \mathrm{P}(\mathcal{D} \mid \theta_P, \theta_O)\mathrm{P}(\theta_P)\mathrm{P}(\theta_O). \tag{2.4}$$

I define priors on the process parameters in Section 2.4, and define the likelihood, $\mathrm{P}(\mathcal{D} \mid \theta_P, \theta_O)$, and priors on parameters for the observation model, $\mathrm{P}(\theta_O)$, in Section 2.5.

2.2 Data sources

The collated set of data, $\mathcal{D}$, considered by this model consists of a subset of the candidate sources described in Section 1.3: household surveys, ANC facility HIV test results, and ART programme patient counts. Here, I describe the various indicators and resolutions of these sources. For clarity, I defer the definition of the likelihood until after the mechanics of the process model have been described (Section 2.5). Table 2.1 outlines the data sources and the indicators they measure.

2.2.1 Household surveys

From household surveys, we obtain estimates of HIV prevalence, ART coverage, and the proportion of recent infection. Let $r \in \{1, ..., R\}$ denote a region, $g \in \{0, 1\}$ denote sex, with 0 representing men and 1 women, and s denote a specific data source. In the case of household surveys, s represents a survey (e.g. the 2015-2016 MPHIA survey in Malawi).

Let $p^{s,a}_{r,t,g}$ be the estimated proportion from survey s of individuals of sex g exhibiting characteristic a among people in region r at time t. For household surveys, a can be HIV prevalence, antiretroviral treatment status, or recent infection, denoted $\{\text{HIV}, \text{ART}, \text{Rec}\}$, respectively. Each survey is assigned to the midpoint of its collection period, so each $s$corresponds to exactly one t, making t redundant.

As discussed in Section 1.3.1, each $p^{s,a}_{r,t,g}$ comes with a design-based variance, $\tau^{s,a}_{r,t,g}$, calculated according the survey's design. The design effect of the proportion is

$$D_{\text{eff}} = \frac{p^{s,a}_{r,t,g}(1 - p^{s,a}_{r,t,g})/T^{s,a}_{r,t,g}}{\tau^{s,a}_{r,t,g}}. \tag{2.5}$$

We can solve for the effective sample size, $T^{s,a}_{r,t,g}$, of $p^{s,a}_{r,t,g}$ as

$$T^{s,a}_{r,t,g} = \frac{p^{s,a}_{r,t,g}(1 - p^{s,a}_{r,t,g})}{\tau^{s,a}_{r,t,g}}, \tag{2.6}$$

and find the corresponding effective number of positive tests: $P^{s,a}_{r,t,g} = p^{s,a}_{r,t,g} T^{s,a}_{r,t,g}$.

From each DHS that conducted HIV testing, we can extract region-/sex-specific pairs of effective positive tests and effective number of tests: $(T^{s,\text{HIV}}_{r,t,g}, P^{s,\text{HIV}}_{r,t,g})$. From each PHIA survey, we have data on HIV positivity, $(T^{s,\text{HIV}}_{r,t,g}, P^{s,\text{HIV}}_{r,t,g})$, treatment status, $(T^{s,\text{ART}}_{r,t,g}, P^{s,\text{ART}}_{r,t,g})$, and recent infection $(T^{s,\text{Rec}}_{r,t,g}, P^{s,\text{Rec}}_{r,t,g})$.

2.2.2 ANC facilities

In the case of HIV testing data from ANC facilities, we have counts of tests and positive tests. Let $T^{s,\text{HIV}}_{r,t,1}$ be the number of HIV tests conducted at ANC facility or group of facilities s during time period t and $P^{s,\text{HIV}}_{r,t,1}$ be the number of those tests that were positive. Each facility or group of facilities s is located in exactly one region r, so r is redundant.

2.2.3 ART programme data

The final source of data used here is the routinely reported ART patient counts described in Section 1.3.3. For facility s, we have the number of ART patients treated at that facility

in time period t, denoted $C^s_{r,t}$. If these series cannot be restricted to ages 15-49, they must be adjusted for the share of adults receiving treatment that are 50 years old and older. I have done so by applying the time-varying estimated proportion of PLHIV on treatment in Malawi that are between 15 and 50 years old from Spectrum to the observed patient count series. Further, these data are typically not available disaggregated by sex. My model does not include any detail about treatment seeking behaviour within region, so I aggregate these series to the region level, giving us $C^s_{r,t}$.

2.3 Model overview

The model predicts the dynamics necessary to simulate an epidemic model as non-linear functions of time, space, and sex (Hastie & Tibshirani, 1986), aggregates the epidemic model's projections to produce estimates, and uses the observation model to relate those estimates to data. For a single draw from the posterior density:

1. A set of process parameters, θ_P, are used to model region-/sex-/time-specific series of HIV transmission rates, ART initiation rates, and initial prevalence.
2. The epidemic model is initialised at the state determined by the estimated initial prevalence from (1) and integrated using the estimated transmission rates, ART initiation rates, and a set of exogenous, fixed parameters.
3. Predictions from the epidemic model are aggregated to produce estimates of HIV prevalence, ART coverage, and ART patients at the same spatio-temporal resolution as each dataset.
4. Predicted HIV prevalence, incidence, and ART coverage are used with an additional set of parameters, θ_O, to evaluate the observation model given a collated dataset $\mathcal{D}$.

Figure 2.1 presents a simplified representation of the model. The first step from the above list is represented by every node to the left of M, the second step is M, and the third step is to the right of M.

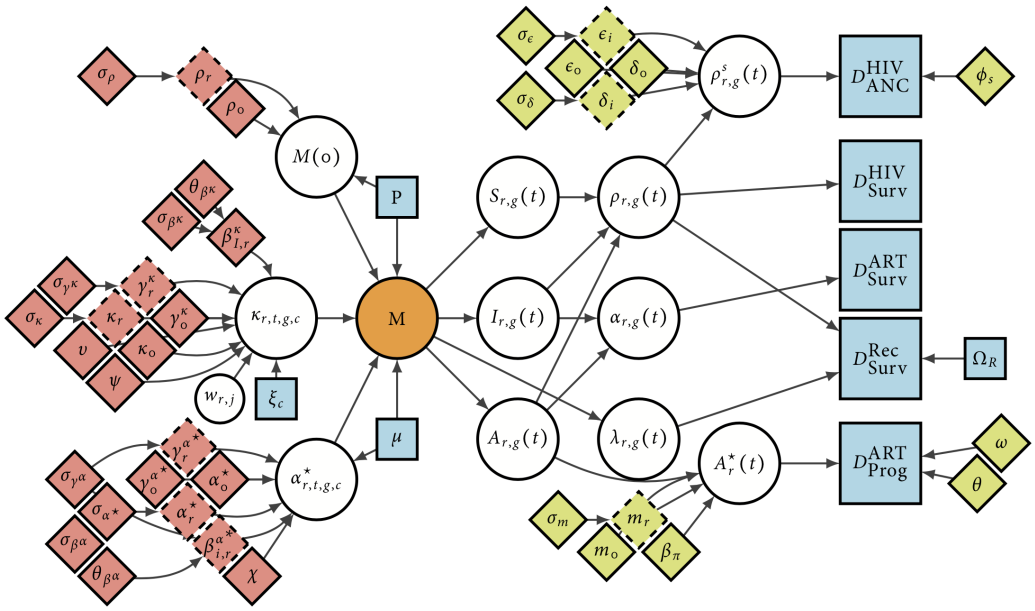

Figure 2.1: A simplified graphical representation of my model of HIV incidence. Diamonds are parameters, white circles are (deterministic) calculations, and blue squares are external data. Red parameters influence the process model, and green/yellow parameters influence the observation model.

2.4 Process model

I use a compartmental epidemic model with inputs predicted by generalised additive models to simulate internally consistent series of HIV incidence, prevalence, and ART coverage among men and women aged 15-49 given a set of process parameters, θ_P. This section provides a detailed description of how θ_P is used to produce those series.

2.4.1 Compartmental model of HIV

I use a deterministic compartmental model of HIV to simulate HIV prevalence and incidence, ART coverage, and the number of PLHIV receiving treatment ($\rho_{r,g}(t)$,$\lambda_{r,g}(t)$, $\alpha_{r,g}(t)$, and $A_{r,g}(t)$, respectively). I model the number of people of sex g in region r at time t by disease status with the following set of ordinary differential equations:

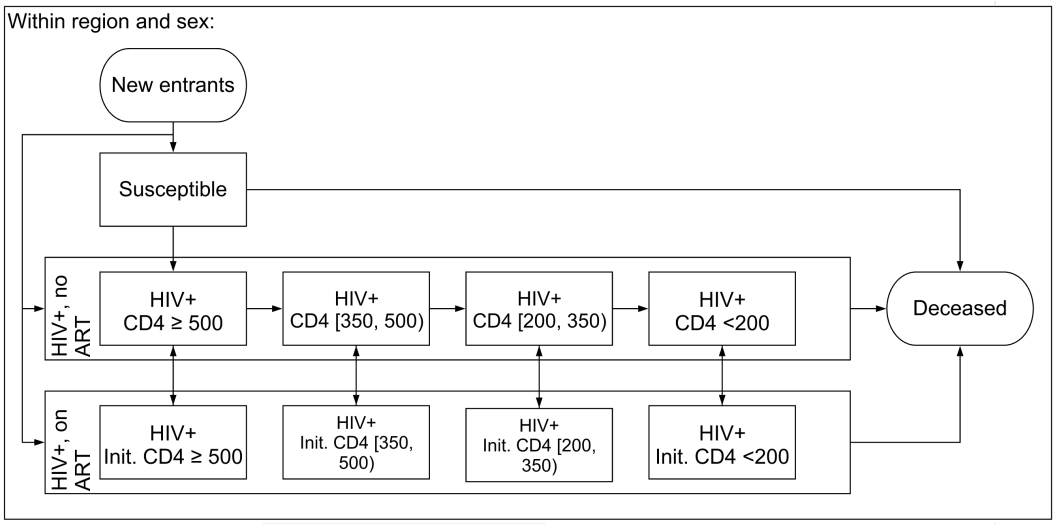

Figure 2.2: Diagram of compartmental model of HIV used in this analysis

$$\begin{aligned}
\frac{\partial S_{r,g}(t)}{\partial t} =& S_{r,g}(t) \cdot (-\lambda_{r,g}(t) - \mu^S_{t,g}) + E^S_{r,t,g} \\
\frac{\partial I_{r,g,c}(t)}{\partial t} =& I_{r,g,c}(t) \cdot (-(\mu^S_{t,g} + \mu^I_{t,g,c}) - \alpha^*_{r,t,g,c} - \iota_{g,c,t}) + \lambda_{r,g,c}(t) S_{r,g}(t) + \\
& \eta A_{r,g,c}(t) + \iota_{g,c-1} I_{r,g,c-1}(t) + E^I_{r,t,g,c} \\
\frac{\partial A_{r,g,c}(t)}{\partial t} =& A_{r,g,c}(t) \cdot (-\mu^S_{t,g} - \mu^A_{t,g,c} - \eta) + \alpha^*_{r,t,g,c} I(t) + E^A_{r,t,g,c}.
\end{aligned} \tag{2.7}$$

I denote the number of susceptibles $S_{r,g}(t)$, the number infected in disease stage c without treatment $I_{r,g,c}(t)$, and the number infected with treatment who began treatment at disease stage c $A_{r,g,c}(t)$. I define c to be one of four CD4 compartments, consistent with those defined by the Thembisa model. Figure 2.2 outlines the structure of disease progression in the model. A susceptible individual can either die at rate $\mu^S_{t,g}$ or become infected by the opposite sex at rate $\lambda_{r,g}(t)$. An infected individual without treatment can die at rate $\mu^S_{t,g} + \mu^I_{t,g,c}$, begin treatment with probability $\alpha^*_{r,t,g,c}$, or progress to the next disease stage at rate $\iota_{g,c,t}$. Finally, an individual on treatment can die at rate $\mu^S_{t,g} + \mu^A_{t,g,c}$ or interrupt treatment with annual probability η. Loss to follow up is difficult to measure, so I have fixed η to be 6% annually, adjusting a published figure to account for improvements in the treatment programme (Yu et al., 2007).

I subdivide PLHIV with and without treatment into four CD4 categories to accurately model the survival distribution from HIV infection to AIDS-related death and to be able to capture changing eligibility for ART in which treatment was restricted to those with the

lowest CD4 counts. A less complex compartmental model that omits these subdivisions could perform as well, particularly in the presence of reliable cause-specific mortality data. However, reliable mortality data are not available in most SSA countries. By subdividing by CD4 category, we can inform predicted mortality and ART initiation with estimated mortality parameters from other studies that are, in theory, biologically constant.

I calculate the CD4 stage progression rates, $\iota_{g,c,t}$, using assumptions described by Johnson & Dorrington (2019). Taking the average time spent in each CD4 category from Table 3.1, I apply Equation 3.1 to the Spectrum-estimated year-/sex-specific average age of PLHIV between 15-49:

$$\iota_{g,c,t} = \iota_c 0.96^g (1+k)^{(x-30)/10}, \tag{2.8}$$

where ι_c is the average annual rate of progression from category c to $c+1$, 0.96 is the progression rate of women relative to men, and $k = 0.18$ is the proportionate increase in progression associated with a ten-year increase age. Following Table 3.1, I fix $\iota = (3.16, 2.13, 3.20)$ to be the expected number of years spent in CD4 category without treatment for the three highest CD4 categories. The final category is terminal, so its expected duration is not defined. The age distribution of PLHIV from Spectrum is itself an estimate and therefore might be inaccurate or contain uncertainty, which is not considered here; the approach described above is a simple solution to account for the fact that the model does not currently include age structure.

To account for the shifting age distribution of PLHIV and gradual improvement of HIV patient outcomes without treatment in SSA, I calculate age-adjusted mortality rates based on the assumptions and results from EPP-ASM (Eaton et al., 2019). I use input mortality rates and predicted counts of PLHIV by age in each CD4 bin to find year-/age-/sex-/CD4-specific expected death counts. I aggregate the death and population counts to align them with the year-/sex-/CD4 groups used here and recalculate the mortality rates.

Each E term in Equation (2.7) is the net number of people ageing in or out of the 15-49 year old population. HIV prevalence varies systematically with age in sub-Saharan Africa, so we must consider the possibility that the distributions of people ageing in and out across compartment vary from those of the general population. For example, if the population of

PLHIV is ageing, we would expect prevalence among people ageing out to increase over time.

I use population data and national age-/sex-specific estimates of the share of people in each disease compartment from Spectrum to obtain regional estimates of the number of people turning 15 and 50 years old by sex, $E^{15}_{r,t,g}$ and $E^{50}_{r,t,g}$, respectively. The national-level estimates from Spectrum inherently weigh each region proportionately to its population, so if the regional distribution of individuals across compartment is correlated with population, the national-level disease status distributions will represent smaller regions poorly. I therefore apply region-specific adjustments to the estimated Spectrum distributions at each time point. Specifically, I adjust the odds from Spectrum of being in substage c of compartment C relative to being susceptible with the model's current regional prediction. Taking people ageing in as an example, the adjusted odds of an individual being in CD4 bin c relative to not being infected with HIV are:

$$\Delta^{15,C}_{t,g,c} = \frac{C_{r,g,c}(t)}{S_{r,g}(t)} \frac{C^{15,\text{Spec}}_{g,c,t}}{S^{15,\text{Spec}}_{g,t}}, \tag{2.9}$$

noting that the denominators have cancelled. Then, fixing $o^{15,S}_{t,g} = 1.0$, I solve for the proportion of people in each compartment:

$$\nu^{15,C}_{t,g,c} = \frac{\Delta^{15,C}_{t,g,c}}{1 + \sum_{J \in I,A} \sum_{i=1}^{4} \Delta^{15,J}_{t,g,i}} \tag{2.10}$$

Finally, I calculate the net numbers of people ageing in and out for a given compartment, C as

$$E^{C}_{r,t,g,c} = \nu^{15,C}_{t,g,c} E^{15}_{r,t,g} - \nu^{50,C}_{t,g,c} E^{50}_{r,t,g}. \tag{2.11}$$

This method accounts for the possibility that the distributions of people ageing into or out of the population across compartment vary from that of the general population.

Several inputs to this model are fixed at values defined by exogenous datasets or other models. Specifically, the progression rates through CD4 categories, mortality rates, region-/sex-specific population size, distribution of entrants and exits across disease stage, and the

rate of loss-to-follow-up among persons on ART are all fixed assumptions derived from other data sources. Because the objective is to infer incidence, which depends explicitly on transmission rates, prevalence, and ART coverage, I focus on inferring the parameters that determine those quantities.

The parameters that govern progression across CD4 categories and AIDS-related mortality are derived from cohort studies and are assumed to be biological constants. The model predicts AIDS mortality conditional on CD4 category, treatment status, sex, and average age of PLHIV, a level of stratification at which the assumption of spatial homogeneity might be reasonable. Treating input mortality and progression parameters as priors and inferring spatially varying rates would be preferable, but population-level measurements of AIDS-related mortality and CD4 counts are scarce in SSA.

I also consider region-level population as fixed, although population estimation is its own uncertain process (United Nations, Department of Economic and Social Affairs, Population Division, 2019). Unlike the mortality and progression parameters, input population estimates are regionally resolved. Without age structure, the population levels projected by the model could differ substantially from other estimates, so I match input population at every time step. An age-structured model could infer fertility and mortality by fitting to census data, allowing the user to estimate population levels directly (Wheldon et al., 2013).

I integrate the epidemic model using the forward Euler method with a step size of 0.25 years. Although forward Euler is a crude numerical integrator, I use it to minimise computational cost. Because HIV changes slowly over the course of years, Euler might be an appropriate choice. I also note that there is a direct correspondence between an epidemic model integrated with forward Euler and widely used discrete-time epidemic models; this model could be viewed as a quarterly discrete-time model governed by annual rates that are being scaled at projection-time.

Aside from numerical error, one deficiency of using the Euler method with relatively large steps is that this model can never actually reach 100% ART coverage. Let h be the size of the time step in years. Then a fixed proportion of people on ART are removed at each time step, so the maximum ART coverage the model can estimate is $1 - h\eta$, where η is the annual proportion of people lost to follow-up. As the step size decreases, this quantity will get closer to 1.0. In cases where ART coverage could be approaching 100%, smaller step

sizes should be considered.

I define prevalence as

$$\rho_{r,g}(t) = \frac{\sum_{c=1}^{4}[I_{r,g,c}(t) + A_{r,g,c}(t)]}{S_{r,g}(t) + \sum_{c=1}^{4}[I_{r,g,c}(t) + A_{r,g,c}(t)]} \tag{2.12}$$

and ART coverage as

$$\alpha_{r,g}(t) = \frac{\sum_{c=1}^{4} A_{r,g,c}(t)}{\sum_{c=1}^{4}[I_{r,g,c}(t) + A_{r,g,c}(t)]}. \tag{2.13}$$

With the compartmental model defined, I will define the models for incidence, ART initiation, and the initial state of the model, denoted $\lambda_{r,g}(t)$, $\alpha^*_{r,t,g,c}$, and $(S_{r,g}(0), I_{r,g}(0), A_{r,g}(0))$, respectively, for all $r \in \{1, ..., R\}$ and $g \in \{0, 1\}$.

2.4.2 Underlying linear models

The epidemic model detailed in Equation (2.7) relies on a number of quantities that we either cannot or do not observe directly. I have assumed that a subset of them (mortality, disease progression, etc.) are fixed and drawn from other data sources and models, but three key components (incidence, ART initiation rate, and the initial state) are inferred from data. Each of these quantities is represented by an underlying generalised additive model (Hastie & Tibshirani, 1986). The epidemic model uses the predictions of each underlying linear model to simulate predicted values that can be compared to observations. In this way, despite not directly observing quantities like region-specific HIV transmission rates, the model is able to infer incidence from measurement of other indicators.

Model of HIV incidence

I model incidence, $\lambda_{r,g}(t)$, as a log-linear function of time-varying transmission rates, opposite-sex prevalence, and ART coverage. Log-linear models fit naturally into the classical ODE-based epidemic modelling framework. Recalling the simple SIR model from Equation (1.3), we defined new infections as $\beta SI/N$, resulting in incidence being $\lambda = \beta I/N$. Because incidence is only constrained to be greater than zero, we can find

$$\log \lambda = \log \beta + \log \frac{I}{N}. \tag{2.14}$$

In this case, N is fixed and I is determined by the system of ODEs, so $\log \beta$ can be viewed as the intercept in a log-linear model of incidence.

Casting the classical SIR model as naturally incorporating linear submodels is a useful change in perspective because we can augment this model with any of the familiar hierarchical modelling tools described in Section 1.5.3. If, for example, we are estimating region-specific epidemics, we can use a hierarchical model to borrow statistical strength across regions:

$$\begin{aligned} \log \lambda_r &= \log \beta_r + \log \tfrac{I_r}{N_r} \\ \log \beta_r &\sim \mathrm{N}(\log \beta_0, \sigma^2) \end{aligned} \tag{2.15}$$

Here, we assume that each region's log-transmission rate, $\log \beta_r$, is normally distributed around a shared mean, $\log \beta_0$.

Because we do not observe λ_r directly, it will be more precisely identified in regions with more observations or in regions with observations that are more easily described by the epidemic model. In other words, this model allows regions with less identifiable series of incidence to share statistical strength from regions with more easily identified incidence series.

Returning to HIV, I model the incidence rate among members of sex g as a log-linear function of prevalence and ART coverage among the opposite sex. Omitting spatial transmission dynamics briefly, the log-incidence rate among members of sex g residing in region r attributable to members of the opposite sex, g^* in CD4 bin c is:

$$\begin{aligned} \log \lambda_{r,g,c}(t) = {} & \log \xi_c + g \cdot (\psi + \upsilon t) + \log \kappa_{r,t} + \\ & \log \left[(I_{r,g^*,c}(t) + (1-\omega) A_{r,g^*,c}(t)) / N_{r,g^*} \right], \end{aligned} \tag{2.16}$$

where ξ_c is the relative infectiousness of stage c, $\psi + \upsilon t$ is an inferred time-varying, sex-specific infectiousness log-ratio for women, $\kappa_{r,t}$ is a region- and time-specific transmission rate, and $\omega = 0.2$ is the relative infectiousness of individuals on treatment. The first three

terms of this equation, $\log \xi_c + g \cdot (\psi + \upsilon t) + \log \kappa_{r,t}$, are equivalent to $\log \beta$ from Equation (2.14), except that they allow for variation in transmission rates across sex, CD4 bin, time, and space. The final term, $\log\left[(I_{r,g^*,c}(t) + (1-\omega)A_{r,g^*,c}(t))/N_{r,g^*}\right]$, is log-transformed prevalence, which is equivalent to $\log I/N$ from Equation (2.14), adjusted for the strong preventative effect on transmission of ART (Eisinger, Dieffenbach, & Fauci, 2019). Excluding the inferred sex ratio of transmission, this formulation is identical to EPP.

The model can be extended to account for spatial mixing allocating the number of potentially infectious contacts from each region to itself and its neighbours. Specifically, we have:

$$\begin{aligned} \log \lambda_{r,g,c}(t) \quad = \quad & \log \xi_c + g \cdot (\psi + \upsilon t) + \log \kappa_{r,t} + \\ & \log \textstyle\sum_{j \in \{k \sim r\}} w_{r,j,g}(I_{r,g^*,c}(t) + (1-\omega)A_{r,g^*,c}(t))), \end{aligned} \tag{2.17}$$

and $w_{r,j,g}$ is the share of contacts among individuals of sex g in r that are with individuals of sex g^* in j. I use the relative infectiousness ratios listed in Table 3.1 in Johnson & Dorrington (2019) to fix the values of ξ_c and set the following priors on the sex ratio of transmission parameters:

$$\psi, \upsilon \sim \mathrm{N}(0, 5). \tag{2.18}$$

The spatial dynamics of transmission in Equation (2.17) are determined by weights, $w_{r,j,g}$, which are defined as

$$w_{r,j,g} = \sqrt{\frac{w(r,j)w(j,r)}{N_{g,r}N_{g^*,j}}}. \tag{2.19}$$

These weights balance contacts between people of sex g from region r with people of sex g^* from region j using the geometric mean of the two contact rates.

I define $w(r,j)$ such that the share, w_0, of contacts coming from r is fixed and the remaining share is divided among its neighbours:

$$w(r,j) = \begin{cases} w_0 & r = j \\ \dfrac{1-w_0}{\|\{j \sim r\} \setminus r\|} & r \neq j. \end{cases} \tag{2.20}$$

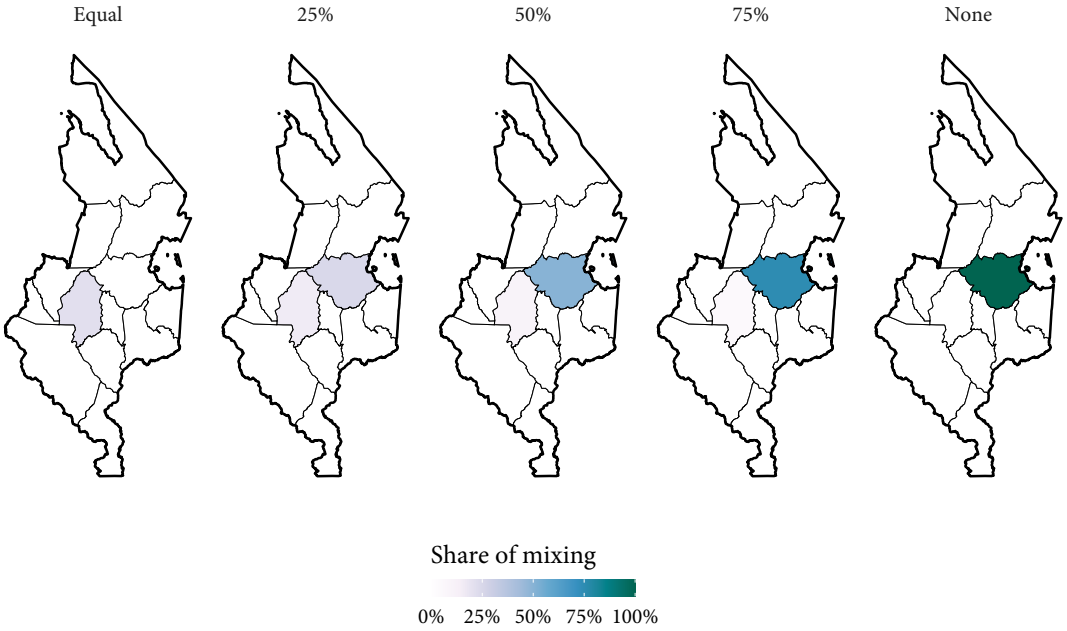

Figure 2.3: Share of mixing between Zomba district and all adjacent districts under five different assumptions about w_0.

where $\{j \sim r\} \setminus r$ is the set of regions adjacent to r exclusive of r. Figure 2.3 illustrates values of $\sqrt{w(r,j)w(j,r)}$ in the Zomba district of Malawi with varying assumptions about w_0. Although Held, Höhle, & Hofmann (2005) and Wakefield, Dong, & Minin (2017) infer the parameters for their spatial epidemic models, I do not attempt to do so here. The Epidemic/Endemic and TSIR models are typically fit to granular surveillance data from pathogens that spread rapidly over space and time. There are no such data on HIV cases in lower-income settings, so I assume that we cannot infer w_0 directly from data.

Mathematically, the incidence rate, $\lambda_{r,g}(t)$, is only constrained to be greater than zero, but numerical simulation of the system of ODEs in Equation (2.7) is not well constrained and negative predictions can disrupt the inference procedure. Therefore, I calculate the HIV infection probability during a single time step of duration h attributable to all disease stages combined by aggregating the stage-specific rates and finding the probability of infection:

$$\lambda_{r,g}(t) = 1 - \exp\left[-h \sum_{c=1}^{4} \lambda_{r,g,c}(t)\right]. \tag{2.21}$$

I use this transformation solely for the sake of avoiding numerical problems during the inference procedure; all reported region-level incidence here are per person-year.

Spatio-temporal HIV transmission rates The model allows the transmission rate of HIV to vary by time, region, sex, and transmitting CD4 category. The relative infectiousness by transmitting CD4 category is based on fixed assumptions, and the other dynamics are inferred. I model the log-transformed region-/time-specific HIV transmission rate, $\log \kappa_{r,t}$, with a hierarchical linear model:

$$\begin{aligned}
\log \kappa_{r,t} &= \kappa_0 + \kappa_r + (\gamma_0^\kappa + \gamma_r^\kappa) \cdot t + \textstyle\sum_{i=1}^{K_\kappa+1} \beta_{i,r}^\kappa \phi_i^\kappa(t) \\
\kappa_0 &\sim \mathrm{N}(0, 5) \\
\kappa_r &\sim \mathrm{N}(0, \sigma_\kappa) \\
\gamma_0^\kappa &\sim \mathrm{N}(0, 5) \\
\gamma_r^\kappa &\sim \mathrm{N}(0, \sigma_{\gamma^\kappa}) \\
\beta_{i,r}^\kappa &\sim \mathrm{ARIMA}_{\sigma_{\beta^\kappa}, \theta_{\beta^\kappa}}(1, d, 0) \\
\sigma_\kappa, \sigma_{\gamma^\kappa}, \sigma_{\beta^\kappa} &\sim \mathrm{N}^+(0, 1) \\
\operatorname{logit} \theta_{\beta_\kappa} &\sim \mathrm{N}(0, \sqrt{1/0.15}) \\
\beta_{1,r}^\kappa &= 0.
\end{aligned} \tag{2.22}$$

where κ_0 is a shared intercept, κ_r is a regional intercept and γ_0^κ and γ_r^κ are mean and region-specific slopes with respect to time. The remainder of the model for $\kappa_{r,t}$ defines a spline model with coefficients distributed according to an autoregressive integrated moving average (ARIMA) model (Hyndman & Athanasopoulos, 2018). In this model, K_κ is the number of knots, $\beta_{i,r}^\kappa$ is a region-specific coefficient for basis function i and ϕ_i^κ is the i'th basis function.

We say $\vec{y}$ follows an ARIMA(p, d, q) model, if the d-order sequential differences in y follow

$$\begin{aligned}
y'_t &= \textstyle\sum_{i=1}^{p} \phi_i y'_{t-i} + \sum_{j=1}^{q} \theta_j \epsilon_{t-j} + \epsilon_t \\
\epsilon_t &\sim \mathrm{N}(0, \sigma),
\end{aligned} \tag{2.23}$$

where y'_t is the t'th entry in the sequence of d-order differences in y, ϕ_i is an autoregressive coefficient, θ_j is a moving average coefficient, and ϵ_t is the "innovation" from $t-1$ to t (Hyndman & Athanasopoulos, 2018). We use y'_t as shorthand for the d-order differences. If $d = 0$, then $y'_t = y_t$, if $d = 1$, then $y'_t = y_t - y_{t-1}$, and so on. Note that if $d = 1$ and $p = q = 0$, then this model is a first-order random walk:

Param	Size	Description	Prior
ψ	1	Log-incidence rate sex ratio	$\mathrm{N}(0, 5)$
υ	1	Log-incidence rate sex ratio slope	$\mathrm{N}(0, 5)$
κ_0	1	Log-transmission rate mean intercept	$\mathrm{N}(0, 5)$
κ_r	R	Log-transmission rate region intercept	$\mathrm{N}(0, \sigma_\kappa)$
σ_κ	1	Log-transmission rate region intercept SD	$\mathrm{N}^+(0, 1)$
γ_0^κ	1	Log-transmission rate mean slope	$\mathrm{N}(0, 5)$
γ_r^κ	R	Log-transmission rate mean regional slope	$\mathrm{N}(0, \sigma_{\beta^\kappa})$
σ_{γ^κ}	1	Log-transmission rate region slope SD	$\mathrm{N}^+(0, 1)$
$\beta_{i,r}^\kappa$	$R \times K_\kappa + 1$	Log-transmission rate regional spline coefficient	$\mathrm{ARIMA}_{\sigma_{\beta^\kappa}, \theta_{\beta^\kappa}}(1, d, 0)$
σ_{β^κ}	1	Log-transmission rate region ARIMA SD	$\mathrm{N}^+(0, 1)$
$\mathrm{logit}\, \theta_{\beta^\kappa}$	1	Log-transmission rate region ARIMA autocorrelation SD	$\mathrm{N}^+(0, \sqrt{1/0.15})$

Table 2.2: Parameters used in the model of HIV transmission rates. Indexed parameters are estimated for all possible values of that index.

$$y_t - y_{t-1} = \epsilon_t, \tag{2.24}$$

where ϵ_t is Gaussian noise.

Returning to Equation (2.22), I specify that only one autoregressive term and no moving average terms may be included but do not specify the order of difference. All else equal, higher order differencing should result in a smoother curve. I assess the choice of d in the model comparison study presented in Chapter 3.

The model of incidence contributes the following parameters to θ_P: a transmission rate sex log-ratio, ψ, a transmission rate intercept, κ_0, a set of regional intercepts, κ_r, a mean transmission rate slope with respect to time, β_0^κ, a set of regional slopes, β_r^κ, and two standard deviations, σ_r^κ and σ_{β^κ}, which are outlined in Table 2.2.

Model of ART initiation

I use a similar log-linear regression approach to model region-/time-/sex-/substage-specific ART initiation rates (an outcome we similarly are not observing directly):

$$
\begin{aligned}
\log \alpha^{\star}_{r,t,g,c} &= \zeta_{c,g} + g \cdot \chi + \alpha^{\star}_0 + \alpha^{\star}_r + (\gamma^{\alpha^\star}_0 + \gamma^{\alpha^\star}_r) \cdot t + \textstyle\sum_{i=1}^{K_\alpha+1} \beta^{\alpha^\star}_{i,r} \phi^{\alpha^\star}_i(t) \\
\alpha^{\star}_0 &\sim \mathrm{N}(0, 5) \\
\alpha^{\star}_r &\sim \mathrm{N}(0, \sigma_{\alpha^\star}) \\
\gamma^{\alpha^\star}_0 &\sim \mathrm{N}(0, 5) \\
\gamma^{\alpha^\star}_r &\sim \mathrm{N}(0, \sigma_{\gamma^\alpha}) \\
\beta^{\alpha^\star}_{i,r} &\sim \mathrm{ARIMA}_{\sigma_{\beta^\alpha}, \theta_{\beta^\alpha}}(1, 2, 0) \\
\sigma_{\alpha^\star}, \sigma_{\gamma^\alpha}, \sigma_{\beta^\alpha} &\sim \mathrm{N}^{+}(0, 1) \\
\operatorname{logit} \theta_{\beta_\alpha} &\sim \mathrm{N}(0, \sqrt{1/0.15}) \\
\beta^{\alpha^\star}_{1,r} &= 0.
\end{aligned}
\tag{2.25}
$$

This model has region-specific log-linear models with respect to time with additional region-specific ARIMA error term. Here, $\zeta_{c,g}$ is a sex-/stage-specific rate of ART initiation, χ is an inferred intercept among women, $\alpha^{\star}_0$ is a mean intercept, $\alpha^{\star}_r$ is a regional intercept, K_α is a number of knots, $\beta^{\alpha^\star}_{i,r}$ is a regional spline coefficient, ϕ_i is a spline basis function, and $\sigma_{\alpha^\star}$, $\sigma_{\gamma^{\alpha^\star}}$, and $\sigma_{\beta^\alpha_r}$ are standard deviations. To prevent the autoregressive model from being rank deficient, I fix the first coefficient in the regional splines to be zero. In the results I present here, I set ϕ to be an order-two spline with annual knots, effectively linearly interpolating between inferred annual values.

Unlike in the model of HIV transmission rates, I have fixed the order of ARIMA differencing to be two in this model. This choice should produce relatively smooth estimates of ART initiation over time, with the capacity to make larger jumps if the data suggest they is necessary. Anecdotally, I have found that, because the ART patient count series are included, this model is more precisely determined by data than the transmission rate model.

For all t before ART was scaled up in any given region, I fix $\phi_i(t)$ to be zero. The baseline ART initiation rate $\zeta_{c,g}$ is defined as $\mu^I_{c,g}/\mu^I_{1,1}$, the ratio of mortality in CD4 stage c relative to women in stage 1. This encodes an assumption that PLHIV at stage c initiate treatment in proportion to the expected mortality in c. This model of ART initiation contributes the following parameters to θ_P: an intercept, $\alpha^{\star}_0$, a set of region random effects, $\alpha^{\star}_r$, set of spline coefficient means, mean and region-specific slopes, $\gamma^{\alpha^\star}_0$ and $\gamma^{\alpha^\star}_r$, a set of regional spline coefficients, $\beta^{\alpha^*}_{i,r}$, three standard deviations, $\sigma_{\alpha^\star}$, σ_{γ^α}, and σ_{β^α}, and an autoregressive

Param	Size	Description	Prior
χ	1	Log-ART initiation rate sex effect	$\mathrm{N}(0,5)$
$\alpha_0^\star$	1	Log-ART initiation rate mean intercept	$\mathrm{N}(0,5)$
$\alpha_r^\star$	R	Log-ART initiation rate regional intercept	$\mathrm{N}(0,\sigma_{\alpha^\star})$
$\sigma_{\alpha^\star}$	1	Log-ART initiation rate regional intercept SD	$\mathrm{N}^+(0,1)$
$\gamma_0^{\alpha^\star}$	1	Log-ART initation rate mean slope	$\mathrm{N}(0,5)$
$\gamma_r^{\alpha^\star}$	R	Log-ART initation rate regional slope	$\mathrm{N}(0,\sigma_{\gamma^\alpha})$
σ_{γ^α}	1	Log-transmission rate region slope SD	$\mathrm{N}^+(0,1)$
$\beta_{I,r}^{\alpha^\star}$	$R \times K_\alpha + 1$	Log-ART initiation rate regional spline coefficient	$\mathrm{ARIMA}_{\sigma_{\beta^\alpha},\theta_{\beta^\alpha}}(1,2,0)$
σ_{β^α}	1	Log-ART initiation rate region ARIMA SD	$\mathrm{N}^+(0,1)$
$\mathrm{logit}\,\theta_{\beta^\alpha}$	1	Log-ART initation rate region ARIMA autocorrelation SD	$\mathrm{N}^+(0,\sqrt{1/0.15})$
δ_0	1	Mean ANC bias	$\mathrm{N}(0,5)$

Table 2.3: Parameters used in the model of ART initation. Indexed parameters are estimated for all possible values of that index.

parameter θ_{β^α}, which are outlined in Table 2.3.

Model of initial state

Region-/sex-specific initial prevalence is modelled with a logit-linear model:

$$\begin{aligned} \mathrm{logit}\,\rho_{r,g}(0) &= \rho_r + g\cdot\epsilon \\ \rho_r &\sim \mathrm{N}(\rho_0,\sigma_\rho) \\ \rho_0 &\sim \mathrm{N}(0,5) \\ \sigma_\rho &\sim \mathrm{N}^+(0,1) \end{aligned} \tag{2.26}$$

where ρ_0 is cross-region logit-transformed mean prevalence at time 0, ρ_r is a regional deviation from ρ_0, ϵ is an intercept for prevalence among women (recalling that $g = 1$ among women), and σ_ρ is a standard deviation for the random effects. I calculate ϵ from Spectrum estimates as the log-ratio of female prevalence to male prevalence.

To maintain consistency with other, national-level estimates of prevalence, I put a prior on initial prevalence among men:

Param	Size	Description	Prior
ρ_0	1	Initial mean prevalence	$\mathrm{N}(0, 5)$
ρ_r	R	Initial regional prevalence	$\mathrm{N}(0, \sigma_\rho)$
σ_ρ	1	Initial prevalence SD	$\mathrm{N}^+(0, 1)$

Table 2.4: Parameters used in the model of the initial state. Indexed parameters are estimated for all possible values of that index.

$$\begin{aligned} \hat{\rho}_{\text{Nat}} &\sim \mathrm{N}(\rho_{\text{Nat}}, 0.005) \\ \hat{\rho}_{\text{Nat}} &= \frac{1}{P_{\text{Nat},0}(0)} \sum_{r=1}^{R} \frac{P_{r,0}(0)}{1+\exp(-\rho_r)}, \end{aligned} \tag{2.27}$$

where $P_{r,0}(0)/(1 + \exp(-\rho_r))$ is estimated male PLHIV in region r. This prior encourages the model to match external estimates of initial prevalence, without sacrificing subnational variation. The inverse logit-transformed mean of the random effects, $1/(1 + \exp(-\rho_0))$, cannot be compared directly to exogenous initial prevalence ρ_{Nat}, because ρ_{Nat} is implicitly population-weighted.

I assume that $t = 0$ is before ART scale-up, so $A_{r,g,c}(0) = 0$ in all cases. Making fixed assumptions about the distribution of PLHIV across disease substage without treatment, I calculate $I_{r,g,c}(0)$ and solve for $S_{r,g,c}(0)$:

$$\begin{aligned} I_{r,g,c}(0) &= b_{g,c} \cdot \rho_{r,g}(0) \cdot P_{r,g}(0) \\ S_{r,g}(0) &= P_{r,g}(0) - \textstyle\sum_{c=1}^{4}(I_{r,g,c}(0)) \\ A_{r,g,c}(0) &= 0. \end{aligned} \tag{2.28}$$

where $b_{g,c}$ is the Spectrum-derived share of PLHIV of sex g in CD4 stage c at time zero. $P_{r,g}(0)$ is exogenously defined population at time zero.

This model adds the following parameters to θ_P: a national mean, ρ_0, regional deviations from the means, ρ_r, and one standard deviation, σ_ρ, which are outlined in Table 2.4.

2.5 Observation model

With each of the components described above, simulating the model for a given set of process parameters yields predicted values for HIV incidence, prevalence, ART coverage,

and ART patient counts for both sexes, and all times and regions. In this section, I describe how those predictions are related to data via a set of observation models.

2.5.1 Household surveys

I assume that large household surveys are representative within each region, so if s is a household survey, $P^{s,\text{HIV}}_{r,t,g}/T^{s,\text{HIV}}_{r,t,g}$, provides an unbiased estimate of true prevalence in demographic segment $\{r, t, g\}$ where P and T are the effective count and effective sample size, respectively. P and T are calculated according to (1.2), so we can assume that $P^{s,\text{HIV}}_{r,t,g}$ is a sample from a binomial distribution with $T^{s,\text{HIV}}_{r,t,g}$ trials each with a probability of $\rho_{r,g}(t)$:

$$P^{s,\text{HIV}}_{r,t,g} \sim \text{Binom}(T^{s,\text{HIV}}_{r,t,g}, \rho_{r,g}(t)). \tag{2.29}$$

Defining a binomial distribution using the effective count and effective sample size is a computationally efficient way to approximate the effect of the complex multi-stage survey design (Chen, Wakefield, & Lumley, 2014) and is increasingly common in recent HIV mapping exercises (Dwyer-Lindgren et al., 2019; Eaton et al., 2021).

I make a similar assumption about survey-estimated ART coverage:

$$P^{s,\text{ART}}_{r,t,g} \sim \text{Binom}(T^{s,\text{ART}}_{r,t,g}, \alpha_{r,g}(t)). \tag{2.30}$$

Recency assays indicate whether an individual was infected in the recent past, so estimated incidence and prevalence must be combined to estimate the proportions that are recent. We can use the estimator from Kassanjee, McWalter, & Welte (2014) as modified by Eaton et al. (2021) to find this proportion:

$$\nu_{r,g}(t) = \frac{\lambda_{r,g}(t) \cdot (1 - \rho_{r,g}(t)) \cdot (\Omega_R - \gamma_R) + \gamma_R \rho_{r,g}(t)}{\rho_{r,g}(t)}, \tag{2.31}$$

where Ω_R is the mean duration of recent infection (fixed at 130/365), and γ_R is the proportion of positive recency assays that are false positives (fixed at 0). As before, I assume that each $P^{s,\text{Rec}}_{r,t,g}$ is a sample from a binomial distribution:

$$P^{s,\text{Rec}}_{r,t,g} \sim \text{Binom}(T^{s,\text{Rec}}_{r,t,g}, \nu_{r,g}(t)). \tag{2.32}$$

Because the mean duration of recent infection and recency assay false positive rate are fixed, the observation models for survey data do not contribute any parameters to the model.

2.5.2 ANC facility data

ANC attendees might not represent the general population well, so we cannot estimate $\rho_{r,1}(t)$ with $P^{s,\text{HIV}}_{r,t,1}/T^{s,\text{HIV}}_{r,t,1}$. Instead, following Bao (2012), I estimate site-specific ANC prevalence as a function of general population prevalence and facility effects

$$\begin{aligned}
\text{logit}\,\rho^s_{r,1}(t) &= \text{logit}\,\rho_{r,1}(t) + \delta_0 + \delta_s + (\epsilon_0 + \epsilon_s)\cdot t \\
\delta_0, \epsilon_0 &\sim \text{N}(0, 5) \\
\delta_s &\sim \text{N}(0, \sigma_\delta) \\
\epsilon_s &\sim \text{N}(0, \sigma_\epsilon) \\
\sigma_\delta, \sigma_\epsilon &\sim \text{N}(0, 1),
\end{aligned} \tag{2.33}$$

where δ_s is a facility-specific random effect, δ_0 is a mean ANC offset, ϵ_0 is a mean slope with respect to time, ϵ_s is a site-specific slope, and σ_δ and σ_ϵ are standard deviations for the site-specific parameters. Bao (2012) do not include slopes with respect time in their ANC observation. Eaton et al. (2014) indicated that we cannot assume that the representativeness of ANC facilities is not changing.

Eaton & Bao (2017) report that Gaussian approximations to standard binomial models do not offer adequate posterior predictive coverage when fit to HIV prevalence data from ANC facilities, so this work includes the option to use one of two possible likelihoods. The first is a standard binomial model

$$P^{s,\text{HIV}}_{r,t,1} \sim \text{Binom}(T^{s,\text{HIV}}_{r,t,1}, \rho^s_{r,1}(t)), \tag{2.34}$$

and the second is a beta-binomial model

Param	Size	Description	Prior
δ_s	S	Site-specific ANC bias	$\mathrm{N}(0, \sigma_\delta)$
σ_δ	1	ANC bias SD	$\mathrm{N}^+(0,1)$
ϵ_0	1	Mean ANC slope	$\mathrm{N}(0,5)$
ϵ_s	S	Site-specific ANC bias slope	$\mathrm{N}(0, \sigma_\epsilon)$
$\phi_{\text{type}[s]}$	2	Type-specific ANC beta-binomial overdispersion	$\mathrm{N}(-1,1)$
σ_ϵ	1	ANC bias slope SD	$\mathrm{N}^+(0,1)$

Table 2.5: Parameters used in the model of the initial state. Indexed parameters are estimated for all possible values of that index.

$$\begin{aligned} P^{s,\text{HIV}}_{r,t,1} &\sim \text{BetaBinom}(T^{s,\text{HIV}}_{r,t,1}, \rho^s_{r,1}(t), \phi_{\text{type}[s]}) \\ \text{logit}\, \phi_{\text{type}[s]} &\sim \mathrm{N}(-1,1) \end{aligned} \tag{2.35}$$

where $\phi_{\text{type}[s]} \in (0,1)$ is a parameter measuring the autocorrelation between each Bernoulli trial. In the beta-binomial case I estimate two separate values of ϕ: one when s is an individual facility and one when s is an aggregate over multiple facilities. I evaluate the effect of the choice of ANC observation in Chapter 3. This model contributes the following parameters to the model: coefficients for the observation model, δ_0, δ_s, ϵ_0, and ϵ_s, and hyperparameters, σ_δ, σ_ϵ, and $\phi_{\text{type}[s]}$, which are outlined in Table 2.5.

2.5.3 ART programme data

The final data source used by the model is programmatic ART patient count time series, which I denote $C_{r,t}$. The compartmental model produces estimates of the number of PLHIV living in r that are on treatment, $A_r(t)$, but these estimates are not directly comparable to the corresponding $C_{r,t}$. While large surveys measure individuals in their regions of residence, ART programme data measure individuals where they seek treatment. Because I fit directly to survey data and am using population estimates defined by residency, I am implicitly modelling individuals in their regions-of-residence, and therefore need to adjust $A_r(t)$ before it can be compared directly to $C_{r,t}$.

Following Eaton et al. (2021), I model the number of PLHIV seeking treatment in region r at time t as

$$A_r^*(t) = \sum_{\{j \sim r\}} \pi_{j \to r,t} A_j(t), \tag{2.36}$$

where $\{j \sim r\}$ is set of regions that are adjacent to r inclusive of r, $\pi_{j \to r,t}$ is the time-varying probability an individual residing in j will seek treatment in r, and $A_j(t)$ is the number of PLHIV on ART who live in j. Note that $A_j(t) = \sum_{g \in \{0,1\}} A_{j,g}(t)$.

I model the odds of moving from j to r relative to staying in r as

$$\begin{aligned}
\log \tfrac{\pi_{r \to j,t}}{\pi_{r \to r,t}} &= \log \mu_{r \to j,t} = m_j + m_0 + \beta_\pi t \\
m_j &\sim \mathrm{N}(0, \sigma_m^2) \\
m_0 &\sim \mathrm{N}(-3, 1) \\
\beta_\pi &\sim \mathrm{N}(0, 5) \\
\sigma_m &\sim \mathrm{N}^+(0, 2),
\end{aligned} \tag{2.37}$$

where m_j is a region-specific "mass" term, m_0 is a mean mass with a prior that ensures that most people will stay in their home regions, and β_π is a time-specific slope. Following the Naomi model, I place an informative prior on m_0 that assumes *a priori* that the majority of people seek treatment in their region of residence. If m_j and β_π are both fixed to be zero and region r has one neighbour, then $m_0 = -3.0$ implies that approximately 95% of individuals residing in r will seek treatment in r.

I allow each π to vary with respect to time to account for national-level changes in ART programmes; across-the-board improvements in treatment provision could result in fewer patients needing to travel to receive adequate care and therefore, a negative value of β_π. Naomi was designed to estimate recent trends, so it covers a much shorter time period and does not need to account for long-term changes in ART programmes.

I use the softmax function to solve for $\pi_{r \to j,t}$ with $\mu_{r \to j,t} = 1.0$:

$$\pi_{r \to j,t} = \frac{\mu_{r \to j,t}}{1 + \sum_{\{k \sim r\} \setminus r} \mu_{r \to k,t}}. \tag{2.38}$$

Then I find $\pi_{r \to r,t} = 1 - \sum_{\{k \sim r\} \setminus r} \pi_{r \to k,t}$.

These data do not have a natural denominator, so we cannot treat them as independent

binomial samples. Instead, I have implemented a negative binomial model with variance that can scale both linearly and quadratically with its mean (Lindén & Mäntyniemi, 2011). Let

$$\begin{aligned} \mu &= A_r^*(t) \\ \sigma^2 &= \mu + \theta_1\mu + \theta_2\mu^2, \end{aligned} \tag{2.39}$$

where $\theta_1, \theta_2 > 0$. We can use μ and σ^2 to find the typical negative binomial parameters: $r = \mu^2/(\sigma^2 - \mu)$ and $p = \mu/\sigma^2$. Then we have

$$C_{r,t} \sim \text{NegBinom}(r, p). \tag{2.40}$$

For a fixed value of θ_2, as θ_1 goes to zero, this distribution converges to a negative binomial with overdispersion θ_2. Conversely, for a fixed value of θ_1, as θ_2 goes to zero, it goes to a quasi-Poisson distribution. As both θ_1 and θ_2 go to zero, it returns to Poisson.

Allowing the variance of $C_{r,t}$ to scale both linearly and quadratically with μ allows this model to scale appropriately across regions of varying sizes. A one-unit change in θ_2 will result in a much larger change in variance in a high-population region than in a low-population region, even though we do not necessarily expect the measurement of ART patients to be quadratically higher variance in the high-population region.

I set the following priors on θ_1 and θ_2:

$$\log \theta_1, \log \theta_2 \sim \text{N}(0, 2) \tag{2.41}$$

The observation model for ART patient counts contributes the following parameters to the full set of parameters: θ_1 and θ_2 from the quasi-negative binomial distribution, one mass, m_r, per region, a mean mass, m_0, a time coefficient, β_t, and one variance σ_m^2, which are outlined in Table 2.6.

Param	Size	Description	Prior
m_0	1	Mean ART attraction mass	$\mathrm{N}(0, 5)$
m_r	R	Regional ART attraction mass	$\mathrm{N}(0, \sigma_m)$
σ_m	1	Regional ART attraction mass SD	$\mathrm{N}^+(0, 1)$
β_π	1	ART attraction slope	$\mathrm{N}(0, 5)$
ω	1	Linear overdispersion term	$\mathrm{N}^+(0, 2)$
θ	1	Quadratic overdispersion term	$\mathrm{N}^+(0, 2)$

Table 2.6: Parameters used in the ART patint count observation model Indexed parameters are estimated for all possible values of that index.

2.5.4 Posterior density

I assume each data point is an independent sample from the distributions listed above. Let $\mathcal{D}^{\text{HIV}}$, $\mathcal{D}^{\text{ART}}$, and $\mathcal{D}^{\text{Rec}}$ be the subsets of $\mathcal{D}$ corresponding prevalence, treatment, and recency data, respectively. Then we have

$$\mathrm{P}(\mathcal{D} \mid \theta_O, \theta_P) = \mathrm{P}(\mathcal{D}^{\text{HIV}} \mid \theta_O, \theta_P) \times \mathrm{P}(\mathcal{D}^{\text{ART}} \mid \theta_O, \theta_P) \times \mathrm{P}(\mathcal{D}^{\text{Rec}} \mid \theta_O, \theta_P). \tag{2.42}$$

We can decompose further by data source to illustrate the varying degrees of dependence on parameters:

$$\begin{aligned} \mathrm{P}(\mathcal{D} \mid \theta_O, \theta_P) \quad = \quad & \mathrm{P}(\mathcal{D}^{\text{HIV}}_{\text{Surv}} \mid \theta_P) \times \\ & \mathrm{P}(\mathcal{D}^{\text{ART}}_{\text{Surv}} \mid \theta_P) \times \\ & \mathrm{P}(\mathcal{D}^{\text{Rec}}_{\text{Surv}} \mid \theta_P) \times \\ & \mathrm{P}(\mathcal{D}^{\text{HIV}}_{\text{ANC}} \mid \theta_O, \theta_P) \times \\ & \mathrm{P}(\mathcal{D}^{\text{ART}}_{\text{Prog}} \mid \theta_O, \theta_P). \end{aligned} \tag{2.43}$$

None of the likelihoods of the survey data depend on inferred parameters, so $D^{\text{HIV}}_{\text{Surv}}$, $D^{\text{ART}}_{\text{Surv}}$, and $D^{\text{Rec}}_{\text{Surv}}$ are independent of θ_{O}.

We can refine this further by identifying the compartmental model outputs that each indicator depends on:

$$\begin{aligned} \mathrm{P}(\mathcal{D} \mid \theta_O, \theta_P) \quad = \quad & \mathrm{P}(\mathcal{D}^{\mathrm{HIV}}_{\mathrm{Surv}} \mid \rho) \times \\ & \mathrm{P}(\mathcal{D}^{\mathrm{ART}}_{\mathrm{Surv}} \mid \alpha) \times \\ & \mathrm{P}(\mathcal{D}^{\mathrm{Rec}}_{\mathrm{Surv}} \mid \lambda, \rho) \times \\ & \mathrm{P}(\mathcal{D}^{\mathrm{HIV}}_{\mathrm{ANC}} \mid \theta_O, \rho) \times \\ & \mathrm{P}(\mathcal{D}^{\mathrm{ART}}_{\mathrm{Prog}} \mid \theta_O, A). \end{aligned} \tag{2.44}$$

Finally, we have the full posterior density (omitting the normalizing constant):

$$\begin{aligned} \mathrm{P}(\theta_O, \theta_P \mid \mathcal{D}) \propto \quad & \mathrm{P}(\mathcal{D}^{\mathrm{HIV}}_{\mathrm{Surv}} \mid \rho) \times \\ & \mathrm{P}(\mathcal{D}^{\mathrm{ART}}_{\mathrm{Surv}} \mid \alpha) \times \\ & \mathrm{P}(\mathcal{D}^{\mathrm{Rec}}_{\mathrm{Surv}} \mid \lambda, \rho) \times \\ & \mathrm{P}(\mathcal{D}^{\mathrm{HIV}}_{\mathrm{ANC}} \mid \theta_O, \rho) \times \\ & \mathrm{P}(\mathcal{D}^{\mathrm{ART}}_{\mathrm{Prog}} \mid \theta_O, A) \times \\ & \mathrm{P}(\theta_O) \times \mathrm{P}(\theta_P). \end{aligned} \tag{2.45}$$

2.6 Inference strategy

The model described above is a complex one that involves substantial computation for a single posterior density evaluation. The posterior log-density is analytically intractable, so I must use an approximate method or a sampling algorithm like MCMC to sample from $\mathrm{P}(\theta_O, \theta_P \mid \mathcal{D})$ and produce posterior estimates of λ.

The No U-Turn Sampler (NUTS) (Hoffman & Gelman, 2011) has proved too intensive for testing, so I have focused on an approximate inference strategy based on the Laplace method as implemented in the **TMB** R package Kristensen et al. (2016). The inference strategy proposed by Skaug & Fournier (2006) uses automatic differentiation to find the joint posterior mode of the model, places a multivariate normal around that mode, and produces samples by taking samples from the multivariate normal.

A key feature of this strategy is the ability to quickly approximate the marginal posterior density with respect to a set of "random" parameters, $\theta_R \subset \theta_O \cup \theta_P$, via the Laplace method. In this case, θ_R is every parameter listed above that is not a hyperparameter. If $f(\theta_F, \theta_R)$ is the negative posterior log-density of our model with "fixed" parameters $\theta_F \in \mathbb{R}^m$ and

random parameters $\theta_R \in \mathbb{R}^n$, then the Laplace method approximates the marginal posterior density of the model with respect to θ_R; that is, it approximates the following integral:

$$L(\theta_F) = \int_{\mathbb{R}^n} \exp(f(\theta_F, \theta_R))d\theta_R. \tag{2.46}$$

If we can find $\hat{\theta}_R$ that minimizes $f(\theta_F, \theta_R)$ for a given set of fixed parameters θ_F via an inner optimisation step, then we can use the Laplace approximation to approximate the marginal density of θ_F with a multivariate normal distribution:

$$L^\star(\theta_F) = 2\pi^{n/2}\det(H(\theta_F))^{-1/2}\exp(-f(\theta_F, \hat{\theta}_R)), \tag{2.47}$$

where $H(\theta_F)$ is the Hessian of $f(\theta_F, \theta_R)$ evaluated at $\hat{\theta}_R$. **TMB** uses automatic differentiation to find $H(\theta_F)$. Because we are minimising the negative log-likelihood, our outer optimisation step is minimising

$$-\log L^\star(\theta_F) = -m\sqrt{2\pi} + \frac{1}{2}\log\det(H(\theta_F)) + f(\theta_F, \hat{\theta}_R). \tag{2.48}$$

To produce posterior samples from $\mathrm{P}(\theta_O, \theta_P \mid \mathcal{D})$, I use TMB to find maximum *a posteriori* estimates of both the random and non-random parameters, θ_O and θ_P, which we will call θ^{MAP}. I find the precision matrix of the model with respect to all non-hyperparameters at θ^{MAP} and invert it to find Σ^{MAP}, the covariance matrix of the model with respect to the non-hyperparameters. I then take samples from the multivariate normal distribution centred at θ^{MAP} with covariance Σ^{MAP} to approximate sampling from $\mathrm{P}(\theta_O, \theta_P \mid \mathcal{D})$, where Σ^{MAP} includes zero variance and covariance for all hyperparameters. Assuming that the hyperparameters have zero variance is not accurate, but I have found that applying the Laplace approximation to hyperparameters often results in implausible predictions.

2.7 Remaining chapters

In subsequent chapters, I describe how I have validated and applied the model proposed here. In Chapter 3, I describe a cross-validation strategy and experiment designed to

help us identify which (if any) of the possible specific configurations of this model is best. In Chapter 4, I test the effects of varying assumptions about spatial transmission on computation time and inferred incidence. In Chapter 5, I select one, suitably good model specification and use it generate descriptive results for the 28 districts of Malawi. Finally, in Chapter 6, I examine how well the model predicts incidence at increasingly long forecasting horizons to assess the benefits of sequentially including administrative data as it becomes available.

Chapter 3

Model Specification Testing

Specifying the model presented in Chapter 2 requires the user to make a number of design decisions that could meaningfully affect inferred HIV incidence. To help make these decisions in an empirically justified way, I have designed a cross-validation scheme that allows the user to compare how well different model specifications predict out-of-sample data. In this chapter, I present a series of model specification experiments applying the cross-validation scheme to data from Malawi and discuss the effects various specifications decisions could have on inferred incidence.

3.1 Introduction

Throughout Chapter 2, I neglected to specify a number of key design decisions that could impact inferred incidence. For example, I did not fix the number of knots in the transmission rate model. This is complex model, so between all combinations of all possible choices there are hundreds or thousands of models that we might consider fitting.

Because we do not fit directly to incidence data, we must carefully assess the extent to which our inferred incidence series are determined by design decisions. Previously proposed models of HIV incidence have typically offered narrow sets of sensitivity analyses, but only a small number of authors have reported comprehensive, cross-validated model specification tests (Dwyer-Lindgren et al., 2019).

To those ends, I designed a model specification experiment to identify which model or set

of models fit best to prevalence and treatment data and to assess how inferred incidence was impacted by model specification choices. I fit every combination of a set of possible choices (146 models in total) to real data from Malawi and compared the posterior predictions using cross-validation over six forecasting horizons. I also identified determinants of convergence failure in the approximate inference strategy from Section 2.6 to try to inform future model fitting and compared posterior distributions from that strategy to those of the No U-Turn Sampler (NUTS).

3.2 Methods

I fit a grid of 146 model configurations to data from Malawi from 1995 through 2020 with six hold-out horizons (1 January 2015, 2016, 2017, 2018, 2019, and 2020). The model configurations varied in terms of the ANC observation model and the transmission rate specification, outlined in Table 3.2. I cross-validated the results by comparing out-of-sample predictions across each of the six horizons and identified the model configurations that most often resulted in the best out-of-sample fit.

3.2.1 Data

I fit every model configuration to data from Malawi collected between 1995 and 2020, described in Section 1.4 and outlined in Table 3.1. Specifically, I fit to seroprevalence data from four household surveys: three DHS and one PHIA survey. From the 2004, 2010, and 2015-16 Demographic and Health Surveys, I used HIV test results, and from MPHIA 2015-2016, I used HIV, ART, and recency test results (Ministry of Health, Malawi, 2018; National Statistical Office/Malawi & ICF, 2017; National Statistical Office - NSO/Malawi & ICF Macro, 2011; National Statistical Office - NSO/Malawi & ORC Macro, 2005). I used ART patient count data provided by the the Malawi DHA and supplemented with data from a published report (Médecins Sans Frontières, Malawi & Ministry of Health and Population, Chiradzulu District, Malawi, 2004). Finally, I included the combined dataset of ANC facility testing data from EPP and from the DHA. Each of the data sources described in Section 2.2 is represented and supplemented with a set of fixed, exogenous assumptions. I used district-level population estimates from the Malawi Census Bureau, disaggregating

Name	Type	Years	Prevalence	Treatment	Recency
2004 DHS	HH Survey	2004	✓		
2010 DHS	HH Survey	2010	✓		
2015-2016 DHS	HH Survey	2015-2016	✓		
MPHIA 2015-2016	HH Survey	2015-2016	✓	✓	✓
UNAIDS ANC Data	Sentinel surveillance	1995-2010	✓		
DHAMIS	Facility reports	2011-present	✓		
DHAMIS	Facility reports	2005-present		✓	

Table 3.1: Population-level data sources from Malawi used in this analysis.

five-year age groups as necessary using Beers interpolation and linearly interpolating between annual estimates to get quarterly estimates (Shryock, Siegel, & Stockwell, 1976). The other exogenous data (mortality rates, progression parameters, etc.) were calculated as described in Chapter 2.

3.2.2 Cross-validation

Because incidence is not measured directly, the correct cross-validation strategy for these data is not obvious. Cross-validation simulates how well a model generalises to new data, so I have designed a strategy that is focused on forecasting the data sources we expect to continue to acquire (Vollmer et al., 2021). To that end, the cross-validation strategy predicts on routinely reported indicators, represented in my model by ART programme data and facility-based ANC prevalence. I constructed cross-validation datasets by holding out all data after one of six forecasting horizons: the first of January in 2015, 2016, 2017, 2018, 2019, and 2020. I compared each model using out-of-sample root mean squared error (RMSE) with respect to observed point estimates and 50%, 80%, and 95% posterior predictive coverage separately for the two datasets (ANC facility data and ART programme data).

This work focuses on estimating HIV incidence over space, so it is worth distinguishing the current cross-validation problem from those often found in spatial statistics. Whereas interpolation over unobserved spatial units is a central goal of a model like the one described

by Dwyer-Lindgren et al. (2019), this model operates on fixed regions. If we were to re-run the various data collection mechanisms, we would obtain new observations of the same geographic units given unobserved progression of the epidemic over time. We would not discover a new district of Malawi. Therefore, I have not used a spatial cross-validation scheme.

3.2.3 Model configurations

I tested every combination of choices for seven different design decisions, outlined in Table 3.2. First, I allowed the likelihood for the ANC facility data to be either binomial or beta-binomial. As described in Section 2.5.2, individual facility series shared one autocorrelation parameter and aggregate series shared another under the beta-binomial model. Second, I tested the value of including a non-linear district-level temporal component in the transmission rate model by fitting Equation (2.22) with intercepts only, intercepts and linear slopes with respect to time, and intercepts, slopes, and latent components. Among the models with latent components for the HIV transmission rate, I tested the effects of excluding the linear slope with respect to time, the order of the spline basis functions (one, two, or three), the distance between knots in the spline design matrices (one year or five years), the order of ARIMA differencing, and, finally, whether to include an autoregressive term. All valid combinations of these choices resulted in 146 different models, which led to 876 models to fit when combined with the six forecasting horizons.

3.2.4 Experiments

I used the 146 models to examine four questions relevant to future applications of this model and other models of this type:

1. What determines whether or not the optimiser will converge for a particular model configuration?
2. Which of the two ANC observation models described in Section 2.5.2 fits better?
3. Is there an optimal transmission rate model configuration or a set of choices that consistently lead to better fit?
4. Is fitting with the No U-Turn Sampler feasible or necessary?

Variable	Value	Description
ANC observation model	Binomial	Binomial ANC observation model
	Beta-binomial	Beta-binomial ANC observation model
ARIMA order	1	One degree of ARIMA differencing
	2	Two degrees of ARIMA differencing
	3	Three degrees of ARIMA differencing
Include slope	Yes	Exclude slope w.r.t. time
	No	Include slope w.r.t. time
Spline interval	1	Knots at one-year intervals
	5	Knots at five-year intervals
Spline order	1	Piecewise constant design matrix
	2	Piecewise linear design matrix
	3	Order-three design matrix
Transm. rate model	Constant	Constant w.r.t. time
	Linear	Linear w.r.t. time
	Latent	Include ARIMA component
Use AR	Yes	Include autoregressive term
	No	Exclude autoregressive term

Table 3.2: Model configuration variables tested in this chapter with descriptions of each value. Unless otherwise specified, every component refers to the transmission rate model.

This exercise required fitting and comparing several hundred models. To facilitate interpretation of the results, I conducted the four analyses sequentially: I addressed each question using the successful or best models from the previous question. For example, I only used the best ANC observation model to compare transmission rate parametrisations.

Identifying computational problems

The approximate inference strategy described in Section 2.6 utilises two optimisation steps: an inner step that finds the optimal random effect values $\hat{\theta}_R$ given the current fixed effect values, θ_F, and an outer step that optimises θ_F given $\hat{\theta}_R$. Anecdotally, the Newton optimiser used for inner optimisation in **TMB** can struggle with extremely complex or weakly identified models because the gradients with respect to certain parameters in θ_R diverge to positive infinity, causing the optimiser to fail. Having observed this behaviour in some model configurations, I first examined whether there were consistent determinants of convergence failure. I plotted failure rates by each value of each configuration variable and looked for outliers.

ANC observation model

Taking only the models that converged from the previous analysis, I compared out-of-sample RMSE and 50%, 80%, and 95% posterior predictive coverages between the binomial and beta-binomial ANC observation models. To identify the best observation model, I produced box plots of out-of-sample RMSE and posterior predictive coverage by observation model and forecasting horizon. Because I fit every valid configuration described in Section 3.2.3, every binomial model had a beta-binomial equivalent to which I could compare it directly. I performed this comparison by making scatter plots of out-of-sample RMSE and posterior predictive coverage of all pairs for which both models converged.

Transmission rate parametrisation

Observing that one ANC observation model was superior to the other, I restricted the models once again to just those with that observation model. Within this subset of 73 models, I compared out-of-sample RMSE and posterior predictive coverage across each of

the transmission rate parametrisation variables. I also found the best model for each time horizon-dataset combination, and found how often each configuration variable appeared in each best model.

Considering certain intuitive theoretical requirements (for example, that HIV transmission will not decrease monotonically to zero), I visually inspected results generated using configuration variables that appeared frequently in best models to characterise the effects of different design decisions. After identifying these characteristics, I verified that the observation model from Section 3.2.4 was still best in the set of selected models. By comparing the models identified as good candidates to previous estimates of national level incidence in Malawi, I identified a single model to take as the fixed model for the rest of the document.

Inference strategy

Finally, I used the **tmbstan** R library (Monnahan & Kristensen, 2018) to fit the final model from Section 3.2.4 using both the approximate strategy described in Section 2.6 and NUTS (Hoffman & Gelman, 2011). I compared the two sets of fits by plotting the posterior means and standard deviations of each parameter. Given the focus on inferring incidence, I also compared point estimates and 95% credible intervals for incidence by sex, region, and time.

3.3 Results

Of the 876 model configuration-horizon combinations, 736 (84%) successfully converged. The largest single determinant of failure was using only a log-linear model for the HIV transmission rate, while the largest determinant of success was using a beta-binomial distribution in the ANC observation model. Figure 3.1 presents failure rates among the 438 model configuration-horizon combinations that used a beta-binomial likelihood. Of the 34 failures in that subset, the largest determinant of failure was forcing the transmission rate model to be log-linear with respect to time.

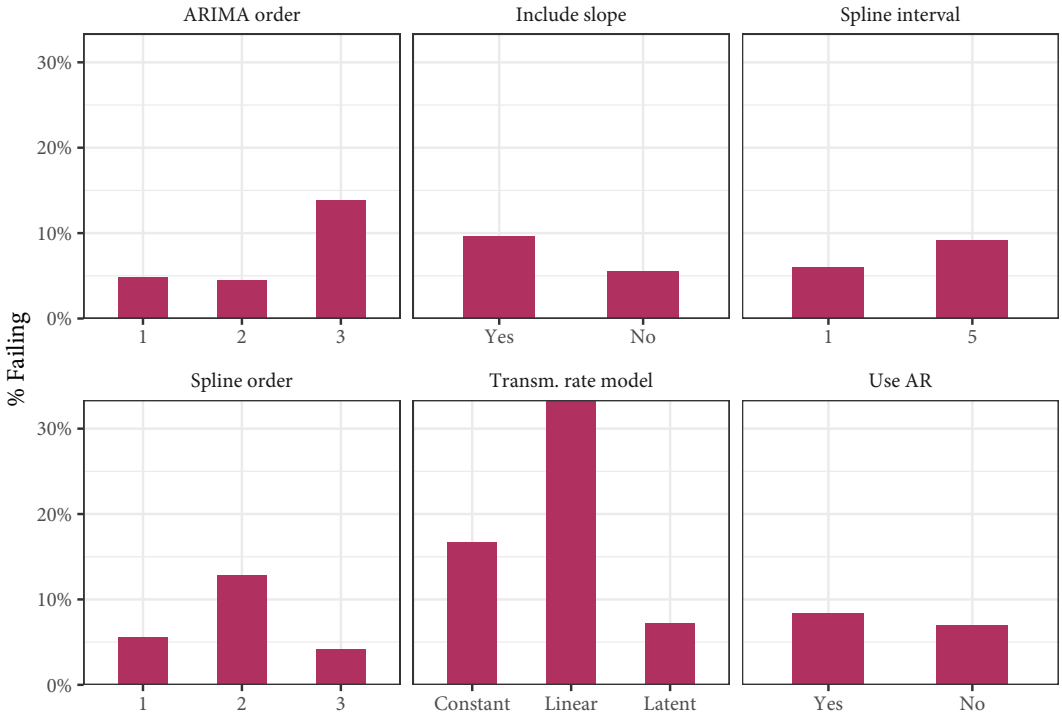

Figure 3.1: Share of configuration-horizon combinations that failed to converge by configuration variable.

3.3.1 ANC observation model comparison

Restricting to the 736 successful models, I examined the impact of the ANC observation model on out-of-sample prediction of ANC prevalence and ART patient counts. Figure 3.2 presents box plots of log-transformed out-of-sample RMSE by ANC observation model and metric, aggregating over model configuration and forecasting horizon. These box plots provide an indication of how these two observation models compare to each other regardless of exact model specification. The beta-binomial observation model fit better to both ANC prevalence and ART patient count data than the binomial model.

Figure 3.3 stratifies each box plot by forecasting horizon. Note that the x-axis of this figure is the date forecasting began, so "01/01/2015" holds out the most data and "01/01/2020" holds out the least. Across both datasets, the beta-binomial observation model resulted in better out-of-sample RMSE in almost every case.

Finally, Figure 3.4 compares out-of-sample RMSEs for every pair models with otherwise identical specifications. A point being below the line of equality indicates that the binomial observation model had higher RMSE compared to the equivalent beta-binomial model.

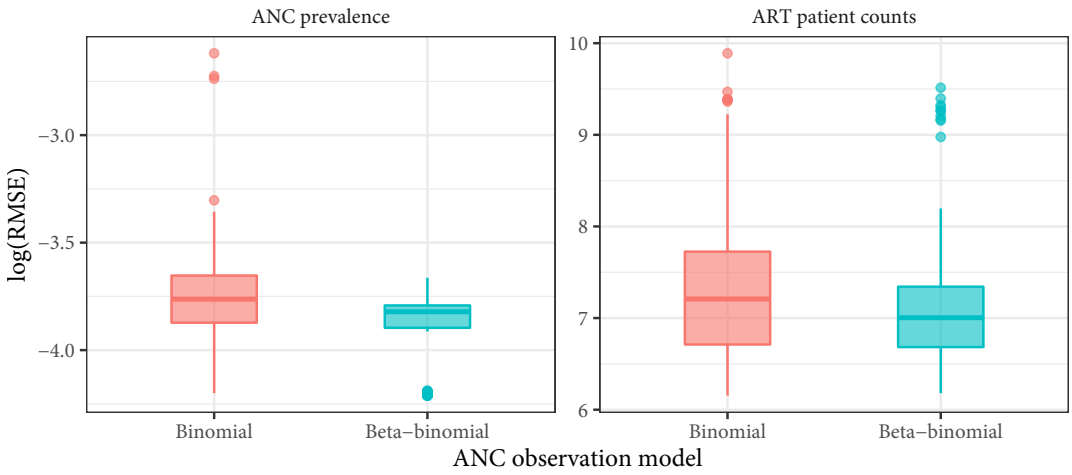

Figure 3.2: Box plots of log-transformed out-of-sample RMSE by ANC observation model and data set.

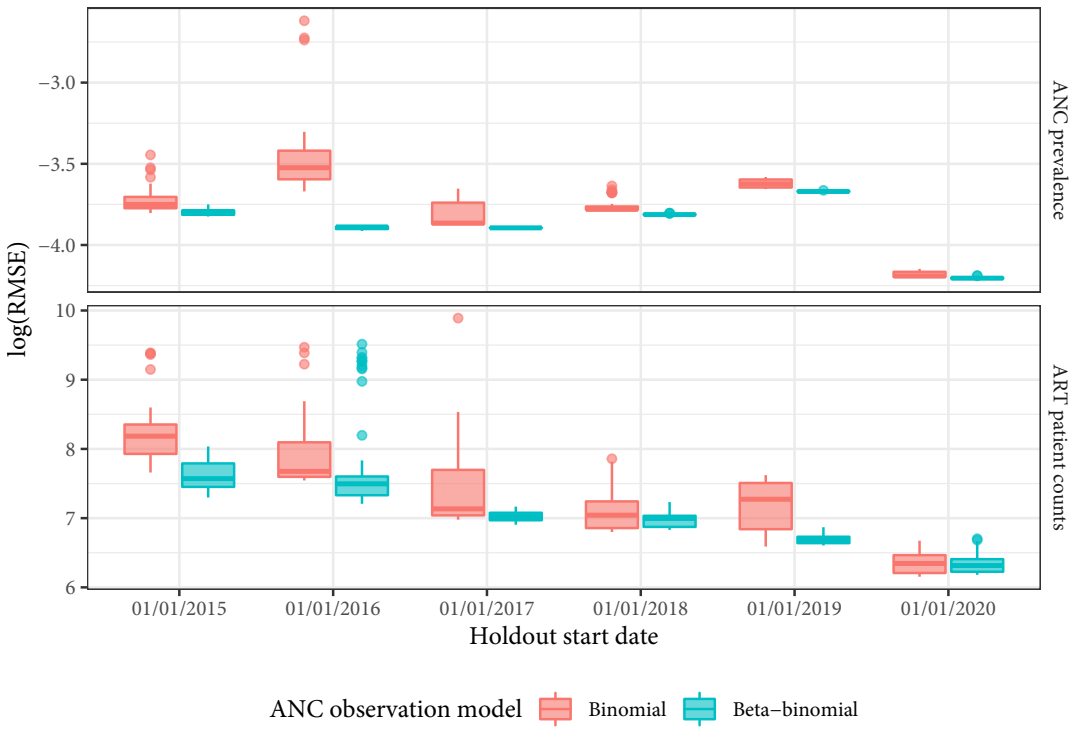

Figure 3.3: Box plots of log-transformed out-of-sample RMSE by holdout start date, ANC observation model, and dataset.

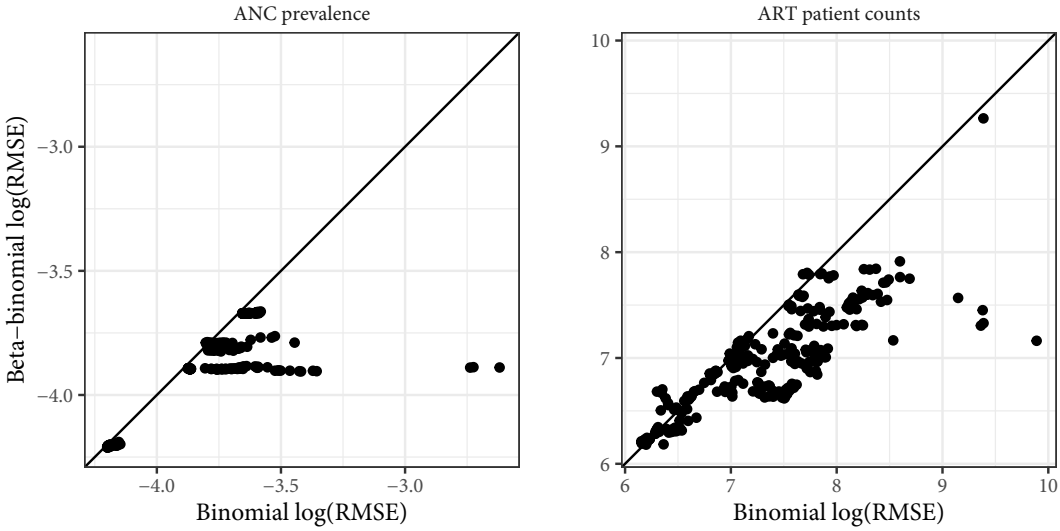

Figure 3.4: Scatter plots of log-transformed out-of-sample RMSE for model configuration pairs that differ only in ANC observation model by dataset. The black line is equality. Points below the line of equality indicate that the beta-binomial observation model was lower and vice versa.

This comparison shows that the beta-binomial observation model generally offered superior out-of-sample posterior predictive fit.

The results were less decisive when using out-of-sample posterior predictive coverage. Figure 3.5 presents box plots of 50%, 80%, and 95% posterior predictive coverage by ANC observation model, metric, and coverage level. Median 50%, 80%, and 95% posterior predictive coverages of ANC data were 65.8%, 89.1%, and 97.1%, respectively, in models using a beta-binomial observation model, suggesting that it overestimated uncertainty. Conversely, the same coverage values were 41.4%, 65.6%, and 81.8% when using a binomial observation model, suggesting that it underestimated uncertainty.

Defining posterior predictive coverage error as the absolute difference between the target coverage level and observed posterior predictive coverage provides a means to compare high and low coverage values to each other. Table 3.3 presents the share of equivalent pairs in which the beta-binomial model had a better value than the binomial model for four error metrics. In general, the better of each pair of models is the one that uses the beta-binomial observation model.

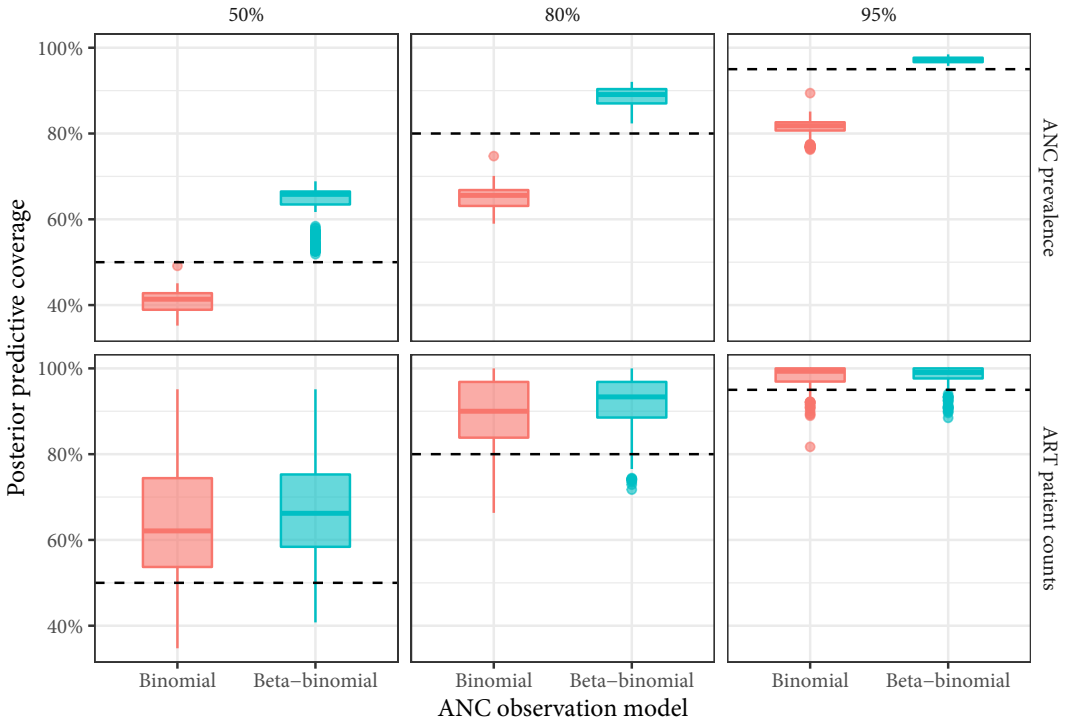

Figure 3.5: Out-of-sample 50%, 80%, and 95% posterior predictive coverages by ANC observation model and dataset. Dashed lines are target coverages.

	50% post. coverage error	80% post. coverage error	95% post. coverage error	RMSE
ANC prevalence	16.9%	99.4%	100.0%	98.7%
ART patient counts	43.2%	51.9%	71.4%	74.4%

Table 3.3: The proportion of configuration pairs in which the beta-binomial ANC observation model was superior to the binomial model by dataset and error metric.

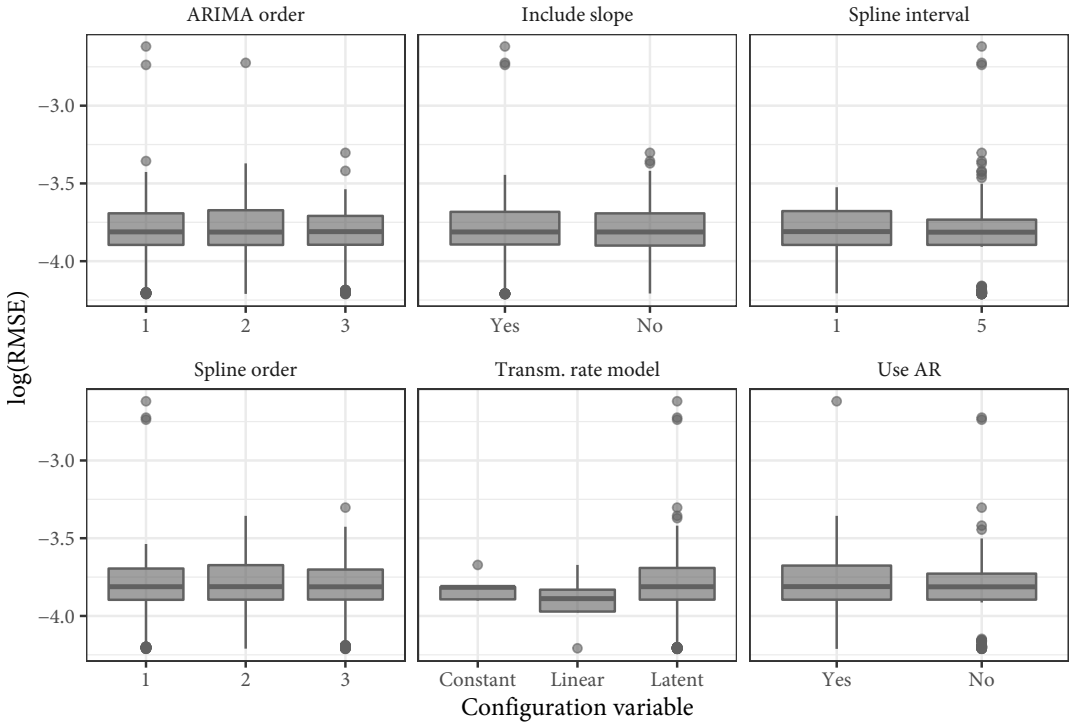

Figure 3.6: Box plots of log-transformed out-of-sample RMSE by configuration variable for ANC prevalence data.

3.3.2 Transmission rate specification

Given the results of the ANC observation model comparison, I restricted all other comparisons to configurations with a beta-binomial ANC observation model, reducing the number of unique configuration-horizon pairs from 876 to 438. Figures 3.6 and 3.7 present log-transformed out-of-sample RMSE values by dataset and configuration variable. Unlike in the ANC observation model comparison, none of these options provided a decisive advantage over the alternatives. Placing knots once every five years resulted in better log-RMSE than placing knots once per year, but these plots offer little insight otherwise.

To identify whether any single configuration was superior to the others, separately by dataset, I found the number of forecasting horizons for which each configurations was in among k best configurations. For example, this analysis would identify any configuration that was among the five best for all six forecasting horizons. No configuration had the best RMSE for more than one horizon-dataset combination, so I found the number of times each model configuration appeared in the best five and ten configurations across all horizons by dataset. Table 3.4 provides the number of configurations that appeared zero

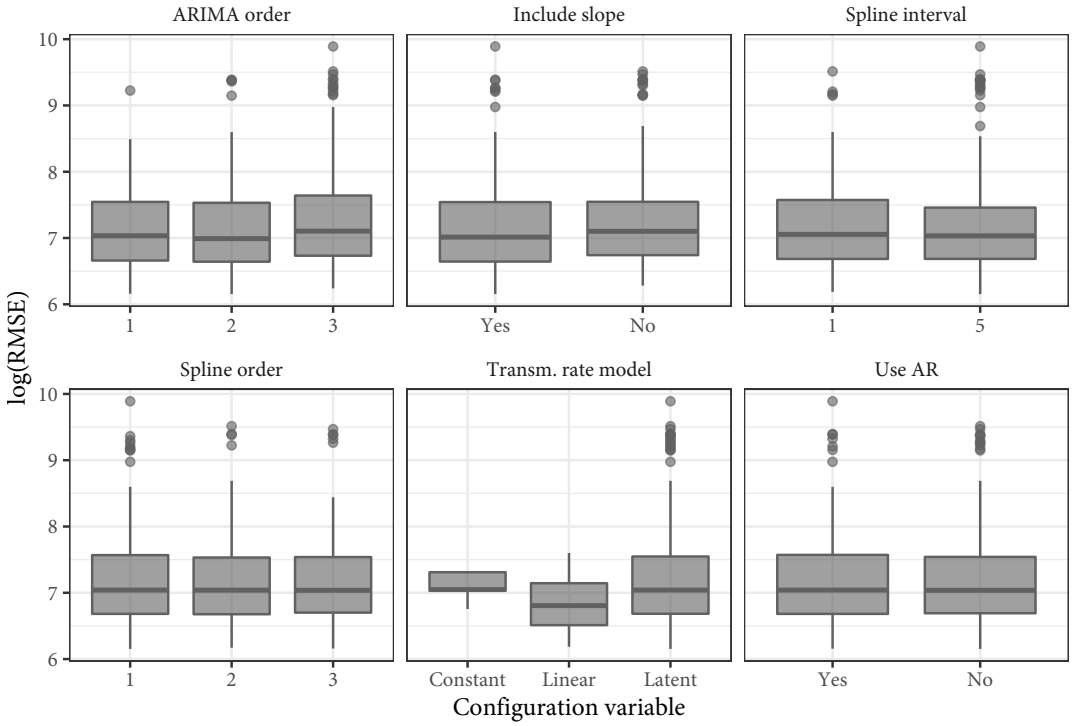

Figure 3.7: Box plots of log-transformed out-of-sample RMSE by configuration variable for ART patient counts.

through six times in the five-best and ten-best by dataset. Column "6" being zero across all rows shows that no single configuration appeared in the top ten for all six horizons. Three configurations were in both the top five and top ten configurations for three of six horizons for ANC prevalence. The top configurations for ART patients were more variable; few model configurations appear in the top ten more than once.

To understand which individual features of each configuration led to better fit, I found the number of times each configuration variable appeared in the single best fit for all 12 horizon-dataset pairs, plotted in Figure 3.8. Five top configurations included a slope in the transmission rate model while seven did not. Every top-ranked model included a non-linear temporal component and most used five-year intervals for spline knots, included autoregressive terms, and used order-three splines.

Informed by the results from Figure 3.8, I subjectively compared the predictions from a handful of models to identify the effects of each decision on predictions. Specifically, I restricted to configurations that included non-linear temporal components and fixed the order of differencing to one, the spline interval to five, and the spline order to two. This left four configurations: with and without a linear temporal term and with and without an

	0	1	2	3	4	5	6
ANC Prevalence							
1	68	6	0	0	0	0	0
5	52	17	2	3	0	0	0
10	37	21	11	3	2	0	0
ART Count							
1	71	3	0	0	0	0	0
5	60	12	2	0	0	0	0
10	44	22	8	0	0	0	0

Table 3.4: The number of configurations that appear n times in the top k models for each forecasting horizon. The 0 column indicates that most models never appeared in the top 1, 5, or 10 models, while the 6 column indicates that no model was in the top 1, 5, or 10 models for all six horizons.

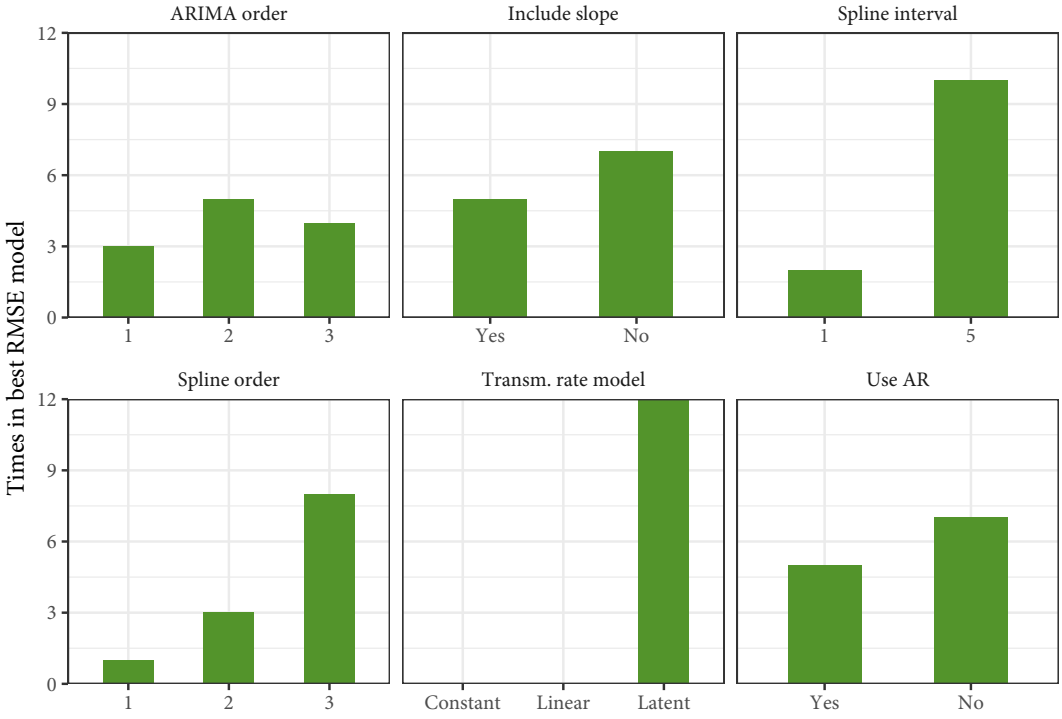

Figure 3.8: Bar plots of the number of times each configuration variable value appears in the best model for each horizon-dataset pair (12 maximum).

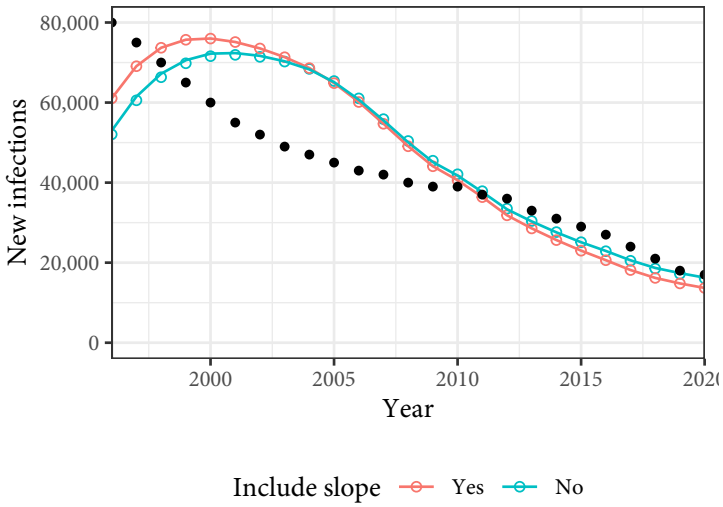

Figure 3.9: Comparison of UNAIDS estimated annual new infections to four models from this analysis. Lines correspond models without autoregressive terms, and open circles correspond to models with autoregressive terms. Black points are UNAIDS estimates.

autoregressive term.

Figure 3.9 compares inferred national new infections incidence in those four models with the shortest forecasting horizon in the analysis (1 January 2020). Including an autoregressive term made only a small difference, but including a linear slope with respect to time led to higher initial incidence and more rapidly declining incidence trend.

At the district level, the inclusion of a linear term in the transmission rate model had a larger effect on incidence than the inclusion of an autoregressive term. I measured the effects of these decisions on spatial variability by finding the coefficient of variation in incidence across regions by time and sex. Figure 3.10 presents estimated coefficients of variation for these four models. The two models with linear terms inferred more spatial variation in earlier years and less spatial variation later, while models with and without autoregressive terms were similar. Among men in 2020, the mean ratio of incidence in models than included linear terms to those that did not was 0.94, indicating that models with linear terms estimated less spatial variability among men in 2020. The same ratio comparing models that did and did not include autoregressive terms was 1.01, indicating that including an autoregressive had only a small effect on spatial variability.

Finally, I compared the four selected models to existing national-level estimates. Figure 3.9 compares national-level new infections as predicted by the four models to corresponding

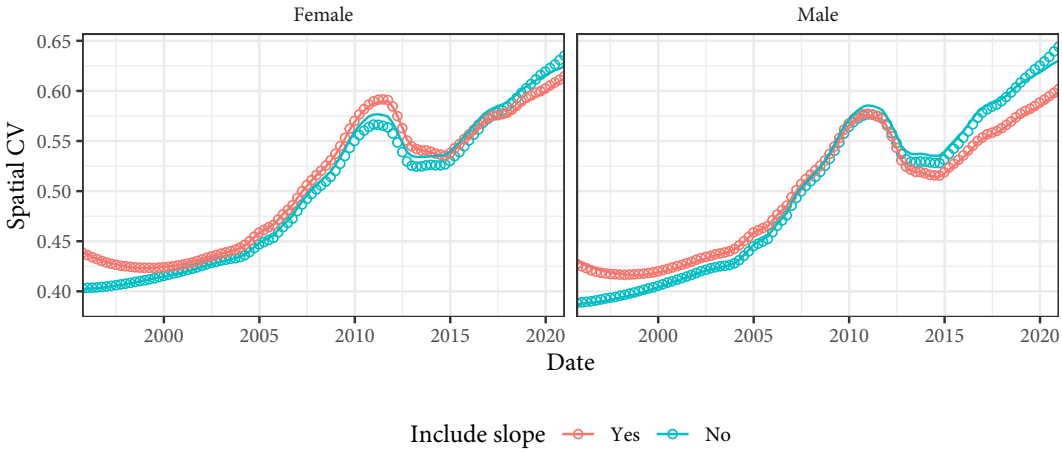

Figure 3.10: Spatial coefficients of variation in incidence rates by sex for four models from this analysis. Lines correspond models without autoregressive terms, and open circles correspond to models with autoregressive terms.

UNAIDS estimates. My model predicted a later, lower peak than the UNAIDS estimates, but the two sets generally converged in recent years. Between 2010 and 2021, the average difference in annual new infections (rounded the nearest hundred) between the UNAIDS estimates and the model without autoregressive and linear terms was 2,310.

There is also evidence that the share of new infections that are among women is increasing in sub-Saharan Africa (Risher et al., 2021). Figure 3.11 plots the sex ratio of incidence over time from each of the four models. While the models without linear terms inferred increasing sex incidence rate ratios (IRRs), matching UNAIDS well in recent years, the models with linear terms inferred decreasing sex IRRs, conflicting with most other data sources. The sex IRR in this work is inferred using a model that allows women's transmissibility relative to men to change linearly over time, while in UNAIDS' model the increasing sex incidence rate ratio in recent years is a fixed assumption.

Figure 3.8 showed that the transmission rate models that were constant or log-linear with respect to time were never identified as best using RMSE, but understanding the extent to which the inferred incidence trends are determined by the transmission rate model, the epidemic model, or the data is important. Figure 3.12 plots inferred incidence in Lilongwe among women and men assuming that the HIV transmission rate was constant with respect to time within each district alongside the model that was selected for the remainder of the analysis. Figure 3.13 plots the inferred transmission rates among men for these two models. Although the transmission rate in the "Constant" model was constant with respect to time,

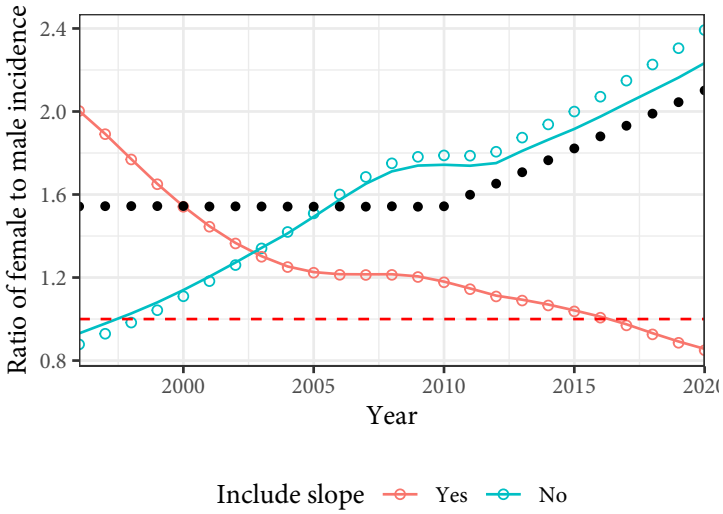

Figure 3.11: The ratio of incidence among women to that among men from four models compared to UNAIDS. Lines correspond models without autoregressive terms, and open circles correspond to models with autoregressive terms. Black points are UNAIDS assumptions.

the two models inferred similar series of incidence, demonstrating that the incidence is either truly constant well-identified by data or determined by the dynamics of the epidemic model.

Given these results, I selected a model with the following configuration to be the default model for all further analysis: no linear term with respect to time, one degree of differencing, an order-two spline with five-year intervals between knots, and no autoregressive term. Figure 3.14 plots illustrative predictions from this model for the Lilongwe district of Malawi, while Chapter 5 provides a more detailed exploration of the results. This model had among the best fit, was consistent with existing national-level estimates from Spectrum/EPP, and estimated regional sex IRRs that were consistent with other epidemiological evidence. Although I was unable to identify a single best model, this configuration balanced the characteristics identified by the empirical model comparison study with desiderata from the subjective comparison. Of particular note is the omission of the linear term in the transmission rate model; the linear trend dominated inferred transmission rates in a way that resulted in implausibly low and steady incidence forecasts.

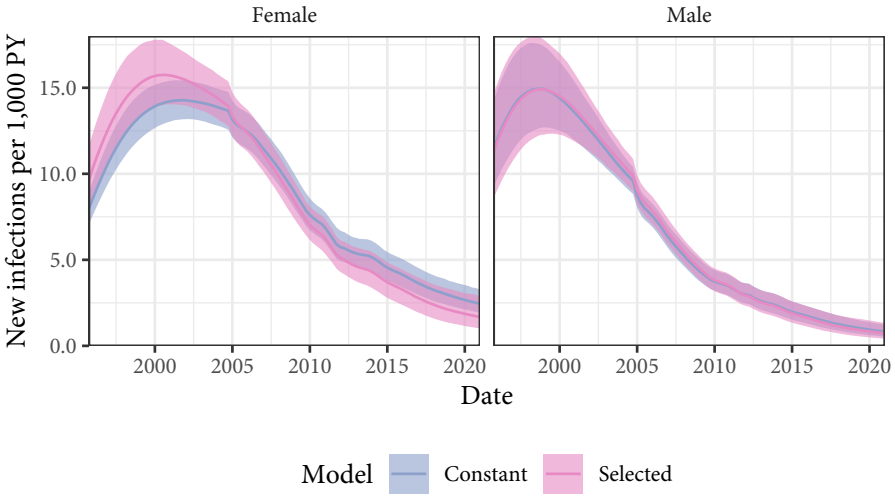

Figure 3.12: Estimated HIV incidence among men and women in Lilongwe for two models with varying transmission rate parametrisations.

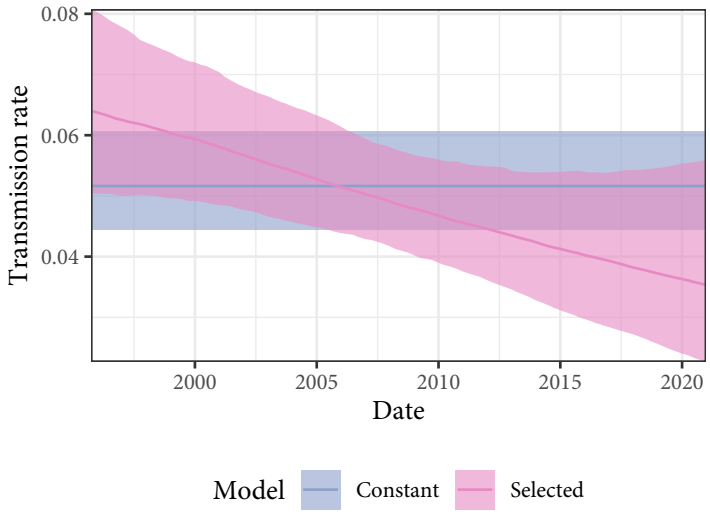

Figure 3.13: Estimated HIV incidence among men in Lilongwe for two models with varying transmission rate parametrisations.

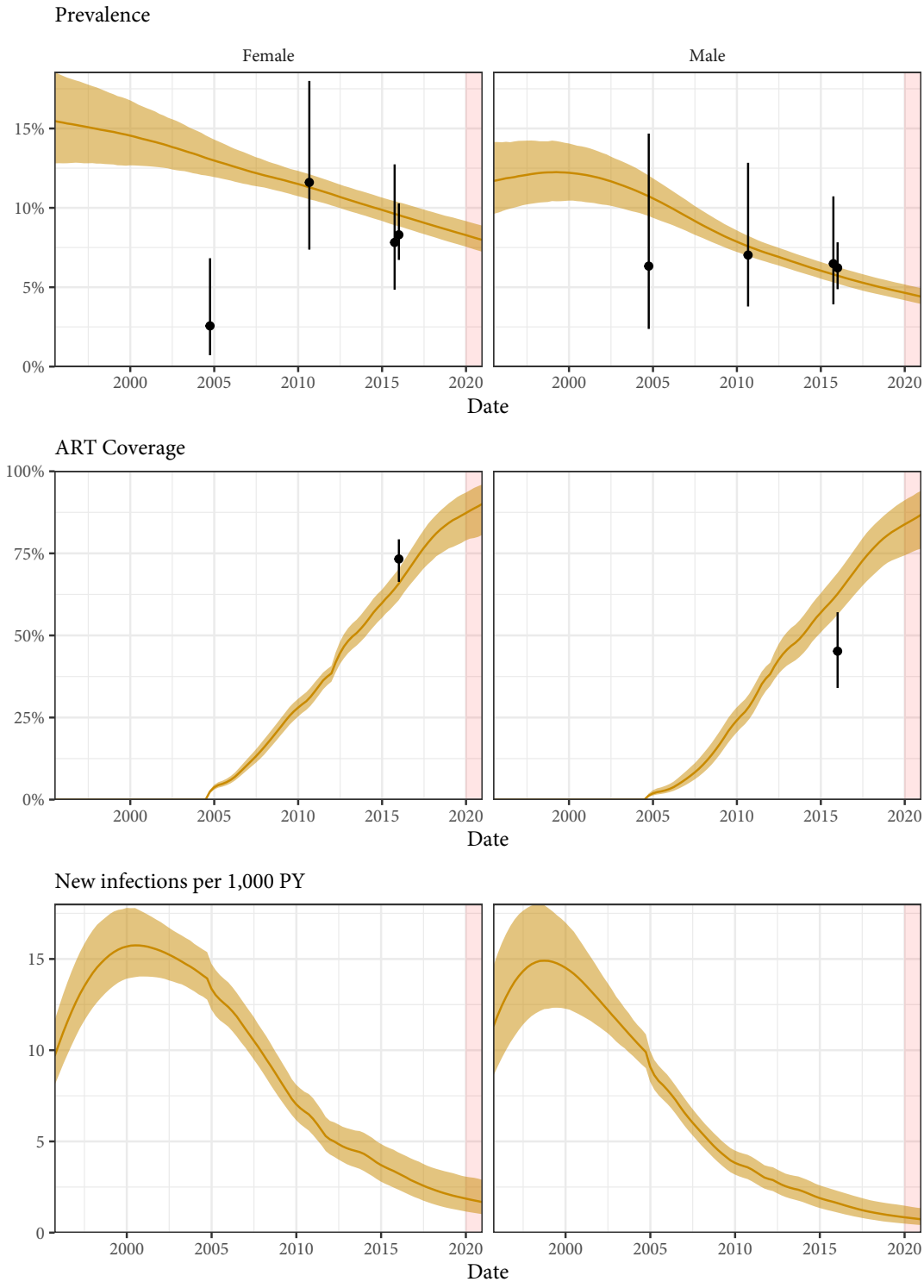

Figure 3.14: Illustrative predictions of HIV prevalence, ART coverage, and HIV incidence in Lilongwe by sex from the final model used in this analysis. Yellow lines correspond to posterior medians, and yellow regions correspond to 95% posterior credible intervals. The red region represents the time point after which all data were held out.

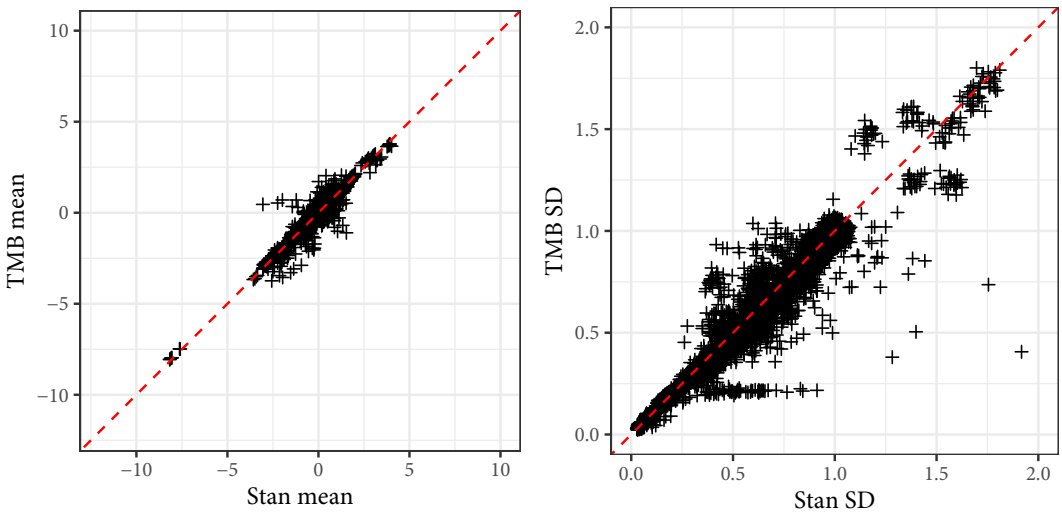

Figure 3.15: Comparison of posterior parameter means and standard deviations between Stan and TMB (the approximate strategy described in Chapter 2.

3.3.3 Inference strategy

Taking the final configuration from Section 3.3.2, I re-fit the model to each of the six forecasting horizons and an additional dataset with no data held out using NUTS. Across the seven horizons, the approximate strategy fit in an average of about 2.5 hours, while NUTS produced samples in an average of about 28.0 hours. Comparing each pair of corresponding fits, Stan took approximately 13 times longer than TMB on average. When fitting with NUTS without holding any data out, the average number of effective samples across all parameters was 2,157, and the median was 2,094. Of the 1,389 parameters, 1,384 (99.6%) had $\hat{R}$ values between 0.9 and 1.1, suggesting that the model had converged.

Figure 3.15 plots posterior means and standard deviations of all model parameters excluding hyperparameters with variance fixed to be zero in the approximate inference strategy. The posterior means correlated closely across the seven horizons, but TMB slightly underestimated parameter variance. The Pearson correlation coefficients between the means and standard deviations were 0.98 and 0.96, respectively. Excluding fixed hyperparameters, the correlation between the posterior standard deviations was 0.98.

Finally, Figure 3.16 compares posterior median incidence and 95% credible interval size across all regions and both sexes from 2010 onward. Again, we can see that the posterior means were similar, while the uncertainty estimates varied. There did not seem to be as

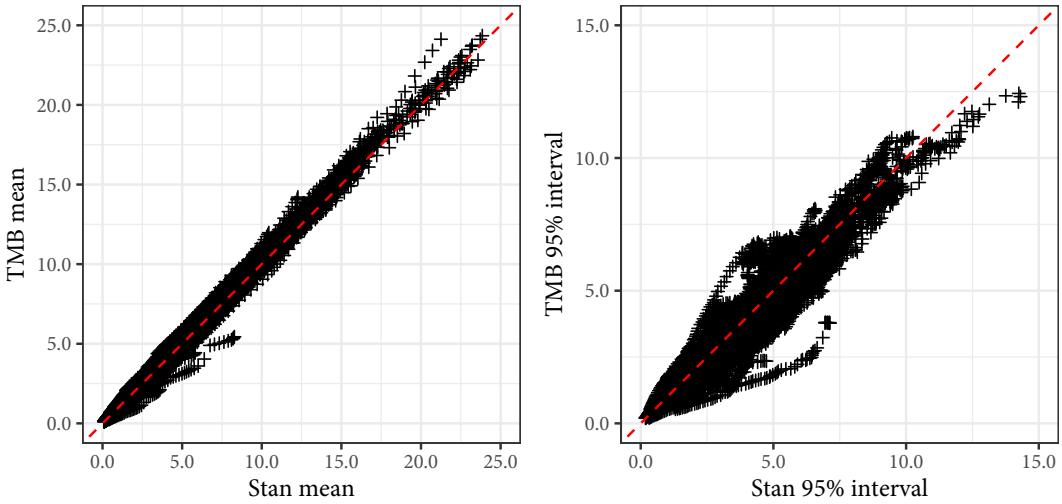

Figure 3.16: Comparison of posterior incidence means and standard deviations between Stan and TMB (the approximate strategy described in Chapter 2.

much systematic variation across the two inference strategies as in the parameter uncertainty. Overall, the correlations between the posterior means and 95% credible intervals of incidence were 1.00 and 0.96, respectively. Stan estimated substantially wider confidence intervals among women in Salima when using the longest forecasting horizon, which manifested as a distinctive line below the line of equality in the right panel of Figure 3.16. The average ratio of the Stan 95% CI intervals of incidence to those from TMB was 1.04, indicating again that Stan estimated slightly wider uncertainty.

3.4 Discussion

I sought to identify an optimal specification for the model described in Chapter 2 by fitting 146 unique model configurations to data from Malawi. Certain decisions were clearly identified by data, while others will need to be made subjectively by future users. I also identified determinants of convergence failure and compared fitting with an approximate inference strategy to fitting with NUTS.

I found that, consistent with Eaton & Bao (2017), an observation model for ANC testing data that incorporated overdispersion relative to a traditional binomial model offered better out-of-sample fit than the binomial model. Whereas Eaton & Bao (2017) added variance to the Gaussian approximation to the binomial distribution, the beta-binomial model

naturally incorporates count data. That said, the beta-binomial model produced posterior coverages that were higher than target values in both ANC facility and ART programme reporting. I have not taken into account autocorrelation across time in either dataset, which could have led the model to overestimate uncertainty.

I also found that region-specific non-linear temporal components improved fit relative to baseline intercept-only and linear transmission rate models. Although we do not expect the transmission rate of HIV to be constant or log-linear with respect to time, these simple models provide informative baseline estimates. Figure 3.12 shows that even inaccurate models can produce credible-looking estimates of incidence. This result indicates that much of the variation in prevalence data in Malawi can be explained by basic epidemic dynamics, as opposed to spatio-temporal variation in transmission rates, and suggests that we must be cautious interpreting inferred incidence.

Beyond the use of a beta-binomial ANC observation model and the inclusion of non-linear temporal components in the transmission rate model, the comparison study did not clearly identify any particular design decisions as superior to others. In one view, this indicates that the model's out-of-sample fit does not depend too closely on the specific modelling decisions. On the other hand, if two, equally good decisions imply meaningfully different incidence series, how does the user decide which to use? Further work quantifying the identifiability of models like these ones would be valuable

The qualitative insights offered by the cross-validation strategy I have described here can help with such subjective decisions. As Figure 3.9 shows, including a slope in the transmission rate model had a substantial effect on predicted incidence. Although current data in Malawi are well-described by rapidly decreasing transmission rates, building into the model a preference for decreasing transmission could reduce its sensitivity to changes in the future. The forecasts also show that, subjectively speaking, the slope-inclusive model understates uncertainty simply because of the dominance of the linear term. I also observed that, even when forecasting for five years, autoregressive terms were not necessary for controlling variance in predicted transmission rates.

I also used this analysis to address two important computational questions. First, I checked whether certain model configuration variables resulted in convergence failure more consistently than others. No single variable seemed to predict convergence failure, but the models

with the simplest transmission rate specifications were less likely to succeed than those with more complex specifications. This could be due to the inability of simpler specifications to accurately describe the data. This analysis did not provide many clear insights, but it could be used to help guide future use of the model.

Finally, I evaluated the validity and necessity of the approximate inference strategy from Section 2.6 relative to NUTS. As expected, fitting with NUTS took substantially longer than fitting with the approximate strategy. The results from the two strategies were, in general, similar, but I observed two notable trends. First, when the two strategies differed, TMB seemed to underestimate parameter uncertainty. This could result in overconfidence in estimated incidence. Second, the hyperparameters that had zero variance in the approximate strategy exhibited substantial variance when fit with NUTS. Ignored correlations between these parameters and others could have wide-reaching effects on point estimates and estimates of uncertainty.

With those caveats in mind, TMB was, on average, 12.8 times faster than NUTS and produced reliable results that were consistent with, if not identical to, those produced by Stan. Assuming that Stan is the gold standard, the results produced by TMB might be sufficiently accurate in some settings. In others, we might consider a strategy where testing is done in TMB and the final fit is produced with Stan.

This model comparison study had several weaknesses that might limit its ability to guide future modelling.[1] First, the experiment detailed here was conducted on data from Malawi, and wider generalisability is uncertain. Similar exercises could be conducted in any setting to which the user might want to apply the model.

I was also able to test only a subset of the many possible configurations supported by the model from Chapter 2. The transmission rate of HIV is the unknown quantity most closely related to HIV incidence, so that was the focus of this analysis. I did not, for example, interrogate the structure of the ART initiation model. Future work in this area could use the cross-validation strategy I outline here to perform even more exhaustive model comparisons.

Second, drawing any single conclusion from this analysis is difficult due to the large number of models fit and the general ambiguity of the results. Beyond a few broad structural

[1] I defer the more general discussion of weaknesses of the model and my thesis to Section 8.3.

decisions, the analysis did not indicate that any given specification choice was categorically better than another. Future users might still need to fit several models and compare the results subjectively to identify the model that best suits their needs. That said, the relatively thorough set of experiments conducted here can offer users some guidance on how these decisions should be made, in addition to relative confidence that the decisions are not being made arbitrarily.

It is also possible that the inference strategy comparison was not strictly fair. I have used the **tmbstan** library to fit the existing TMB model with NUTS. Anecdotally, I have found that, given statistically equivalent models written in TMB and Stan, **tmbstan** will fit the TMB model noticeably slower than **rstan** will fit the Stan model (both using NUTS). We might be able to fit with NUTS more efficiently with a native Stan version of the model.

Additionally, these models were fit using an approximate inference strategy, although I found that the estimates from the approximate strategy correlated closely with those produced by NUTS. Given the additional computational burden incurred by NUTS, fitting 876 times without the approximate strategy was not practical. However, as shown in Figure 3.16, I found that a subset of TMB fits were closely correlated with the corresponding NUTS fits, suggesting that the comparison was fair.

Despite these weaknesses, this model comparison study can help guide the use of the model, as well as the design of future models of HIV burden. I tailored the cross-validation strategy to the specific data streams that we expect to continue to be available, providing an empirically justified basis for otherwise-arbitrary design decisions.

As I have demonstrated here, the small decisions necessary to parametrise models like this one can have dramatic effects on inferences, so I suggest that thorough model comparison exercises should not be viewed as post-hoc sensitivity analyses and relegated to supplementary material. They should, instead, be seen as an essential step in the model development process.

Chapter 4

Effect of Spatial Transmission on Inferred HIV Incidence

Chapter 3 addressed a number of empirically testable model design questions, but it notably did not examine the effects of spatial mixing. For computational efficiency, every configuration in that chapter assumed that incidence in a region was a function of prevalence in that region alone. In this chapter, I examine the effects of incorporating various assumptions about the spatial dynamics of HIV transmission and discuss the advantages and disadvantages of including these dynamics in the model.

4.1 Introduction

HIV transmission is a fundamentally spatial process: a susceptible individual cannot be infected with HIV without coming into contact with the virus. At the population level, the spatial dynamics of HIV are apparent on both the continental scale as the virus moved from central Africa to southern Africa over the course of decades (Faria et al., 2014) and the local scale as clusters of infections might emerge among people who inject drugs over the course of weeks (Conrad et al., 2015).

These dynamics are not considered by any of the spatially structured models described in Section 1.5.3 at least in part because population-level data attributing new HIV infections to specific areas are virtually non-existent. Other work has estimated the influence of

mobility on HIV in SSA, but these studies used methods and data that limited their ability to produce predictions for the general population (Kate Grabowski et al., 2020; Okano et al., 2020; Ratmann et al., 2020; Valdano et al., 2021). Recently, viral genetic sequences have been used to identify transmission pairs in prevention trials (Hall et al., 2021; Novitsky et al., 2020). However, genetic data are not widely available, and the methods for incorporating them into large-scale, population-level epidemic models are not obvious. Population-level prevalence data necessarily reflect underlying spatial dynamics, but because an observed HIV infection in a particular location could be attributable to any number of regions, these data do not facilitate inference of spatial dynamics.

Even if reliable data on transmission between regions were available, inferring transmission rates in a spatially structured epidemic model would still present a substantial computational problem. Allowing the epidemic to spread over regions makes incidence in each region a function of past incidence in all other regions, inducing complex covariance between the parameters used to infer transmission rates when we condition on data. This is a particular concern for the approximate inference strategy described in Section 2.6 because the efficiency of that method depends on the density of the model's Hessian, H from Equation (2.48). If incidence in each region depends on past incidence in all other regions the Hessian with respect to the transmission rate parameters will be dense and inference will be slowed.

Beyond computational concerns, the effects of including spatial mixing dynamics in this model are unknown. None of the models described in Section 1.5 explicitly incorporated transmission over spatial units, so it is impossible to say in advance what the effects of including or excluding spatial dynamics will be. Depending on the specific data the model is fit to, spatial mixing could result in more or less uniform distributions of incidence.

I conducted a final model comparison study examining the effects of incorporating spatial transmission dynamics into the model described in Chapter 2. I fit the final model from Chapter 3 to the same datasets with a series of assumptions about the degree of spatial mixing across districts. I used these results to measure the effects of spatial mixing on both national and subnational incidence, in addition to measuring the impact on computation time. Using the cross-validation strategy described in Chapter 3, I examined whether one assumption about the extent of mixing was better supported by data than the others. For that

particular question, I hypothesised that, because the available data reflect the underlying spatial dynamics of HIV only indirectly, the optimal spatial transmission assumption would not be identifiable from the available data.

4.2 Methods

I fit the final model from Section 3.3.2 with five assumptions about the spatial dynamics of HIV to data with the same six forecasting horizons. Using the same cross-validation strategy described in Section 3.2.2, I identified whether any assumption offered superior out-of-sample fit and compared estimates of HIV incidence over time for alternative spatial transmission assumptions

4.2.1 Data

The data used in this analysis are described in Section 1.4. I refer the reader to Table 3.1 for an overview of the data sources incorporated into the observation model. As in that chapter, I constructed a cross-validation dataset by identifying all data collected after six annual forecasting horizons (1 January 2015 through 1 January 2020) as out-of-sample.

4.2.2 Model configuration

I fit the final model configuration described in Section 3.3.2 with TMB. This model used the beta-binomial ANC observation model and specifies the transmission rate model with no linear term with respect to time, one degree of differencing, an order-two spline with five-year intervals between knots, and no autoregressive term. To re-emphasise the conclusion of Chapter 3, this is not necessarily the best model, but it offered competitive out-of-sample fit and resulting estimates were consistent with other national-level estimates.

4.2.3 Assumptions about spatial dynamics

I fit this model with five assumptions about the proportion of contacts people residing in region r have with people residing in region j. Recalling the log-linear model of incidence

from Equation (2.17), the proportion of contacts among women in region r that are with men in CD4 bin c is

$$\sum_{j \sim r} \sqrt{\frac{w(r,j)w(j,r)}{N_{j,0}N_{r,1}}}(I_{j,0,c} + (1-\omega)A_{j,0,c}), \tag{4.1}$$

where $j \sim r$ denotes the set of regions that are adjacent to r inclusive of r. For adjacent regions r and j, $w(r,j)$ is defined as

$$w(r,j) = \begin{cases} w_0 & r = j \\ \dfrac{(1-w_0)}{\|\{j \sim r\} \setminus r\|} & r \neq j \end{cases}, \tag{4.2}$$

where w_0 is the share of contacts that are made with individuals from the home region. The remaining share is divided equally among the neighbours of the home region. I used four values for w_0 (100%, 75%, 50%, and 25%) and a special case in which contact with each neighbour is as likely as contact with the home region, which can be expressed as

$$w(r,j) = \frac{1}{\|\{j \sim r\} \setminus r\|}. \tag{4.3}$$

4.2.4 Model comparisons

I estimated the effects of spatial transmission on computation time, inferred incidence, and out-of-sample fit to data. I measured the number of hours required to fit each model and sample from the approximate posterior distributions and compared the resulting times. To assess the effects on inferred incidence, I examined the absolute and relative differences in both national- and district-level incidence. Finally, I used the cross-validation strategy described in Section 3.2.2 to measure the effects of spatial transmission on out-of-sample fit to ANC faciltiy data and ART programme data.

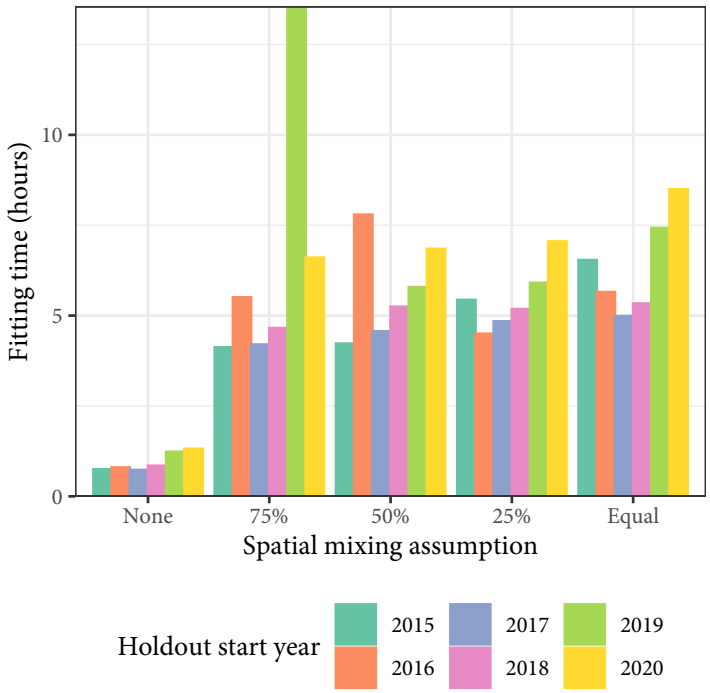

Figure 4.1: Fitting time for five assumptions about spatial transmission by forecasting horizon relative to a model with no spatial mixing.

4.3 Results

Figure 4.1 plots the times to fit and sample from each model. The models with spatial dynamics took approximately six times longer than the model with no spatial transmission. Excluding the two outlying times, the relative execution times seemed to increase relative to the baseline model as the homogeneity of spatial mixing increased, from a mean of 6.9 times longer than baseline to a mean of 5.7 times longer than baseline.

Figure 4.2 presents posterior median quarterly incidence risk per 1,000 people at the national-level. Differences in the absolute level of incidence were larger in earlier years and smaller in later years when data were more plentiful. The estimated posterior median incidence among women ranged from 3.7 to 4.3 in 1999 and from 0.5 to 0.6 in 2021.

Framed in relative terms, these differences increased. Figure 4.3 plots the percent change in quarterly incidence risk between each spatial mixing assumption and the model with no spatial mixing (all when including data up to the beginning of 2020). Although the absolute differences were lowest in recent years, the relative differences were at least as large. The model that fixed only 25% of contacts to be from the home region estimated

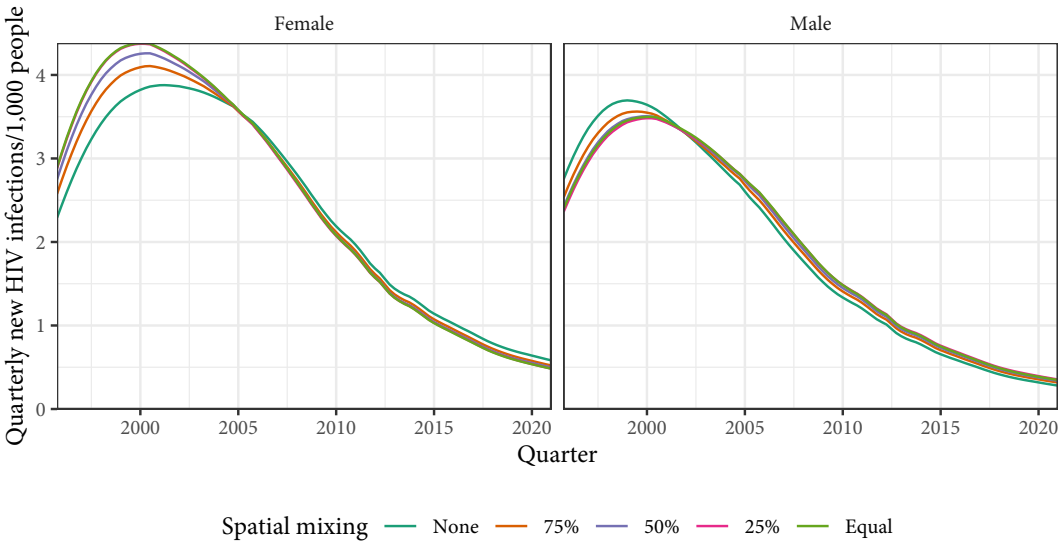

Figure 4.2: Posterior median incidence risk by sex and time for five assumptions about spatial transmission from models that included data up to the beginning of 2020.

26.2% higher incidence among men at the end of 2020 and 14.2% lower incidence at the beginning of the projection. Trends were similar but reversed among women. The impact of spatial mixing on district-level incidence was similar but spatially heterogeneous. Figure 4.4 illustrates this heterogeneity by comparing estimated incidence in Nkhata Bay and Ntcheu, the districts in which the 25% mixing assumption had the largest and smallest effects on incidence in men in 2020, respectively. Incidence among men in Nkhata Bay at the end of 2020 ranged from 0.8 per 1,000 person-years without spatial mixing to 1.4 with homogeneous mixing across districts. In contrast, the same range in Ntcheu was from 0.7 to 0.7. These effects were consistent across all 28 districts. Figure 4.5 plots the percent change in inferred incidence relative to no spatial mixing by region, sex, and date in the model with the latest forecasting horizon. Values greater than 0.0 indicate that the spatial mixing assumption resulted in higher incidence than the baseline assumption of no spatial mixing. There was considerable heterogeneity in the effects of incorporating spatial transmission across time, sex, and space. Among men, the standard deviation of percent changes in 2021 ranged from 0.22 when weight was allocated equally across regions to 0.11 when w_0 was 75%. Consistent with the national-level result in Figure 4.3, Figure 4.5 clarifies that the smaller absolute differences still resulted in greater relative differences. The consistent spike in 2005 was Chiradzulu; in models with spatial transmission, the preventative effects of rapid increases in treatment coverage in that district have been muted by relatively slow

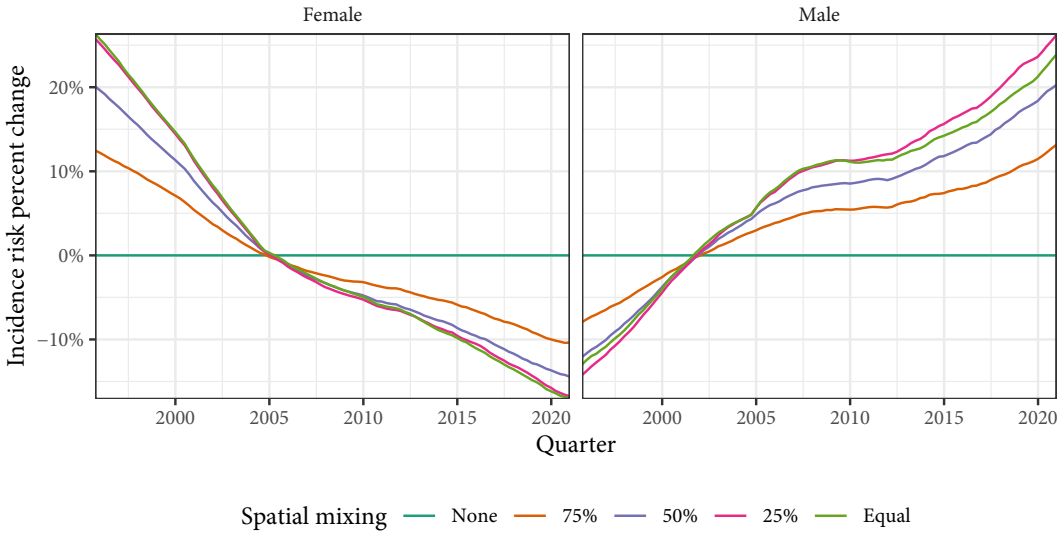

Figure 4.3: Incidence ratios by sex and time for five assumptions about spatial transmission from models that included data up to the beginning of 2020 relative to a model with no spatial mixing.

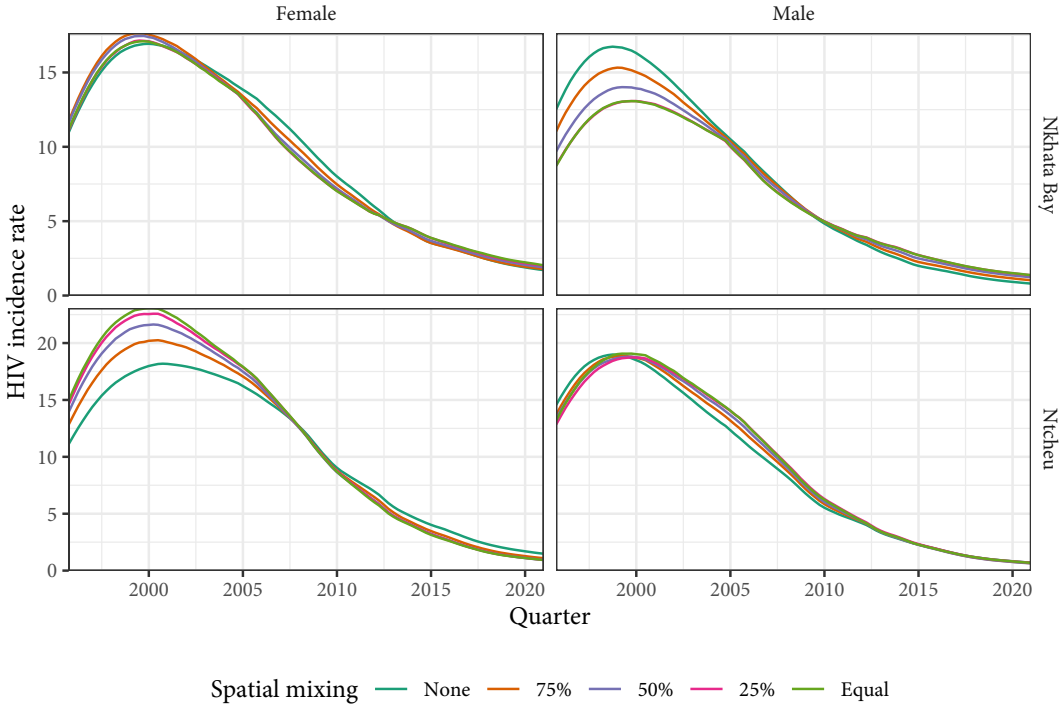

Figure 4.4: Posterior median incidence in two districts of Malawi by sex and time for five assumptions about spatial transmission from models that included data up to the beginning of 2020.

increases in adjacent regions.

To visualise the patterns in the effects of mixing more clearly, Figure 4.6 presents hexagonal tile maps of the incidence ratios from the model that included data up to the beginning of 2020 by sex at the latest projected time (Q4 2021) relative to the model with no spatial mixing. Including spatial transmission dynamics decreased inferred incidence among women and increased it among men. When w_0 was 75%, the mean percent change in incidence across districts among men was 10.4%, while among women it was -12.1%.

The observed sex disparity could be attributable to any number of dynamics, but one hypothesis is that mixing has effectively served as a smoother. By allowing the epidemic in one district to be influenced by its neighbours, the model required less sex disparity in incidence to fit to the same prevalence data. This theory is corroborated by Figure 4.7, which plots the sex IRRs inferred by the five models that used data up to the beginning of 2020. The model with no spatial mixing estimated that the sex IRR of HIV was higher and increased more quickly over the projection period. This observation supports but does not confirm the hypothesis that spatial mixing smoothed the estimates in this study.

The cross-validation scheme did not lend much support to one spatial mixing assumption or another. Figure 4.8 presents out-of-sample RMSE by holdout start date and degree of spatial mixing. The mean ratio of RMSE across the four spatial models and six horizons to the model with no mixing was 1.0 with respect to the ANC data and 1.2 with respect to the ART patient counts. The model without spatial transmission fit best to out-of-sample ART patient counts, but there was very little difference in fit to ANC prevalence.

4.4 Discussion

In this chapter, I evaluated the impact of incorporating spatial transmission dynamics into the final model identified in Chapter 3. I found that spatial mixing increased computation time by a factor of about six and had moderate, heterogeneous effects on incidence estimates. I was not able to definitively identify whether any of the models tested here provided better out-of-sample fit than the others.

Incorporating spatial mixing dynamics substantially increased the time needed to fit the

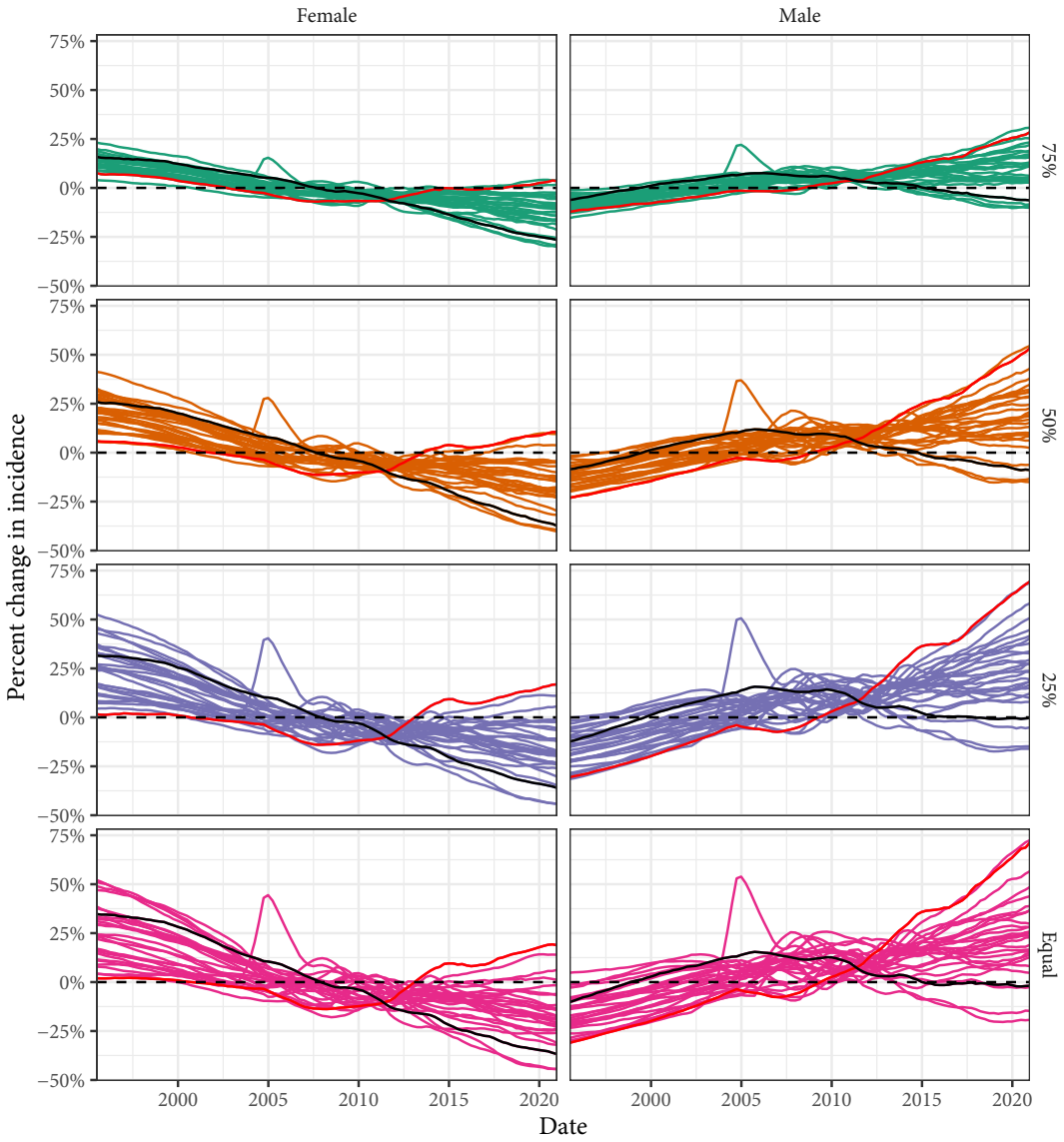

Figure 4.5: Percent change in estimated estimated by sex and district for four assumptions about spatial transmission compared to a model with no spatial transmission. Each line represents a district. Ntcheu and Nkhata Bay are overlaid in black and red respectively.

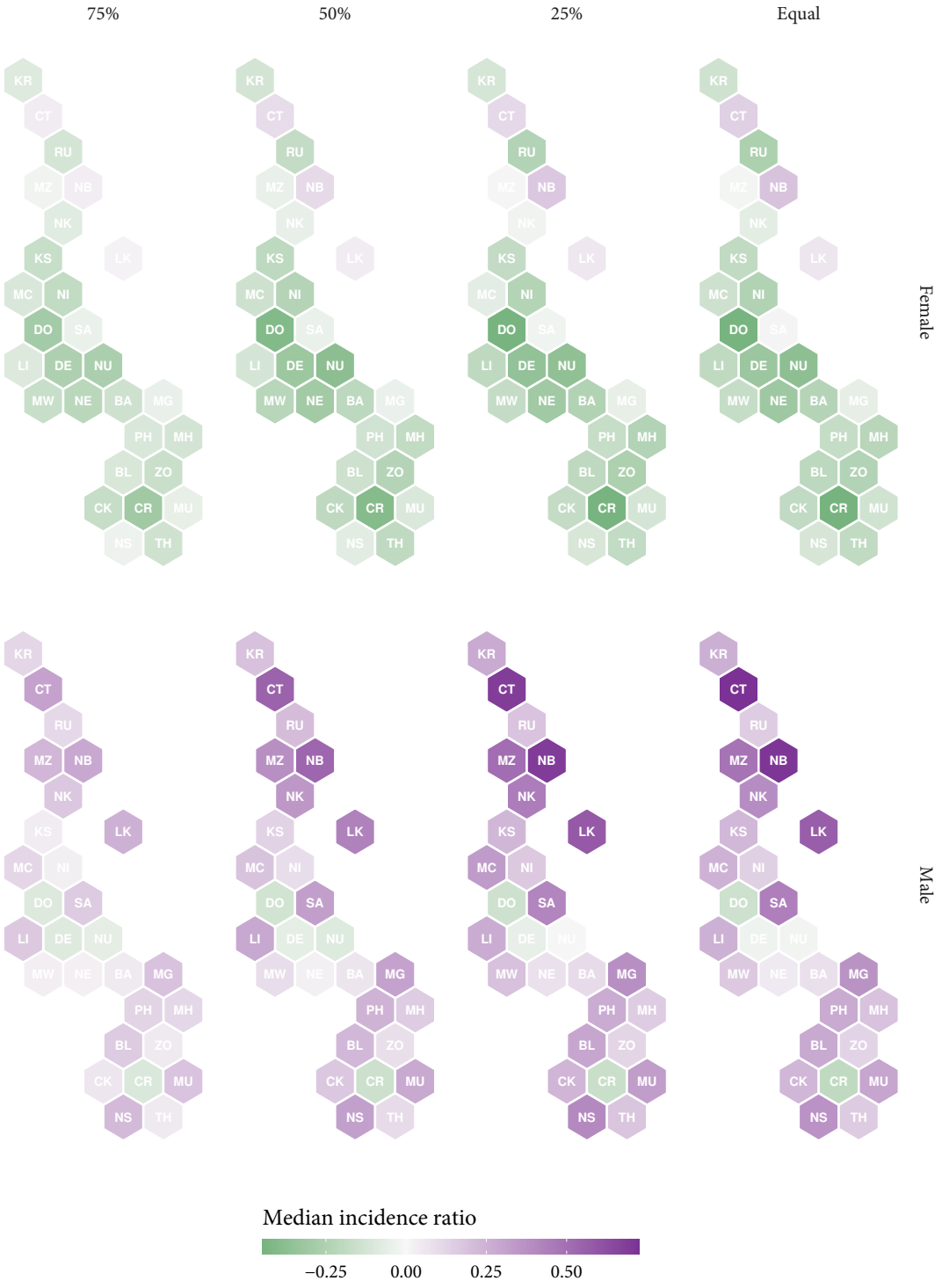

Figure 4.6: Percent change in posterior median incidence under four assumptions about spatial transmission relative to a model without spatial transmsision by sex at the end of 2021.

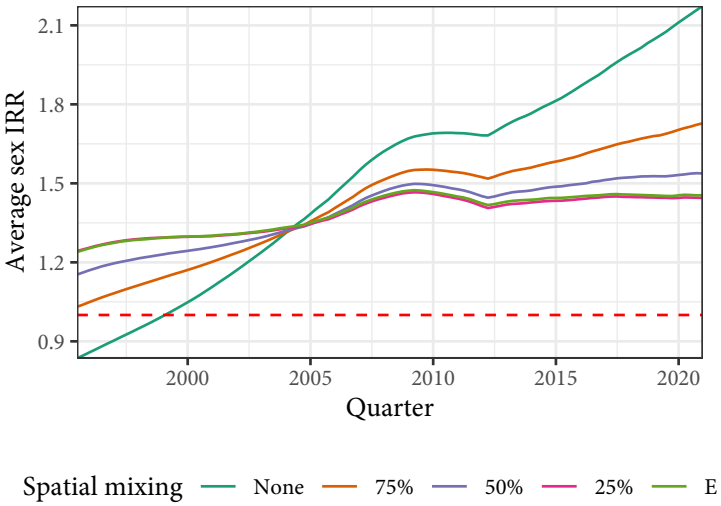

Figure 4.7: Inferred ratio of female incidence to male incidence under five assumptions about spatial transmission averaged across all districts. The red dashed line represents equal incidence between men and women.

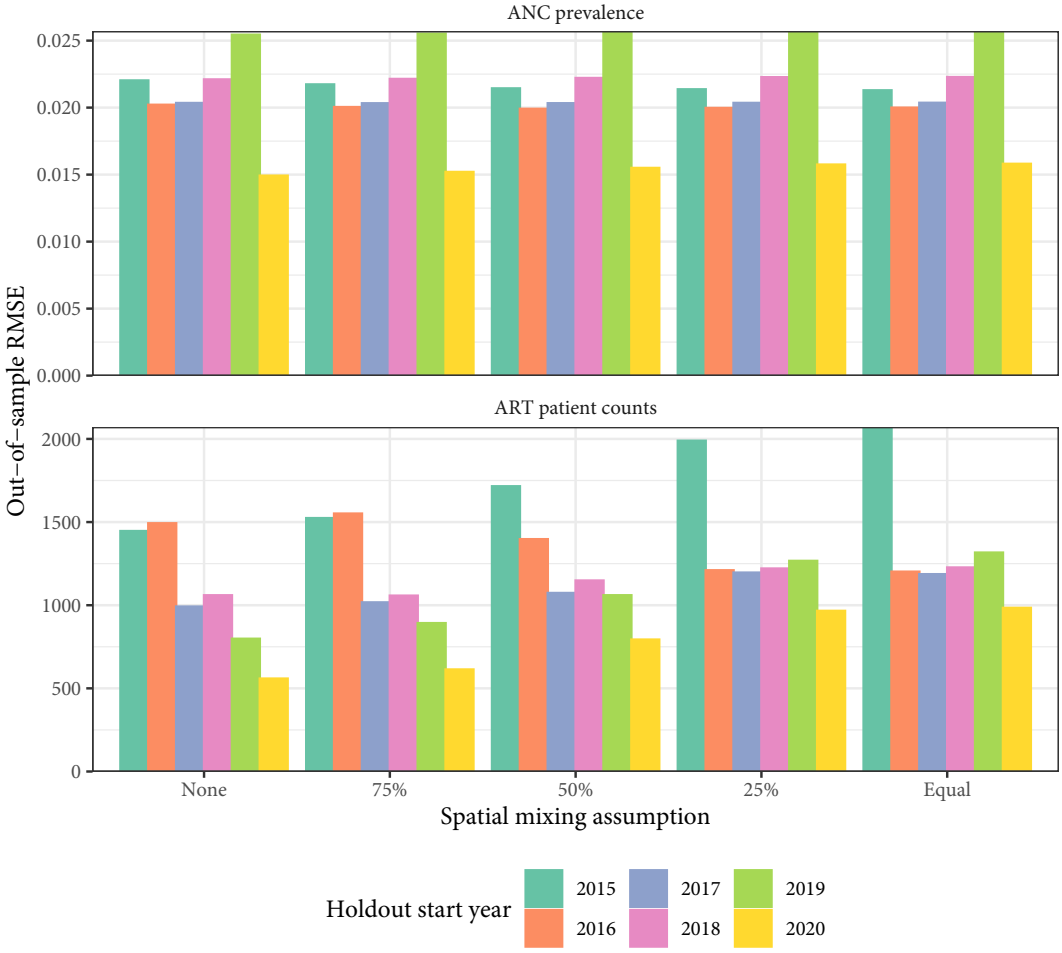

Figure 4.8: Out-of-sample RMSE by holdout start date for five assumptions about spatial transmission.

model and sample from the posterior. This was expected because allowing incidence in one region to depend on past incidence in other regions creates dependence between the parameters used to estimate transmission rates across space, resulting in an approximately six-fold slowdown.

The incorporation of spatial mixing had heterogeneous effects on inferred incidence across region, time and sex. Models with spatial transmission inferred increased incidence among men and decreased incidence among women in recent years. Changes were consistently larger in certain regions of Malawi; in Chiradzulu, the effects of early, rapid ART scale-up were dampened by relatively slower scale-up in nearby districts. In general, the net effect of spatial mixing was to smooth incidence across space and sex.

There were notable differences in inferred sex IRRs over time; models that included spatial transmission estimated more gradual changes in the sex IRRs, resulting in lower IRRs at the end of the projection. Given the consistency between the inferred IRRs from the previous and UNAIDS estimates, a more gradual change does not seem desirable. In total, the average district-level percent change in incidence when moving from a model with spatial transmission to one with 75% of contacts staying within the home region was about 10% for men and -12% for women.

To help make decisions about spatial mixing in an empirically justified way, I measured the impact of varying assumptions about spatial dynamics on out-of-sample fit. All models fit comparably well to out-of-sample ANC data, while the model without spatial transmission fit better than the others to ART patient counts. In total, the cross-validation exercise seemed to indicate weakly that the best-fitting model excluded spatial dynamics. The ambiguity of the cross-validation exercise was not surprising. Prevalence and patient count data do reflect the true, underlying spatial dynamics of the epidemic but can only do so obliquely.

These results do not lend themselves cleanly to one final recommendation. HIV transmission is fundamentally spatial, but the model that the cross-validation exercise identified as best had no spatial transmission. This assumption is reasonably consistent with recent analyses of genetic data from prevention trials, which suggest that HIV transmission is highly local in sub-Saharan Africa (Hall et al., 2021). I also observed in Chapter 3 that inferred sex IRRs from the model without spatial mixing dynamics aligned closely with

external national level estimates and that spatial mixing smoothed those IRRs dramatically. For these reasons, in the rest of the document, I will defer to the cross-validation results and use the model that omits spatial dynamics. Future users will need to weigh the substantial additional computational cost with the theoretical benefits to decide for themselves whether to model spatial transmission directly.

This analysis is subject to several important weaknesses. First, I tested only one, relatively simple formulation for the spatial weights. This parametrisation could be replaced with that of Wakefield, Dong, & Minin (2017) or Held, Höhle, & Hofmann (2005) or with any of the models described by Meredith et al. (2021). All of those parametrisations allow for transmission between any pair of regions, not just neighbours. It is difficult to say in advance what the effect of this change would be, but given the smoothing observed here, it is possible that those results would be smoothed further. Given the ambiguity of the cross-validation exercise, I suspect that the parameters for these models would be poorly identified, but alternative parametrisations are still worth considering. We could also incorporate external information (genetic data, mobile phone records, etc.) that might lead to a more empirically justified set of weights.

Second, without direct measurement of incidence, the results of the comparison conducted here are difficult to interpret. It is impossible to know which of the incidence curves produced here are closest to the truth. We can only say that these assumptions have had an impact, not that any assumption was more accurate than another.

Despite these weaknesses, the comparisons I have presented in this chapter offer valuable insight into the effects of incorporating spatial transmission dynamics into large-scale models of population-level HIV. Future work investigating how to inform our assumptions about spatial mixing with real-world data and how to identify the parameters governing spatial dynamics is necessary.

Chapter 5

Estimates of Key HIV Indicators in Malawi, 1995–2021

Using the results from the specification tests in Chapters 3 and 4, I produced descriptive results in Malawi from 1995-2021. In this chapter, I describe trends in HIV prevalence, incidence, and treatment coverage over time, sex, and space and provide comparisons to national and subnational estimates published by UNAIDS. By estimating district-level trends in HIV epidemic indicators, this model facilitates identification of spatial gaps in treatment programmes. The previous chapters constitute a thorough discussion of the motivation and methodology for this model, so I have included no Introduction section and only a minimal Methods section.

5.1 Methods

I fit the model described in Chapter 2 to data from Malawi collected between 1995 and 2020 with forecasted inputs for 2021. The data were those described in Section 3.2.1, and the model configuration was the one used in Chapter 4 without any spatial transmission. No data were held out because doing so would have unnecessarily restricted the model's ability to forecast. I sampled from the full joint posterior using NUTS. I compared the national-level estimates of incidence, prevalence, and ART coverage to those generated by UNAIDS and the subnational estimates of incidence to those generated by Naomi (UNAIDS, 2020). Whenever applicable, I present posterior median estimates with 95% credible intervals in

the following format: `median (95% CI)`. I measured trends in epidemic indicators using percent change.

5.2 Results

5.2.1 Fit to data

The model fit well to most available data. Figure 5.1 presents estimated prevalence, ART coverage, ANC prevalence, and ART patients counts alongside the data used to fit to each metric for Lilongwe District. Outside of the 2004 DHS, the model matched the prevalence data closely and fit similarly well to each ANC site.[1] It smoothed across sex in ART coverage, resulting in slightly lower coverage among women and slightly high coverage among men. However, it fit essentially perfectly to the ART patient count series. I observed qualitatively similar fits across the 28 districts. Figures A.1 to A.28 present this plot for all 28 districts.

5.2.2 National-level estimates

The model estimated that, at the end of 2021, 7.4% (7.2% to 7.7%) of adults aged 15-49 in Malawi were living with HIV, of whom 91.9% (88.9% to 94.1%) were on ART. There were an estimated 14,200 (11,600 to 17,200) new infections in this population in 2021. Between the beginning of 2010 and the end of 2021, adult prevalence decreased by 33.7% (30.7% to 36.5%), from 11.1% (10.9% to 11.4%) to 7.4% (7.2% to 7.7%). Over that period, ART coverage increased by 271.5% (256.2% to 286.2%), from 24.7% (24.0% to 25.4%) in 2010 to 91.9% (88.9% to 94.1%) in 2021. Finally, population-level incidence risk (defined as new infections in a year divided by susceptible population at the start of the year) decreased by 76.0% (71.8% to 80.1%) between 2010 and 2021, from 6.6 (6.3 to 6.9) per 1,000 people to 1.6 (1.3 to 1.9) 1,000 people.

The model inferred that the ratio of female to male incidence increased substantially over the projection period, with an increasingly large share of new infections occurring in

[1] There is a known issue in the 2004 DHS in Lilongwe. The response rate for HIV testing there was only 39% (National Statistical Office - NSO/Malawi & ORC Macro, 2005), suggesting that the survey estimates could be substantially biased.

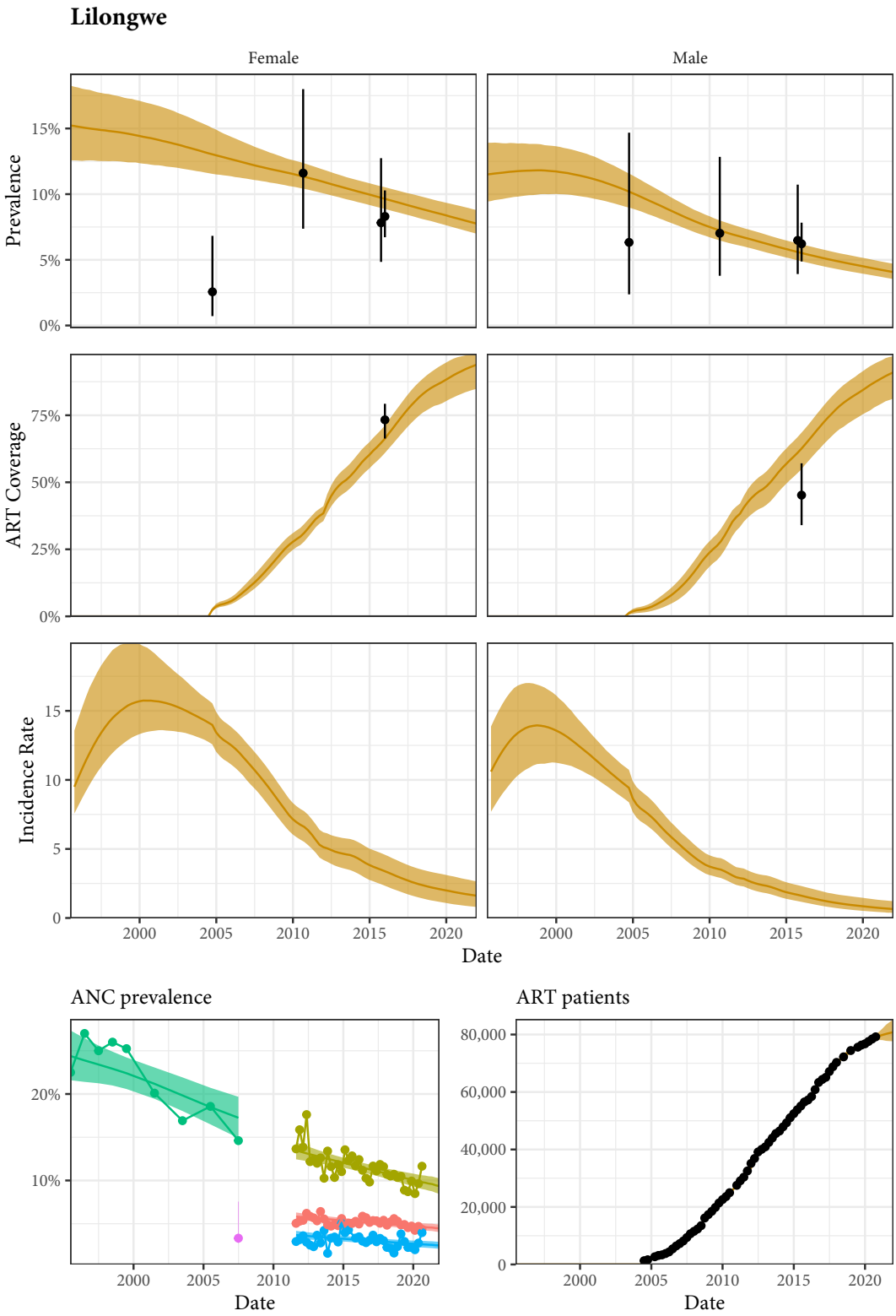

Figure 5.1: Estimated prevalence, ART coverage, ANC prevalence, and ART patient counts in the Lilongwe district of Malawi with household survey data, ANC facility data, and programmatic reporting data (points). Different colours on panel "ANC prevalence" indicate different ANC facilities.

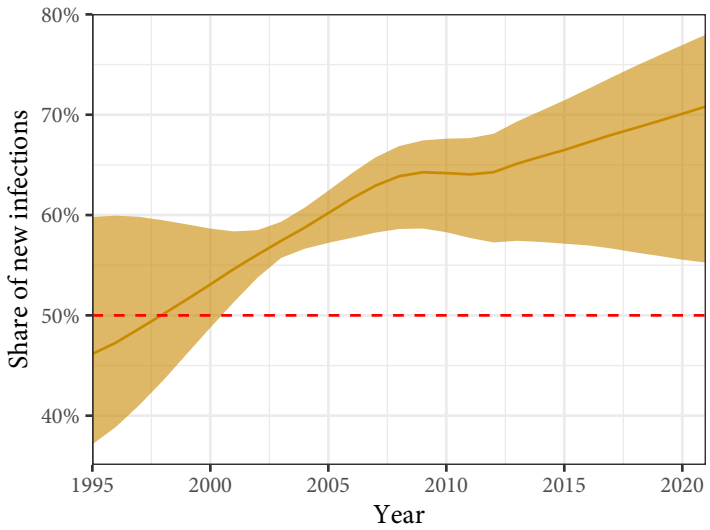

Figure 5.2: Estimated share of annual new HIV infections occurring among women in Malawi

women. Figure 5.2 presents the national-level share of new infections in ages 15-49 that were among women from 1995 through 2021. The model estimated that women accounted for 60.2% (57.2% to 62.4%) of new infections in 2005 (the year after the first household survey), compared to 70.8% (55.3% to 78.0%) in 2021. The model estimated that men LHIV in 2021 were as likely as women LHIV to be on treatment. The ratio of female to male coverage was 1.0 (1.0 to 1.1) in 2020, lower than UNAIDS' estimate of 1.1 Figure 5.3 presents this ratio from 2010 to 2021. This result is consistent with the smoothing in ART coverage observed in Figure 5.1.

Comparison to UNAIDS estimates

Figure 5.4 compares national-level prevalence, ART coverage, and annual new infections from my model to UNAIDS. Although all three metrics aligned closely in recent years, there were significant differences earlier in the epidemic. In 1996, the first full year of projection, my model predicted that there were 52,200 (48,300 to 56,600) new infections among adults aged 15-49, while UNAIDS predicted 80,000. UNAIDS predicted 17,400 new infections in 2020, and my model predicted 15,200 (12,600 to 18,300) new infections.

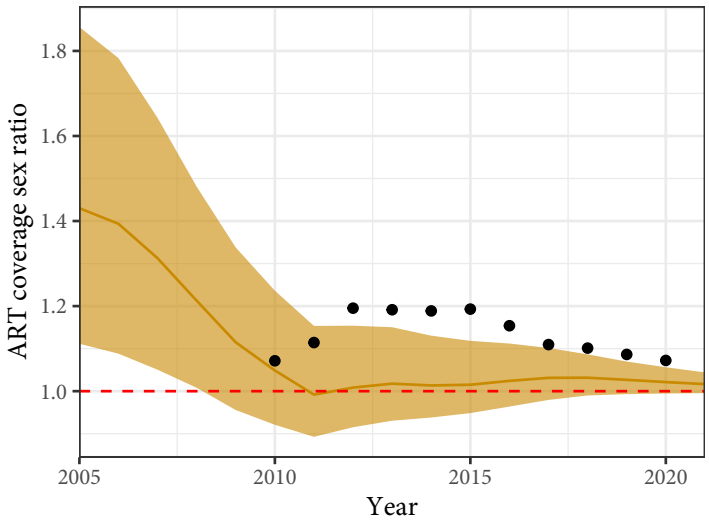

Figure 5.3: Estimated ratio of ART coverage among women to men in Malawi. Points are UNAIDS esimated ART coverage ratios.

5.2.3 District-level estimates

Figure 5.5 presents plots of the posterior distributions of prevalence, ART coverage, and incidence at the end 2021 for both sexes combined by district. Across the 28 districts, median prevalence was 6.8%, ranging from 14.9% (13.7% to 16.3%) in Mulanje to 2.2% (1.8% to 2.8%) in Ntchisi. Median ART coverage was 93.9% and proportionately more uniform, ranging from 96.6% (90.3% to 97.8%) in Mulanje to 84.2% (69.5% to 94.5%) in Zomba. Median incidence risk was 1.4. Incidence was highest in Mulanje at 3.5 (2.5 to 5.2) new infections per 1,000 people and lowest in Ntchisi at 0.4 (0.2 to 0.8).

Tables 5.1 through 5.6 present estimated prevalence, ART coverage, and incidence among adults aged 15-49 by sex and district of residence for the 28 districts of Malawi in 2010 and 2021, as well as the percentage change between the two time points. Figure 5.6 provides hexagonal tile maps of these values.[2] I provide a table of ISO-3166-2 codes for each district in Table A.1. Figures A.1 through A.28 provide estimates of incidence over time by sex for all 28 districts.

Figure 5.7 presents hexagonal tile maps of percent change in prevalence, ART coverage, and incidence between 2010 and 2021. There were substantial decreases in prevalence and incidence in all districts between 2010 and 2021, although these decreases were spatially

[2] Without any quantification these maps might be misleading; I provide them only to give the reader a loose sense of systematic spatial variation.

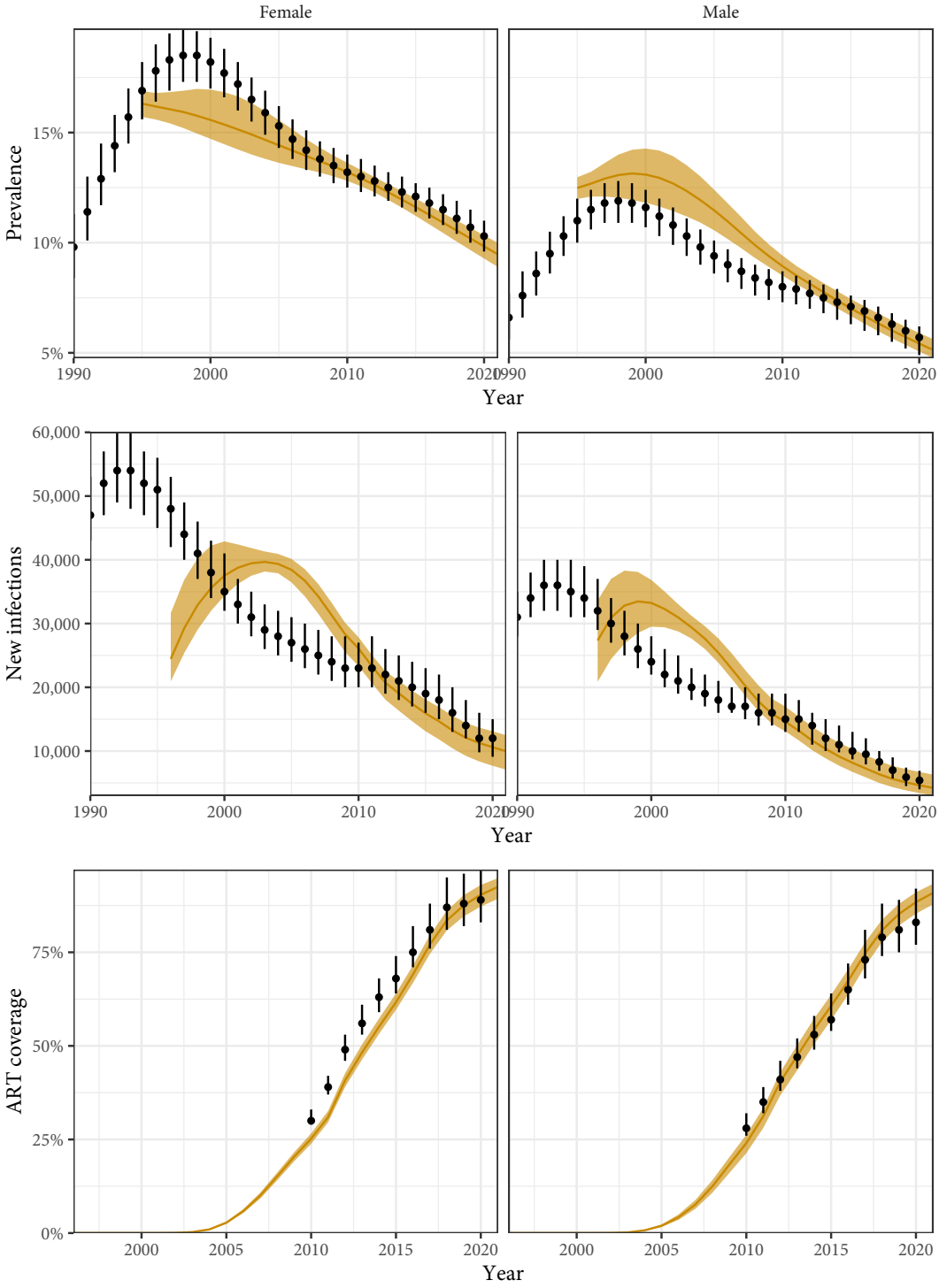

Figure 5.4: Comparison of estimated national-level annual prevalence, new infections, and ART coverage between UNAIDS (point ranges) and the model presented here (yellow regions). Note that UNAIDS ART coverage is among all adults.

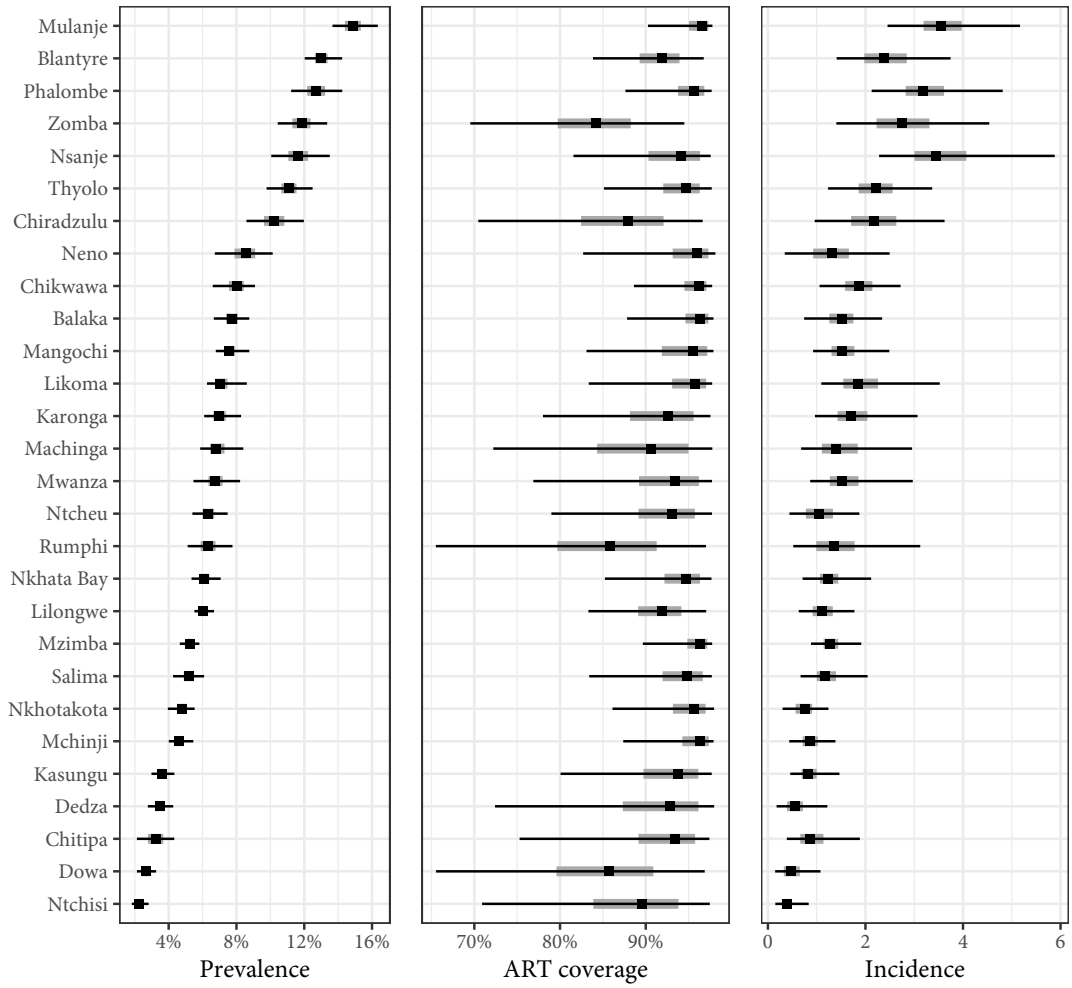

Figure 5.5: Ladder plots of posterior median (points) estimates of incidence risk, ART coverage, and prevalence with 95% and 50% credible intervals (lines and shaded regions, respectively). Districts are sorted vertically from highest median prevalence to lowest.

	Prevalence 2021	Prevalence % change
National	9.5% (8.9% to 10.0%)	-28.1% (-33.3% to -24.4%)
Northern		
Chitipa	4.2% (2.8% to 5.6%)	-21.6% (-37.0% to -1.9%)
Karonga	8.8% (7.6% to 10.7%)	-24.4% (-33.4% to -13.3%)
Likoma	9.2% (8.0% to 11.5%)	-13.0% (-27.7% to 3.7%)
Mzimba	6.7% (5.9% to 7.4%)	-24.4% (-31.8% to -17.2%)
Nkhata Bay	7.7% (6.7% to 9.0%)	-32.8% (-41.7% to -24.3%)
Rumphi	8.0% (6.4% to 10.0%)	-32.3% (-43.6% to -20.6%)
Central		
Dedza	4.4% (3.4% to 5.4%)	-36.1% (-47.6% to -25.8%)
Dowa	3.3% (2.6% to 4.2%)	-36.2% (-47.4% to -24.2%)
Kasungu	4.6% (3.7% to 5.6%)	-28.9% (-38.5% to -18.1%)
Lilongwe	7.8% (7.0% to 8.8%)	-32.0% (-39.7% to -24.4%)
Mchinji	5.8% (5.0% to 6.9%)	-34.5% (-45.4% to -26.4%)
Nkhotakota	6.0% (4.9% to 7.1%)	-40.4% (-49.5% to -32.9%)
Ntcheu	8.1% (6.8% to 9.7%)	-32.5% (-44.9% to -21.9%)
Ntchisi	2.8% (2.3% to 3.6%)	-38.5% (-49.2% to -27.2%)
Salima	6.5% (5.3% to 7.7%)	-28.1% (-38.1% to -17.8%)
Southern		
Balaka	9.6% (8.2% to 11.0%)	-31.4% (-40.9% to -23.8%)
Blantyre	16.7% (15.1% to 18.7%)	-26.8% (-34.7% to -18.1%)
Chikwawa	10.2% (8.3% to 11.7%)	-25.4% (-35.3% to -17.2%)
Chiradzulu	12.9% (10.6% to 15.3%)	-28.6% (-39.6% to -18.7%)
Machinga	8.4% (7.1% to 10.6%)	-32.9% (-41.7% to -22.2%)
Mangochi	9.3% (8.2% to 10.9%)	-32.7% (-40.5% to -24.9%)
Mulanje	18.3% (16.6% to 20.3%)	-20.5% (-27.4% to -14.0%)
Mwanza	8.6% (7.0% to 10.6%)	-25.0% (-35.4% to -12.5%)
Neno	10.6% (8.2% to 12.8%)	-39.6% (-49.3% to -31.0%)
Nsanje	14.6% (12.6% to 17.0%)	-19.4% (-28.4% to -7.6%)
Phalombe	15.6% (13.7% to 17.8%)	-23.7% (-31.2% to -15.6%)
Thyolo	13.6% (11.8% to 15.4%)	-30.7% (-38.9% to -23.1%)
Zomba	14.9% (12.8% to 17.1%)	-29.1% (-38.1% to -20.1%)

Table 5.1: Estimated prevalence among women in Malawi in 2021 and percent change between 2010 and 2021 with 95% credible intervals.

	Prevalence 2021	Prevalence % change
National	5.1% (4.8% to 5.6%)	-42.5% (-46.6% to -37.1%)
Northern		
Chitipa	2.2% (1.3% to 3.1%)	-42.0% (-57.9% to -17.8%)
Karonga	4.8% (4.1% to 5.8%)	-41.9% (-48.6% to -30.8%)
Likoma	5.0% (4.3% to 6.2%)	-33.2% (-44.9% to -14.1%)
Mzimba	3.6% (3.1% to 4.3%)	-40.1% (-46.9% to -28.4%)
Nkhata Bay	4.2% (3.6% to 5.1%)	-48.3% (-55.7% to -36.2%)
Rumphi	4.4% (3.6% to 5.5%)	-47.6% (-54.1% to -34.2%)
Central		
Dedza	2.4% (1.9% to 2.9%)	-51.3% (-59.0% to -43.2%)
Dowa	1.8% (1.5% to 2.2%)	-49.7% (-57.1% to -41.5%)
Kasungu	2.5% (2.0% to 3.1%)	-41.6% (-49.1% to -30.3%)
Lilongwe	4.1% (3.5% to 4.7%)	-45.2% (-51.0% to -37.5%)
Mchinji	3.2% (2.7% to 3.9%)	-46.6% (-54.2% to -37.8%)
Nkhotakota	3.3% (2.7% to 3.9%)	-51.9% (-58.3% to -44.9%)
Ntcheu	4.3% (3.5% to 5.2%)	-49.4% (-59.0% to -40.9%)
Ntchisi	1.5% (1.2% to 1.9%)	-51.7% (-60.1% to -41.8%)
Salima	3.5% (2.7% to 4.3%)	-44.1% (-53.4% to -30.7%)
Southern		
Balaka	5.2% (4.4% to 6.1%)	-47.3% (-54.3% to -38.0%)
Blantyre	8.9% (7.9% to 10.2%)	-40.2% (-46.7% to -32.4%)
Chikwawa	5.5% (4.5% to 6.5%)	-39.8% (-47.9% to -29.4%)
Chiradzulu	6.9% (5.7% to 8.3%)	-46.9% (-54.3% to -37.6%)
Machinga	4.6% (3.9% to 5.8%)	-48.7% (-55.2% to -39.4%)
Mangochi	5.1% (4.4% to 6.2%)	-48.8% (-55.8% to -39.1%)
Mulanje	10.5% (9.2% to 12.2%)	-38.1% (-44.4% to -27.9%)
Mwanza	4.6% (3.6% to 5.7%)	-42.3% (-50.3% to -28.6%)
Neno	6.1% (4.8% to 7.3%)	-52.1% (-58.1% to -45.7%)
Nsanje	7.9% (6.5% to 9.9%)	-36.8% (-47.2% to -21.6%)
Phalombe	9.0% (7.6% to 10.8%)	-39.5% (-46.4% to -29.0%)
Thyolo	7.7% (6.6% to 9.0%)	-44.6% (-50.5% to -37.0%)
Zomba	8.1% (7.0% to 9.5%)	-44.7% (-51.1% to -36.0%)

Table 5.2: Estimated prevalence among men in Malawi in 2021 and percent change between 2010 and 2021 with 95% credible intervals.

	ART coverage 2021	ART coverage % change
National	92.5% (89.3% to 94.8%)	266.3% (249.0% to 284.8%)
Northern		
Chitipa	94.8% (77.6% to 97.6%)	599.3% (276.2% to 1345.6%)
Karonga	93.7% (79.3% to 97.7%)	263.4% (215.4% to 333.0%)
Likoma	96.4% (84.7% to 97.8%)	209.5% (170.6% to 268.8%)
Mzimba	96.9% (90.8% to 97.8%)	242.6% (198.3% to 341.4%)
Nkhata Bay	95.6% (86.2% to 97.8%)	386.7% (250.5% to 582.6%)
Rumphi	86.9% (66.1% to 97.5%)	147.1% (95.1% to 188.6%)
Central		
Dedza	93.8% (73.1% to 98.1%)	268.7% (203.8% to 366.7%)
Dowa	87.0% (65.3% to 97.5%)	140.9% (92.2% to 176.0%)
Kasungu	94.8% (81.3% to 97.8%)	313.6% (230.3% to 434.1%)
Lilongwe	93.9% (84.8% to 97.8%)	223.4% (190.6% to 261.8%)
Mchinji	96.8% (88.2% to 98.0%)	321.8% (217.1% to 485.3%)
Nkhotakota	96.3% (86.4% to 98.1%)	217.8% (173.3% to 280.1%)
Ntcheu	94.4% (80.7% to 97.9%)	458.4% (273.6% to 999.8%)
Ntchisi	91.0% (72.2% to 97.8%)	310.3% (225.9% to 466.2%)
Salima	95.9% (85.2% to 97.8%)	465.6% (343.8% to 667.5%)
Southern		
Balaka	97.1% (89.3% to 98.0%)	246.4% (197.7% to 345.2%)
Blantyre	93.9% (85.3% to 97.7%)	243.4% (209.5% to 280.1%)
Chikwawa	96.7% (89.6% to 97.9%)	298.6% (237.6% to 393.6%)
Chiradzulu	90.1% (72.8% to 97.4%)	131.1% (80.9% to 185.2%)
Machinga	91.9% (73.7% to 97.9%)	437.5% (336.9% to 563.5%)
Mangochi	96.3% (84.5% to 98.0%)	663.3% (465.8% to 1016.7%)
Mulanje	96.9% (91.0% to 97.9%)	467.0% (397.5% to 574.5%)
Mwanza	94.8% (78.6% to 97.9%)	217.5% (160.4% to 318.0%)
Neno	96.4% (83.1% to 98.2%)	235.6% (188.6% to 291.9%)
Nsanje	95.2% (83.5% to 97.7%)	313.4% (238.3% to 521.6%)
Phalombe	96.2% (88.2% to 97.7%)	960.8% (758.2% to 1479.5%)
Thyolo	95.6% (85.5% to 97.9%)	324.7% (253.8% to 446.4%)
Zomba	86.2% (71.2% to 96.1%)	318.4% (247.3% to 384.3%)

Table 5.3: Estimated ART coverage among women in Malawi in 2021 and percent change between 2010 and 2021 with 95% credible intervals.

	ART coverage 2021	ART coverage % change
National	90.8% (87.8% to 93.3%)	277.6% (246.9% to 323.8%)
Northern		
Chitipa	93.1% (74.3% to 97.6%)	725.3% (292.4% to 1980.3%)
Karonga	92.6% (76.8% to 97.7%)	253.7% (196.7% to 336.6%)
Likoma	95.8% (82.2% to 97.9%)	200.2% (161.4% to 312.3%)
Mzimba	96.3% (88.1% to 97.9%)	257.6% (191.9% to 391.5%)
Nkhata Bay	94.9% (84.8% to 97.8%)	440.3% (251.5% to 765.0%)
Rumphi	87.3% (64.7% to 97.2%)	139.0% (88.9% to 185.2%)
Central		
Dedza	93.0% (73.1% to 98.0%)	278.1% (195.4% to 425.7%)
Dowa	86.7% (65.3% to 97.0%)	130.8% (81.2% to 183.1%)
Kasungu	93.8% (79.1% to 97.9%)	334.5% (224.8% to 527.2%)
Lilongwe	91.0% (81.2% to 97.0%)	261.7% (206.0% to 354.1%)
Mchinji	96.6% (87.1% to 98.0%)	326.5% (196.0% to 552.8%)
Nkhotakota	95.7% (85.9% to 98.0%)	221.5% (165.1% to 311.4%)
Ntcheu	93.0% (78.7% to 97.8%)	535.2% (263.7% to 1445.9%)
Ntchisi	90.2% (70.4% to 97.6%)	350.4% (228.0% to 611.5%)
Salima	94.7% (82.3% to 97.8%)	494.2% (329.7% to 804.6%)
Southern		
Balaka	96.0% (85.7% to 97.9%)	262.5% (191.0% to 434.1%)
Blantyre	90.8% (81.9% to 96.8%)	282.8% (221.8% to 367.5%)
Chikwawa	96.1% (87.4% to 97.9%)	286.9% (215.8% to 427.5%)
Chiradzulu	87.6% (69.0% to 96.8%)	125.8% (70.7% to 199.5%)
Machinga	90.3% (70.5% to 97.8%)	509.6% (360.8% to 762.0%)
Mangochi	95.3% (81.7% to 97.9%)	787.5% (493.1% to 1391.3%)
Mulanje	96.6% (89.6% to 97.9%)	498.4% (377.5% to 659.3%)
Mwanza	93.1% (74.4% to 97.9%)	231.3% (156.9% to 396.8%)
Neno	96.2% (83.6% to 98.2%)	231.9% (178.4% to 308.3%)
Nsanje	93.6% (78.1% to 97.7%)	326.8% (228.1% to 648.9%)
Phalombe	96.0% (87.3% to 97.9%)	1000.7% (646.8% to 1492.8%)
Thyolo	94.8% (84.8% to 97.8%)	320.4% (226.6% to 467.6%)
Zomba	84.5% (68.5% to 94.7%)	356.0% (267.6% to 480.9%)

Table 5.4: Estimated ART coverage among men in Malawi in 2021 and percent change between 2010 and 2021 with 95% credible intervals.

	Incidence rate 2021	Incidence rate % change
National	2.1 (1.5 to 2.7)	-74.0% (-79.9% to -68.9%)
Northern		
Chitipa	1.2 (0.5 to 2.5)	-75.4% (-85.8% to -52.2%)
Karonga	2.5 (1.2 to 4.5)	-72.3% (-84.5% to -55.3%)
Likoma	3.0 (1.6 to 5.5)	-59.8% (-79.2% to -24.2%)
Mzimba	1.9 (1.2 to 2.6)	-72.1% (-81.6% to -59.3%)
Nkhata Bay	1.8 (1.0 to 2.9)	-78.0% (-87.8% to -63.8%)
Rumphi	2.0 (0.7 to 4.6)	-65.6% (-84.6% to -31.8%)
Central		
Dedza	0.7 (0.2 to 1.7)	-80.6% (-93.4% to -64.1%)
Dowa	0.6 (0.2 to 1.6)	-72.7% (-91.0% to -47.8%)
Kasungu	1.2 (0.6 to 2.1)	-73.9% (-86.3% to -58.6%)
Lilongwe	1.6 (0.8 to 2.7)	-76.5% (-87.8% to -64.0%)
Mchinji	1.2 (0.5 to 2.0)	-79.9% (-90.7% to -69.4%)
Nkhotakota	1.0 (0.3 to 1.8)	-79.7% (-91.6% to -68.3%)
Ntcheu	1.4 (0.5 to 2.8)	-83.8% (-94.0% to -70.9%)
Ntchisi	0.5 (0.2 to 1.2)	-79.9% (-92.0% to -65.3%)
Salima	1.6 (0.8 to 2.6)	-78.5% (-88.4% to -63.8%)
Southern		
Balaka	2.1 (0.9 to 3.3)	-78.2% (-89.3% to -66.9%)
Blantyre	3.8 (1.9 to 6.5)	-76.4% (-87.6% to -62.8%)
Chikwawa	2.8 (1.3 to 4.1)	-73.0% (-86.5% to -61.3%)
Chiradzulu	3.1 (1.2 to 5.5)	-73.0% (-87.9% to -55.1%)
Machinga	1.9 (0.8 to 4.2)	-80.5% (-91.1% to -65.8%)
Mangochi	2.0 (1.1 to 3.4)	-83.1% (-90.4% to -73.9%)
Mulanje	5.5 (3.3 to 8.0)	-76.7% (-85.5% to -67.3%)
Mwanza	2.2 (1.1 to 4.3)	-71.1% (-83.8% to -47.4%)
Neno	2.0 (0.4 to 3.9)	-78.4% (-92.7% to -65.0%)
Nsanje	5.1 (3.1 to 8.7)	-70.9% (-81.7% to -51.5%)
Phalombe	5.0 (2.9 to 7.4)	-79.3% (-87.3% to -69.9%)
Thyolo	3.1 (1.4 to 5.0)	-80.9% (-90.6% to -71.8%)
Zomba	4.1 (1.8 to 7.2)	-76.0% (-88.4% to -62.3%)

Table 5.5: Estimated incidence rate among women in Malawi in 2021 and percent change between 2010 and 2021 with 95% credible intervals. Note national level incidence is measured per person, not per person-year.

	Incidence rate 2021	Incidence rate % change
National	1.0 (0.7 to 1.4)	-80.2% (-84.3% to -74.0%)
Northern		
Chitipa	0.5 (0.2 to 1.4)	-81.9% (-89.8% to -56.9%)
Karonga	1.0 (0.6 to 2.2)	-80.8% (-87.8% to -63.0%)
Likoma	1.0 (0.5 to 2.5)	-77.0% (-86.1% to -47.4%)
Mzimba	0.7 (0.5 to 1.5)	-80.7% (-87.0% to -63.4%)
Nkhata Bay	0.7 (0.4 to 1.7)	-84.2% (-90.5% to -67.2%)
Rumphi	0.8 (0.3 to 2.2)	-77.5% (-88.3% to -42.2%)
Central		
Dedza	0.3 (0.1 to 0.7)	-85.7% (-93.3% to -73.5%)
Dowa	0.3 (0.1 to 0.6)	-81.1% (-91.6% to -64.4%)
Kasungu	0.5 (0.3 to 1.0)	-82.1% (-88.8% to -66.3%)
Lilongwe	0.6 (0.4 to 1.2)	-82.2% (-89.1% to -70.2%)
Mchinji	0.5 (0.3 to 1.0)	-85.9% (-92.0% to -74.0%)
Nkhotakota	0.4 (0.2 to 0.8)	-85.2% (-92.0% to -75.6%)
Ntcheu	0.6 (0.3 to 1.1)	-88.1% (-94.2% to -80.0%)
Ntchisi	0.2 (0.1 to 0.5)	-85.4% (-92.7% to -71.8%)
Salima	0.8 (0.4 to 1.6)	-82.2% (-89.1% to -64.1%)
Southern		
Balaka	1.0 (0.5 to 1.9)	-82.8% (-89.9% to -70.0%)
Blantyre	1.5 (0.9 to 2.6)	-82.1% (-88.7% to -71.3%)
Chikwawa	1.1 (0.7 to 2.1)	-80.9% (-88.3% to -68.3%)
Chiradzulu	1.4 (0.7 to 2.6)	-80.8% (-89.2% to -66.5%)
Machinga	0.9 (0.5 to 2.1)	-85.0% (-91.5% to -71.1%)
Mangochi	1.0 (0.6 to 2.0)	-86.5% (-91.4% to -75.2%)
Mulanje	2.5 (1.6 to 4.5)	-83.1% (-88.4% to -71.0%)
Mwanza	0.9 (0.5 to 2.2)	-79.5% (-87.4% to -55.9%)
Neno	0.7 (0.2 to 1.6)	-86.2% (-93.6% to -75.4%)
Nsanje	2.4 (1.3 to 4.9)	-77.8% (-86.0% to -58.2%)
Phalombe	2.1 (1.3 to 4.0)	-86.1% (-90.4% to -76.0%)
Thyolo	1.5 (0.9 to 2.6)	-84.2% (-90.4% to -75.3%)
Zomba	1.8 (0.9 to 3.4)	-82.1% (-89.6% to -69.0%)

Table 5.6: Estimated incidence rate among men in Malawi in 2021 and percent change between 2010 and 2021 with 95% credible intervals. Note national level incidence is measured per person, not per person-year.

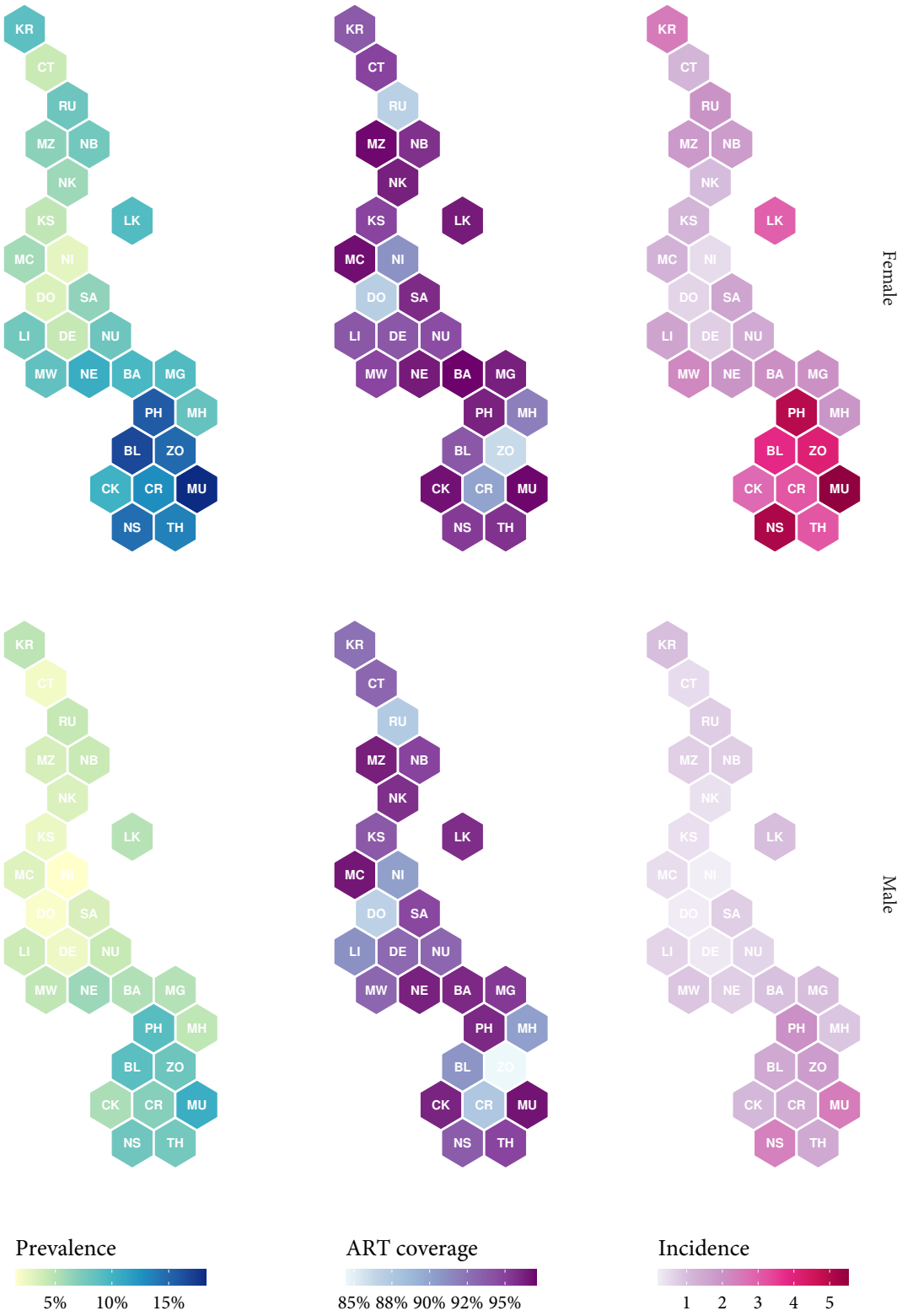

Figure 5.6: Posterior median HIV prevalence, ART coverage, and HIV incidence by sex in Malawi at the end of 2021.

heterogeneous. Across the 28 districts, the median decrease in prevalence was 33.6%, and the median decrease in incidence was 76.0%. The smallest decrease in estimated prevalence was in Likoma with a 21.4% (6.3% to 33.3%) decrease, while the largest was Neno with a 42.9% (36.4% to 50.8%) decrease. Estimated incidence decreased most slowly in Likoma by 65.5% (35.2% to 79.3%) between 2010 and 2021 and most rapidly in Ntcheu by 83.3% (72.0% to 92.8%).

These changes in incidence were matched by dramatic improvements in treatment coverage in every district. Between 2010 and 2021, across both men and women, every district of Malawi at least doubled ART coverage. The median increase in ART coverage was 261.9%. The smallest estimated change in ART coverage was an increase of 116.5% (66.9% to 169.7%) in Chiradzulu, and the largest was a 756.5% (635.6% to 893.7%) increase in Phalombe. Figure 5.8 presents a ladder plot of the distributions of percent changes in all 28 districts.

Chiradzulu and Phalombe are notable because they had the highest and lowest ART coverages, respectively, in 2010 for both men and women, suggesting that the largest improvements in that period were in the districts that were lowest to begin with. Figure 5.9 formalises this intuition by comparing ART coverage in 2010 to the percent change between 2010 and 2021. Low ART coverage in 2010 was strongly associated with large increases between 2010 and 2021. There is a natural reciprocal relationship in this plot because districts with higher 2010 coverage have lower maximum possible percent changes. The line on the figure represents the theoretical maximum increase in ART coverage we could possibly observe given the 2010 posterior median ($1/x - 1$).

Changes in prevalence and incidence were spatially heterogeneous, while changes in ART coverage were consistently high. Figure 5.10 plots the percent change in prevalence, ART coverage, and incidence by region between 2010 and 2021. Prevalence and incidence both decreased most dramatically in the Central region. The median percent change in prevalence across districts within the Central Region was -37.6%, compared to -29.6% and -33.1% in the Northern and Southern Regions, respectively. Similarly the median percent changes in incidence were -78.3%, -72.5%, and -76.2% in the Central, Northern, and Southern Regions, respectively. ART coverage changed more uniformly across region, at 257.7%, 226.3%, and 283.4% in the Central, Northern, and Southern Regions.

In Section 1.1, I stated that the model be sufficiently precise if, in every study region, it

Figure 5.7: Posterior median percent change in HIV prevalence, ART coverage, and HIV incidence by sex in Malawi between the beginning of 2010 and the end of 2021.

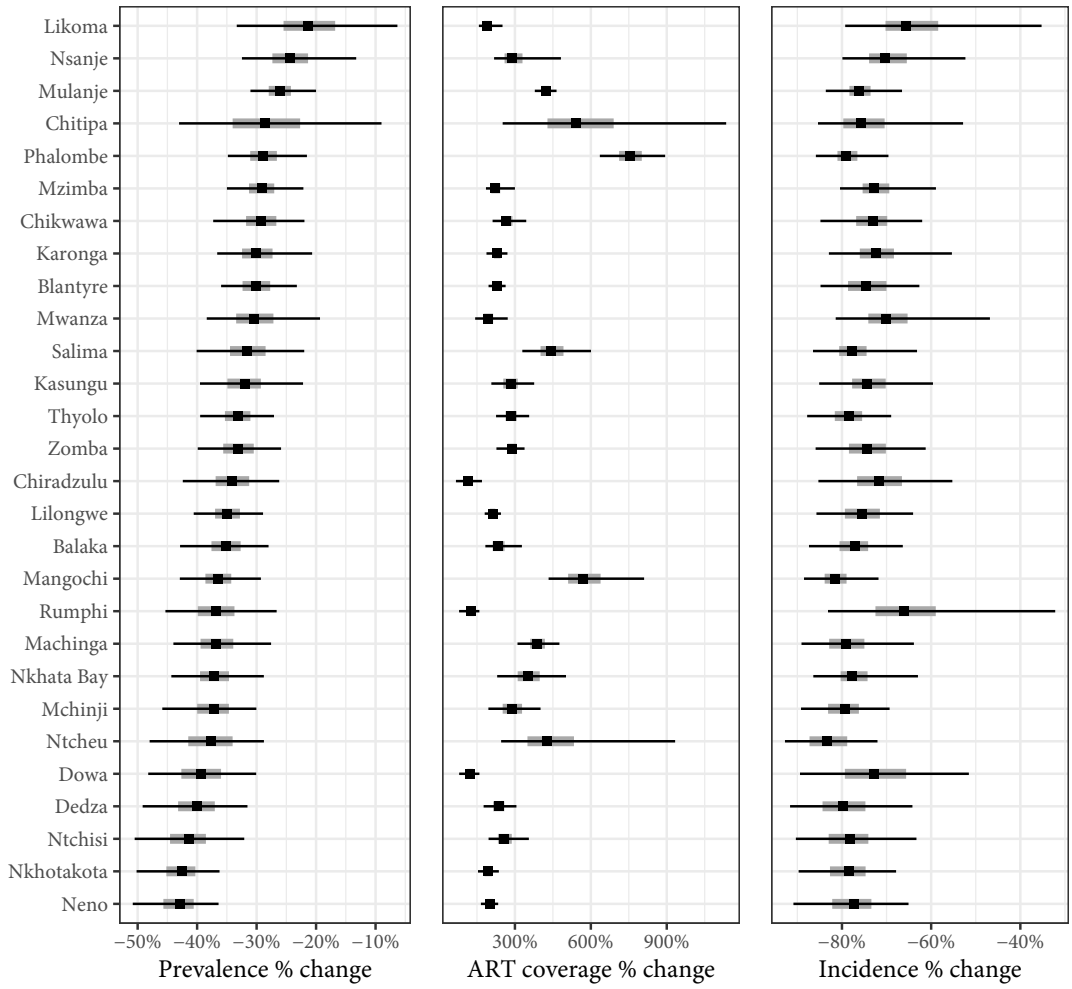

Figure 5.8: Ladder plots of posterior median (points) percent changes in incidence risk, ART coverage, and prevalence with 95% and 50% credible intervals (lines and shaded regions, respectively). Districts are sorted vertically from highest median percent change in prevalence to lowest.

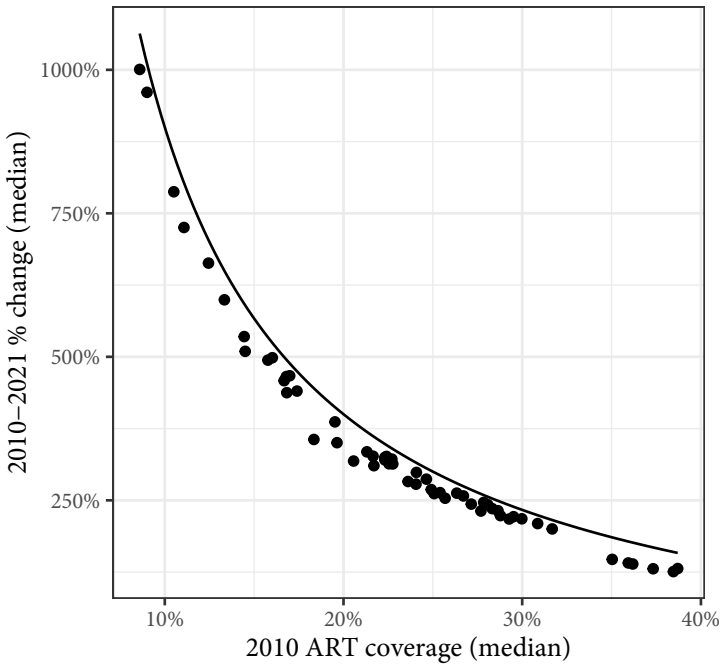

Figure 5.9: Posterior median ART coverage in 2010 compared to the percent change in ART coverage between 2010 and 2021 by sex and district. The black line represents the percent difference between 100% and the 2010 value (that is, the maximum possible change).

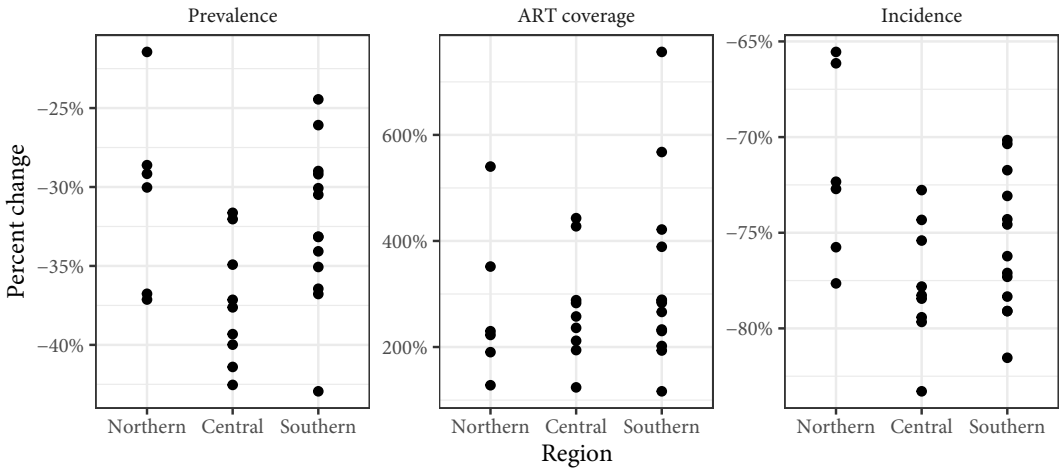

Figure 5.10: Posterior median percent change in district-level prevalence, ART coverage, and incidence grouped by region.

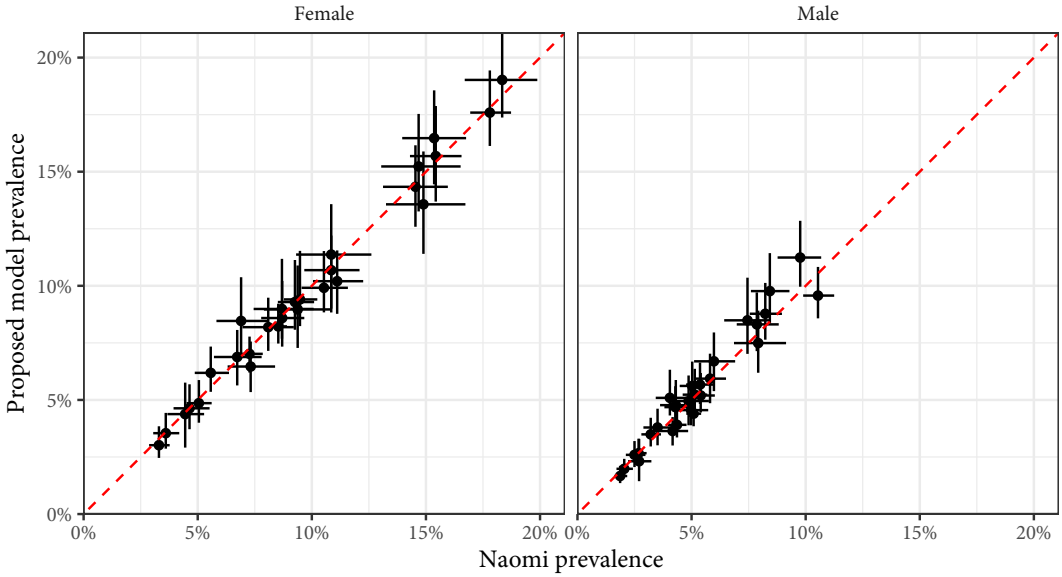

Figure 5.11: Comparison of district-level prevalence inferred by the proposed model to Naomi among adults aged 15-49 in 2020 by sex. Horizontal and vertical line ranges correspond to 95% CIs for Naomi and the proposed model, respectively.

estimated that the posterior probability of incidence having decreased by 50% or more between 2010 and 2021 was less than 20% or greater than 80%. The model met that goal in this analysis. The posterior probability of a 50% or greater decrease in incidence was greater than 80% in every district of Malawi. In fact, this posterior probability was only less than 95% in two districts: Rumphi and Likoma.

Comparison to Naomi estimates

Figures 5.11, 5.12, and 5.13 compare estimated district-level prevalence, ART coverage, and incidence, respectively, from Naomi and the proposed model. The two sets of estimates correspond well across district and sex for prevalence and incidence, particularly when considering uncertainty, and vary more for ART coverage. Figure 5.12 highlights that my model estimated higher ART coverage in both men and women, with a clearer difference among men. It also estimated more slightly more extreme incidence than Naomi with lower estimates in lower districts and higher estimates in higher districts. Pooling both sexes together, the correlation coefficients between the point estimates were 0.99 for prevalence, 0.38 for ART coverage, and 0.93 for incidence.

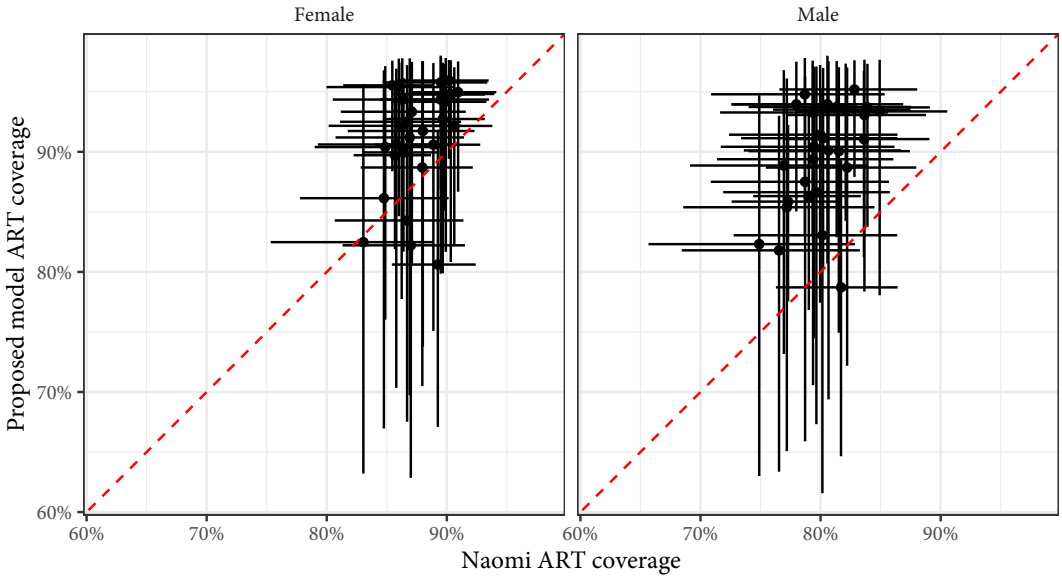

Figure 5.12: Comparison of district-level ART coverage inferred by the proposed model to Naomi among adults aged 15-49 in 2020 by sex. Horizontal and vertical line ranges correspond to 95% CIs for Naomi and the proposed model, respectively.

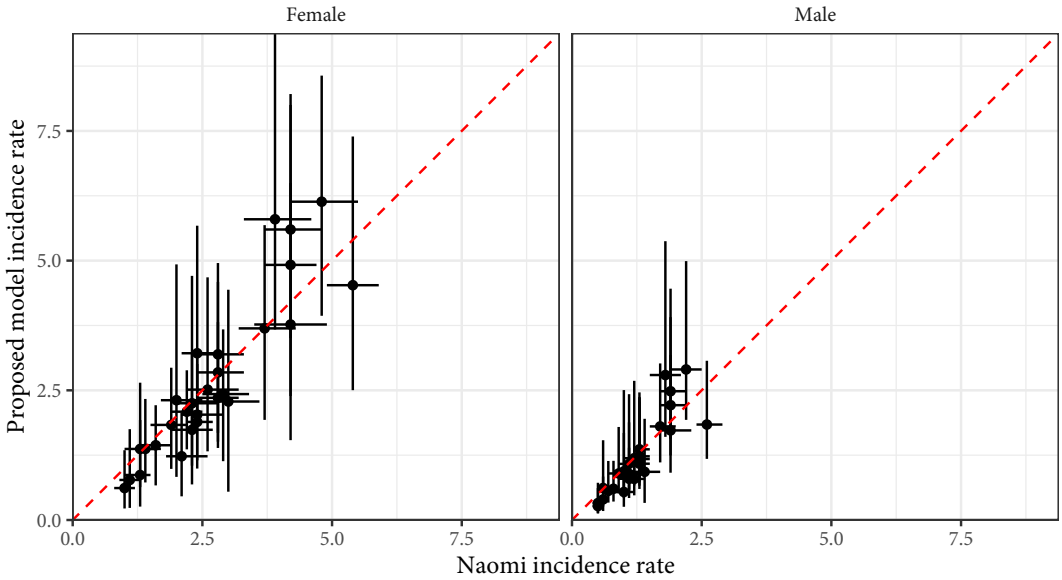

Figure 5.13: Comparison of district-level incidence inferred by the proposed model to Naomi among adults aged 15-49 in 2020 by sex. Horizontal and vertical line ranges correspond to 95% CIs for Naomi and the proposed model, respectively.

5.3 Discussion

In this analysis, I fit the model described in Chapter 2 to data from Malawi collected between 1995 and 2020, described trends across space, time, and sex in HIV prevalence, incidence, and ART coverage, and compared the estimates to existing national-level estimates. The model fit well to the available data, but there were substantial differences between the model's estimates and UNAIDS' estimates at the peak of the epidemic. The two sources agreed in recent years, when data were plentiful. The model estimated with a high degree of certainty that, over the past decade, the rapid scale-up of treatment programmes in Malawi has resulted in marked increases in ART coverage and commensurate decreases in incidence that were generally uniform across space.

These estimates highlight the continued success of Malawi's HIV treatment programme. Estimated national ART coverage increased by 271% percent between 2010 and 2021, and changes were greatest in the most under-served areas. Figure 5.9 shows that districts with the largest increases between 2010 and 2021 were those with the lowest coverage in 2010. In fact, this analysis suggests that the increases in every district between 2010 and 2021 were only slightly below the rates needed to reach 100% coverage by 2021. Far from identifying spatial gaps in changes in treatment, the model estimated that initial gaps have been accounted for over the past ten years. Note that I have used all PLHIV as the denominator when calculating ART coverage. Some of the lower treatment coverage in earlier periods reflect restricted treatment eligibility, but measuring ART coverage in terms of all PLHIV is more consistent with current recommendations.

Despite beginning projections much later than EPP, my model succeeded in synthesising HIV seroprevalence data from household surveys and ANC facilities in a similar way to EPP. It effectively applied the temporal trends from the ANC data to the absolute levels from the surveys. In my model, however, these prevalence data sources are complemented by data measuring ART coverage, ART patients counts, and recency testing.

Notably, my model was able to replicate district-level ART patient count series almost perfectly. Although EPP does reproduce input ART patient counts exactly, it does so by forcing the compartmental model to match the counts at every point in time. This strategy creates stubborn numerical problems and requires the user to have in their possession

a complete and perfectly correct time series of ART patient counts. The strategy I have proposed here does not require complete time series, allows for the possibility of noise, and avoids inducing the numerical problems. This strategy also allows the user to handle outlying or otherwise erroneous data points by simply omitting them.

The near-perfect fit to ART patient count data was not accompanied by similar fit to ART testing data. The model seemed to smooth ART coverage across sex, underestimating in women and overestimating in men despite inferring that women were approximately twice as likely to initiate treatment given CD4 stage. Without sex-specific patient counts, it is difficult to identify the source of this problem, but it could be partially ameliorated with a more flexible model for the sex ratio of ART initiation.

Comparing this model's national-level estimates to existing estimates, I observed discrepancies earlier in the epidemic when data were sparse. My model estimated a later and lower peak in incidence than UNAIDS, although the two sets of results converged in recent years. It is possible that this is the result of the model taking a "minimum intervention" approach to modelling the transmission rate of HIV. Because data from the 1990s and early 2000s are so sparse, the model has limited evidence supporting deviation from the relatively low transmission rates needed to fit to recent data. In other words, the model can fit well to the earliest data with relatively modern transmission rates, which are more precisely identified.

The interpretation of these results is limited by a number of important weaknesses. There were large differences between my model and external estimates in earlier years. These differences could be attributable to several factors. The two sets of estimates converge as more data become available, suggesting that these differences are might be attributable to differing start dates and transmission rate parametrisations. By beginning projections in the 1970s, UNAIDS is able to incorporate the earliest available ANC data, which encourages their models to exhibit exponential growth in the 1970s and 1980s and gradual declines in prevalence from the 1990s and onwards.

Second, this model does not use every possible data source. I have aggregated data from hundreds of ANC facilities into just a handful of district-level series and, in doing so, might have smoothed over important variation. Fitting to each facility separately has an intuitive appeal, but it would increase the complexity of evaluating of the ANC likelihood by a factor of several hundred. The beta-binomial ANC likelihood should partially alleviate these

problems. ANC facilities also provide ART testing data (Eaton et al., 2021), which I have not incorporated here.

Third, the model smoothed ART coverage across sex, estimating that men were as likely as women to be on treatment. These estimates conflicted with UNAIDS, which estimated that women were 7% more likely to be receiving ART than men. This discrepancy could be due to the relative rigidity the sex ratio of treatment initiation in the ART initiation model, which is log-linear with respect to time, or to sex differences in the assumptions about mortality without treatment, which contribute substantially to treatment initiation. In either case, allowing more flexibility in the model of ART initiation could resolve these issues.

Finally, I have simplified the spatial dynamics of HIV in two ways that could affect the results. First, as discussed in Chapter 4, I have assumed that there is no cross-district transmission, which is unlikely to be true. However, recent one recent study in sub-Saharan Africa found that only about 13% of new infections with a set of communities originated from outside of each community (Hall et al., 2021). Additionally, the cross-validation exercise in Chapter 4 indicated that omitting spatial transmission was a better choice than including it.

The second simplification is that I have defined the ART attendance model using first-degree adjacency; PLHIV can only seek treatment in the districts that are adjacent to their home regions. This approach fails to capture the complex reality of ART attendance and might lend too much importance to the frequently arbitrary administrative boundaries. Instead, we could define any number of alternative distance metrics or use a gravity model that allows people to seek treatment in any region (Meredith et al., 2021).

Despite these limitations, the results presented in this chapter offer two key insights. First, by modelling ART initiation and building a generative model for ART patient counts, we can incorporate programmatic treatment data into the model in a natural way without sacrificing fit. Finally, consistent with the national-level estimates from UNAIDS, I found that ART coverage has increased dramatically in every district of Malawi, resulting in proportionate decreases in incidence.

The spatial heterogeneity in these estimates emphasises the need for spatially resolved

models of incidence. Even in a generalised epidemic, I observed a 13-fold difference in general-population incidence between the highest and lowest districts. The view of HIV in Malawi as a single, homogeneous epidemic papers over spatial variation that is becoming increasingly important to policy-makers. I observed substantial variation in incidence over space that would has been masked by national-level models. Effective HIV policy-making depends increasingly on measurement to local trends, and the models used to estimate HIV burden must respond to those needs.

Chapter 6

Nowcasting HIV Incidence with Incremental Data

In this chapter, I consider model fitting processes that are specifically designed to incorporate routinely reported data at greater temporal granularity. I estimate the effects of incrementally including quarterly ANC and ART facility data and measure the ability of the model to forecast the epidemic in Malawi.

6.1 Introduction

Although nationally representative household surveys have been integral to historical estimates of HIV incidence, routine reporting systems offer broader reach and far denser temporal resolution. A small subset of routinely reported data are incorporated into current HIV estimation processes, but the vast majority of are ignored.

Programmatic measurements of HIV indicators are increasingly incorporated into queryable databases, some automatically via electronic medical record systems and some via manual input. I have taken advantage of these systems in a minor way by fitting to district-level aggregates of routine testing data from all ANC facilities in Malawi, but inferential models of HIV could be integrated into digital data warehouses far more deeply. In the simplest case, a model of HIV incidence would run once every quarter, automatically pulling the latest programmatic data from an authoritative source.

There are certain practical questions to answer before implementing such a system. First, we would necessarily produce a new set of historical estimates every time we re-fit the model. Ideally, our estimate of HIV incidence for a fixed time point would converge to a single, well-identified value as more data were added, but this has not been tested. Large changes in estimates for a fixed date from one quarter to the next would be difficult to justify and explain.

Second, because this work includes a generative model for ART patient counts, it lends itself naturally to forecasting. The only input that must be complete over the projection period is population counts. However, as the length of the forecast increases, the precision in estimated incidence will decrease. To aid future users in the design of estimation strategies, we must identify how far in advance (if at all) this model can reliably forecast incidence. The unifying question across the two points is: how much data after a fixed time point do we need before we can view our estimates of incidence as trustworthy?

In this analysis, I used data from Malawi to estimate the effects of incrementally including quarterly data from ANC and ART facilities on inferred incidence. By simulating data from a pre-fit model and fitting to the simulations, I was able ask a number of questions about how each additional data point affected estimated HIV incidence. First, I measured how including new data affected uncertainty in past estimates of HIV incidence, hypothesising that additional data would noticeably impact estimates even at the beginning of the projection period. Second, I estimated how many years of data were required after a fixed date to recover true incidence on that date at a given level of precision. Finally, I assessed how accurately the model was able to forecast incidence under the correct specification. I did not have specific hypotheses for the second and third questions.

6.2 Methods

I simulated quarterly data from 2021 through the end of 2024 using the model fit in Chapter 5 to produce a dataset of ART patient counts and ANC prevalence with known incidence. Chapter 2 details the methodology for that model. I combined the existing data in Malawi from 1995 through 2020 with the simulated data from 2021 through 2024 to construct a partially synthetic dataset. I re-fit the model to the combined dataset, holding out data in

quarterly increments from the beginning of 2021 to the end of 2024. I compared estimated incidence to true incidence to assess how well each forecasting horizon recovered true incidence. To illustrate how future data impacted current estimates, I compared inferred incidence at fixed time points across the 16 forecasting horizons.

I measured the precision of estimated incidence as the absolute width of the 95% posterior credible interval. To measure relative precision, I found the ratio of the size of the credible interval for each forecasting horizon relative to the credible interval for the shortest forecasting horizon (i.e. the model with the least data held out).

I estimated the amount of data necessary after a fixed date to estimate incidence with a given level of precision on that date. That is, how many years of data collection do we need after time t to cover true incidence at time t α% of the time? To answer this question, I found the 95% posterior coverage of true incidence across district, sex, and forecasting horizon by the amount of time between each estimate date and the final in-sample data point. I selected 95% as the target level of precision only because it is widely used as a standard level of uncertainty. Other targets could be selected in future work.

To assess the model's ability to recover true incidence given each forecasting horizon, I found the share of district-sex-time combinations in which true incidence was within the 95% credible interval of predicted incidence. I stratified this measurement of 95% posterior coverage of true incidence by forecasting horizon to try to identify an ideal horizon that balanced accuracy and precision.

6.3 Results

Two forecasting horizons (1 April 2023 and 1 April 2024) failed to converge. Figure 6.1 presents box plots of 95% CI ratio relative to the shortest forecasting horizon pooled across quarter, district, and sex for six estimate years. Each panel groups together estimated incidence for a given year. Higher x values indicate earlier forecasting horizons (more data included), and higher values y values indicate wider credible intervals relative to the model that included the most data. As the amount of data included in the model increased, the credible intervals of estimated incidence contracted, an effect that was more pronounced in estimates for more recent years. Assuming the earliest forecasting horizon (1 January

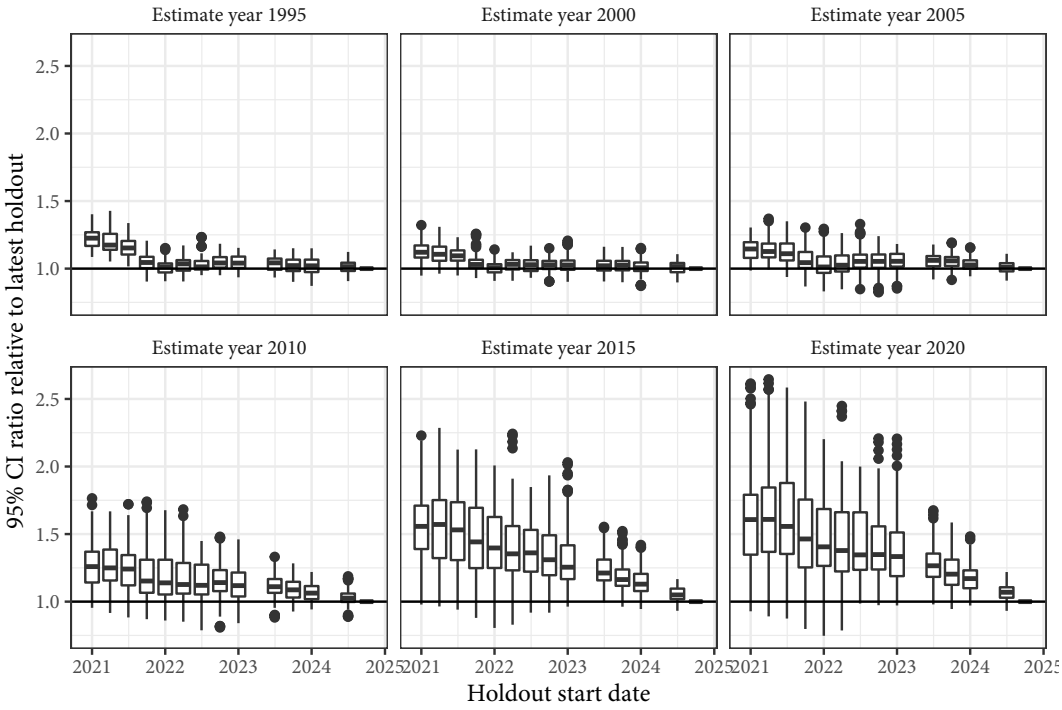

Figure 6.1: Box plots of the ratios of the 95% credible intervals on posterior incidence by holdout start date compared to the intervals for the latest start date by estimate year.

2021), the mean 95% credible interval ratio increased from 1.2 for incidence in 1995 to 1.6 for incidence in 2020.

This trend is particularly clear in the Karonga District, presented in Figure 6.2. For the same estimate dates, models with less data predicted incidence less precisely. Among both women and men, the precision of estimated incidence increased as more data were included, although the effects are lessened in earlier estimate years. For example, among women in 2020, the size of the 95% CI of incidence ranged from 5.2 when holding out the most data to 2.2 when holding out the least.

Figure 6.3 present the data from Figure 6.1 as a heat map, with estimate year increasing along the x-axis and holdout start dates increasing along the y-axis. Darker green values indicate greater uncertainty relative to the shortest forecasting horizon. In general, the CI ratio decreases as the holdout start date increases with each column, showing that, as more simulated data were included, the precision of estimates at fixed dates increased. For example, the mean 95% credible interval ratio for incidence at the end of 2020 decreased from 1.6 when only one quarter of simulated data was included to 1.1 when all but two

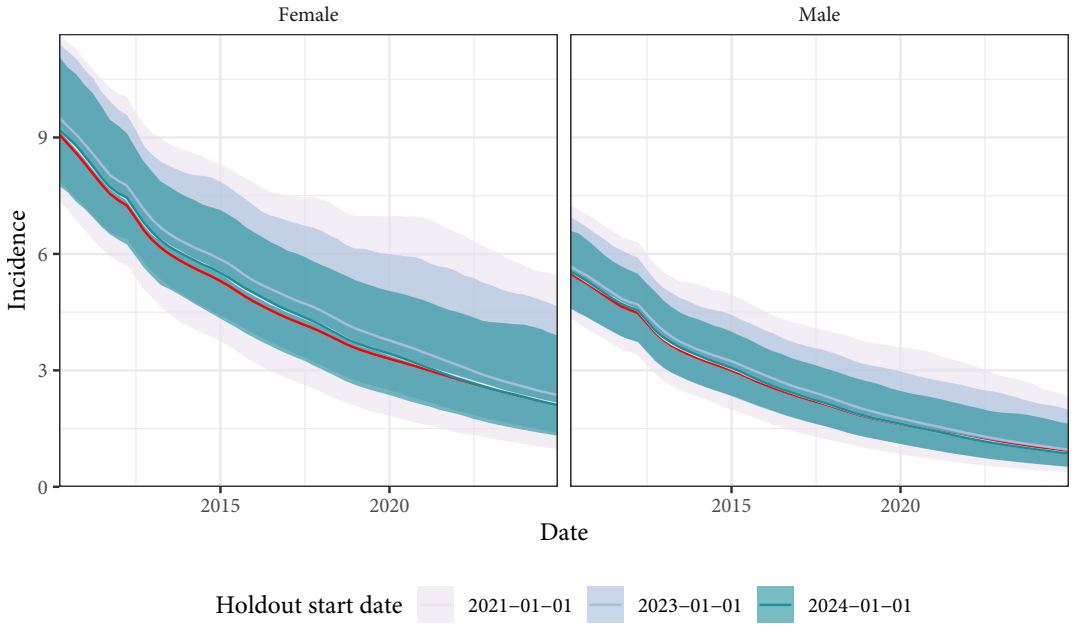

Figure 6.2: Estimated HIV incidence (medians and 95% CIs) in Karonga from 2010 through 2021 by sex and selected holdout start date. The red line is median incidence from the model fit from which the data in this chapter were simulated. Paler blue lines and regions included less data than darker blue lines and regions.

quarters of simulated data were included.

Because the simulated data were generated using a known epidemic, we can find the share of observations (district-/quarter-/sex-combinations) in which true incidence fell within the estimated posterior 95% credible interval for each forecasting horizon. I estimated how much data were necessary after a fixed point in time to reach 95% posterior coverage of incidence. Figure 6.4 plots 95% posterior coverage of true incidence at fixed dates stratified by the difference between each date and the date of the last data point. Negative values of x were aggregated over estimate dates that were later than the last held out data point (forecasts), while positive values were aggregated over estimates dates that were before the last data point. Posterior coverage was highest when aggregating over forecasted dates and decreased steadily as more data were added. Posterior coverage was closest to the target of 95% when the model had six quarters of data after each date, suggesting that the desired level of precision for a given date was achieved 1.5 years after that date. Although Figure 6.4 suggests that incidence was estimated most reliably for time periods before the final data point, I still assessed the ability of the model to forecast incidence. Figure 6.5 presents posterior coverage of true incidence by holdout start date for time periods

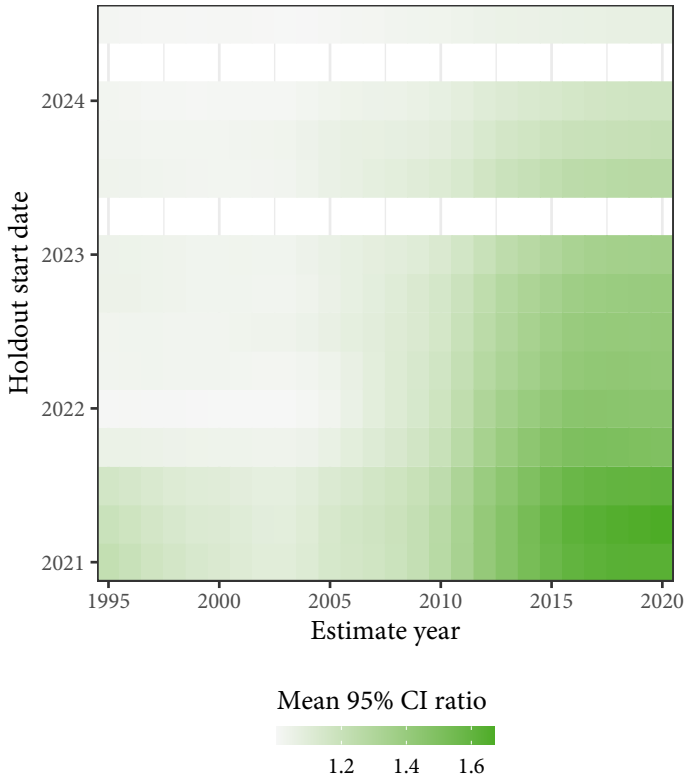

Figure 6.3: Mean 95% CI ratio for each holdout start date relative to the latest holdout start date by estimate date. Higher values indicate that adding new data has a greater effect on precision.

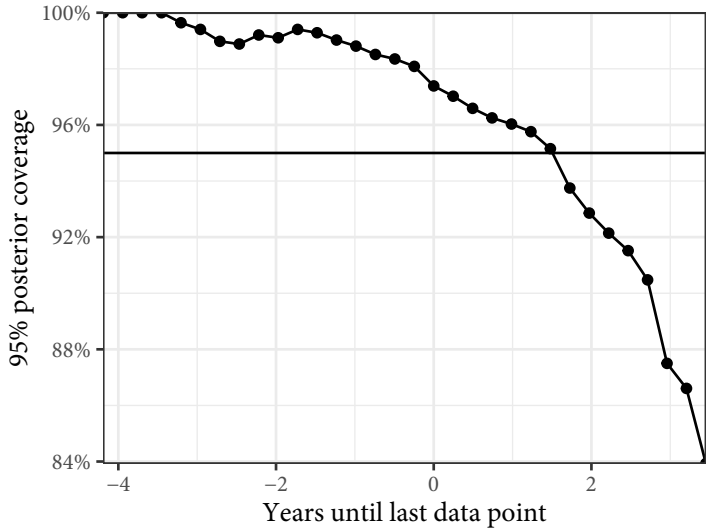

Figure 6.4: 95% posterior coverage of true incidence by distance between estimate date and the final quarter of data for estimate dates later than 2020. Negative x values are aggregated over estimate dates that are beyond the last data point (i.e. forecasted).

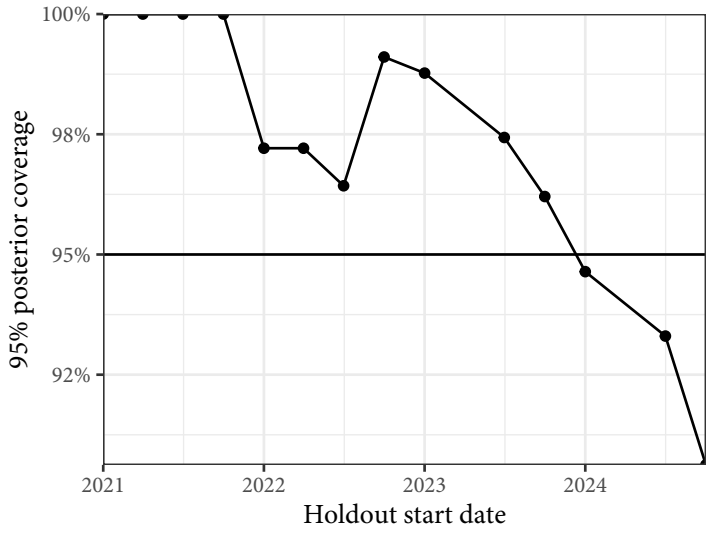

Figure 6.5: 95% posterior coverage of true incidence by holdout start date for the time periods fit to simulated data

with simulated data. Coverage decreased as more data were included, crossing the target coverage of 95% when holdout began at the beginning of 2024. Coverage was 96.2% when forecasting began in the fourth quarter of 2023 and decreased to 94.6% when forecasting began in the first quarter of 2024. Notably, 95% posterior coverage of incidence crosses the target threshold after more than 50% of the simulated data have been included. Because this analysis continued to add data that the original simulation did not have, the model was able to estimate past incidence more precisely than the original fit.

Figure 6.6 presents 95% posterior coverage of incidence disaggregated by estimate quarter. As in Figure 6.3, higher values on the y-axis indicate less data held out and higher values on the x-axis indicate later estimate quarters. Cells below the red dashed line correspond to forecasted estimates. The model achieved target coverage above the line of equality, meaning that no forecasted estimates were properly calibrated.

6.4 Discussion

In this analysis, I tested the ability of the model described in Chapter 2 to provide quarterly estimates of HIV incidence by successively including data simulated from a previously fit model. Each additional quarter of data increased the precision of estimates across the whole projection period, although the effects were larger for more recent periods. The

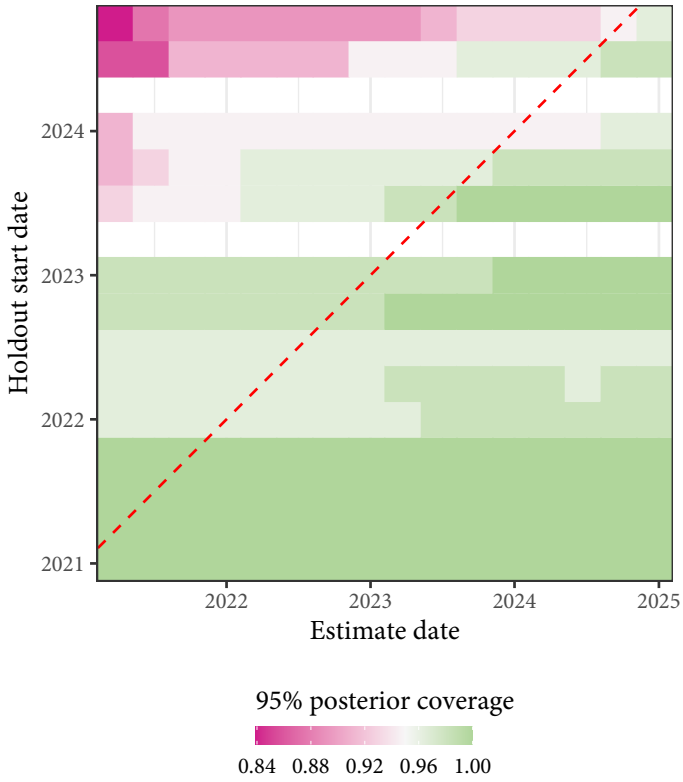

Figure 6.6: 95% posterior coverage of true incidence by holdout start date and estimate year for the time periods fit to simulated data. Cells below red line of equality correspond forecasted incidence estimates (estimate dates greater than holdout start dates).

model was able to infer incidence at 95% level of precision or better at a given date when it was presented with between one and two years of data collected after that date.

Figure 6.1 shows that sequentially incorporating quarterly programmatic data into this model improved the precision of recent estimates and had a noticeable effect on earlier estimates. An estimation process that automatically adds new data on a regular basis or as it becomes available would fit neatly into increasingly widespread digital reporting systems. This analysis shows that such a process could have benefits beyond limiting the need for human intervention.

Under the correct model specification, the model was able to estimate incidence from simulated data in Malawi after approximately 1.5 years of data collection. Prevalence and ART patient counts fundamentally reflect past incidence, so the ability to reliably estimate incidence at a given date within two years of that date indicates good performance. However, the model failed to achieve 95% posterior coverage of incidence for all forecasted time points. Even under a correctly specified model in a highly regular epidemic, incidence was too uncertain to forecast reliably.

This study is subject to important weaknesses that limit its generalisability. First, by appending simulated data to real data, as opposed to using a fully synthetic data set, I have limited our ability to interpret incidence as "true" incidence. Because I re-fit the model over the entire period, estimated incidence series between 1995 and 2020 are not necessarily comparable between this analysis and the analysis from Chapter 5. I appended the simulated data to real data from Malawi to investigate the effects of incrementally including quarterly data in Malawi as directly as possible, but using purely synthetic data would allow us to study how the model performs in forecasting in greater generality. I plan to conduct a larger simulation study investigating the ability of models like this one to detect changes in true incidence in the future.

Second, the scope of both the simulated and real data I have used here is limited. Data in Malawi are highly regular, making them a relatively easy case for forecasting. Future work will test these methods on fully simulated data designed to exhibit greater spatiotemporal heterogeneity. Further, to limit computation time, I have only used one set of simulations sampled using the mode parameter set from Chapter 5. Working with a larger set of simulated data would offer more robust results.

Estimating how much data are needed after a given date to reliably estimate incidence at that date is a central question in HIV incidence modelling. Loosely speaking, we must be able to assess whether a given incidence is trustworthy. This study found that, given data from Malawi, my model was able to infer incidence accurately after about 1.5 years. Future work will examine the performance of this model using a greater number of fully synthetic datasets, but the results presented here hint at the potential impact of deeper integration between HIV incidence models and digital data collection mechanisms.

Chapter 7

Modelling Self-Reported Sexual Partner Age Distributions

Although I have omitted age structure from the compartmental model described in Chapter 2, explicitly modelling age dynamics of HIV transmission could greatly increase the accuracy of my predictions. To model the rate of sexual HIV transmission in one age group attributable to another age group, I would need to make an assumption (or set of assumptions) about the rate of sexual partnership between all pairs of age groups. Data on sexual partnership formation are available from a variety of sources, but they must be smoothed and interpolated before being integrated into an epidemic model. In this chapter, I describe a distributional regression strategy for modelling self-reported sexual partner age distributions. The estimates from such a model could be integrated into an age-specific epidemic model. Alternatively, because both models are Bayesian, we could estimate the parameters for both models simultaneously. This more complex use case would allow us to incorporate the uncertainty from the partner age distribution model seamlessly into the epidemic model.

This chapter has been reproduced from an article originally published in *eLife Sciences* on 24 June 2021 (Wolock et al., 2021). The authors of the original article were myself, Dr Seth Flaxman, Dr Kathryn Risher, Tawanda Dadirai, Prof Simon Gregson, and Dr Jeffrey Eaton. I designed and performed the experiments described in this chapter and wrote the article published in *eLife*. The Appendix associated with that article has been reproduced in Appendix B, and references to sections of that appendix have been updated appropriately

in the text.

7.1 Introduction

Patterns in sexual mixing across ages determine patterns of transmission of sexually transmitted infections (STIs). Consequently, sexual age-mixing has been of great interest to researchers studying the human immunodeficiency virus (HIV) since the beginning of the global epidemic. Anderson et al. (1992) used a model of partnership formation to predict that mixing between young women and older men would amplify the already-substantial effect on HIV on population growth. Garnett & Anderson (1994) used a mathematical model to show that patterns of age-mixing could substantially influence the magnitude and timing of hypothetical epidemic trajectories, while Hallett et al. (2007) demonstrated that delaying sexual debut and increasing age-similar partnerships could reduce an individual's risk of HIV infection in a highly endemic setting.

These modelling studies have been complemented by analyses of survey and population cohort data on age-mixing patterns. Gregson et al. (2002) observed that individuals with older partners were at greater risk of HIV infection. Ritchwood et al. (2016) and Maughan-Brown, Evans, & George (2016) found that larger age differences were associated with more risky sexual behaviour in surveys of young South African people. On the other hand, Harling et al. (2014) found that age-disparate relationships were not associated greater risk of HIV acquisition in young women in South Africa.

These results underscore the importance of considering age-mixing dynamics when designing and evaluating HIV prevention strategies, and, consequently, the importance of measuring them accurately. For example, an intervention aiming to prevent new HIV infections among young women could be undermined by high prevalence among older men. Identifying changes in sexual partner age distributions and attributing them to interventions might even be a valuable end by itself, in which case accurate measurement must be complemented by an effective modelling strategy.

Data about sexual partner age-mixing are routinely collected by long-term cohort studies (such as those that comprise the ALPHA Network) and large-scale household surveys (such as the Demographic and Health Surveys) (Reniers et al., 2016; "The DHS program," 2021).

Typically, these data consist of the respondent's age and sex and the ages of their sexual partners in the last 12 months. These data are highly variable, skewed, and often deviate substantially from conventional parametric distributions, such as the normal distribution or the gamma distribution (Beauclair, Hens, & Delva, 2018).

One may consider statistical modelling approaches for the distribution of partner age as a function of respondent age and sex. Some notable previous approaches to modelling partner age distributions include Hallett et al., who used a log-logistic distribution to model partner age differences for women aged 15 to 45 years, assuming that the partner age difference distributions did not vary over respondent age. More recently, as an input to a model of *Chlamydia trachomatis*, Smid et al. (2018) fit skew normal distributions to each age-/sex-specific partner age distribution and used a secondary regression model to smooth the estimated skew normal parameters across respondent age. They observed substantial changes in the estimated skew normal parameters with respect to respondent age. Although this method allows for non-linear variation across respondent age, their two-stage estimation process makes uncertainty propagation complex. Replacing this process with a single "distributional" regression model, in which all distributional parameters (e.g. the location, scale, skewness, etc.) are modelled as functions of data (Kneib & Umlauf, 2017), allows for complex variation across respondent age while still robustly incorporating uncertainty. Another elegant approach has been the development of exponential-family random graph models (ERGMs) to infer full partnership networks from indivuals reports of the partnerships ('ego-centric' observations of the network) (Krivitsky & Morris, 2017). These stochastic methods, along with the broader suite of ERGMs (Hunter et al., 2008; Hunter, Goodreau, & Handcock, 2008; Krivitsky & Handcock, 2014; Krivitsky, Handcock, & Morris, 2011), can model social network data accurately with robust incorporation of covariates, and tools exist to incorporate their estimates into epidemic models (Jenness, Goodreau, & Morris, 2018; Morris, 1993).

More broadly, no previous work has systematically evaluated the wide variety of distributions potentially available to model partner age distributions. These distributions are skewed, heavy-tailed, and otherwise dissimilar to conventional statistical distributions due to personal preferences, social dynamics, demographic change, and any number of other factors. We were specifically interested in distributions that introduce parameters to

control tail weight, which may capture intergenerational mixing that could sustain endemic HIV and STI transmission (Akullian et al., 2017; Harling et al., 2014; Schaefer et al., 2017). This led us to test the ability of the four-parameter "sinh-arcsinh" distribution originally proposed by Jones & Pewsey (2009) to fit to these data.

We hypothesized that integrating the sinh-arcsinh distribution into a distributional modelling framework would allow us to replicate observed partner age distributions more accurately than prior modelling strategies. We tested this theory by comparing a variety candidate strategies, which varied along three dimensions: the parametrisation of the dependent variable, the choice of distribution, and the method for incorporating variability across respondent age and sex.

7.2 Methods

We conducted two model comparison experiments to identify which of a set of strategies best replicated partner age distributions. First, in our probability distribution comparison, we identified which of a set of distribution-dependent variable combinations fit best to age-/sex-specific data subsets, and then, in our distributional regression evaluation, we tested whether distributional regression methods could be used to estimate age-/sex-specific partner age distributions by sharing strength across observations. We divided the model comparison into two separate experiments to make the probability distribution comparison as fair as possible (accounting for the possibility that certain distributions would perform particularly well under certain regression specification).

7.2.1 Data

We analysed data on sexual partner age distributions from three sources: the Africa Centre Demographic Information System, a health and demographic surveillance site in uMkhanyakude district, South Africa collected by the African Health Research Institute (AHRI) (Gareta et al., 2021; Gareta, Dube, & Herbst, 2020a, 2020b), the Manicaland General Population Cohort in Zimbabwe (Gregson et al., 2017), and the 2016-2017 Demographic and Health Survey (DHS) in Haiti (Institut Haïtien de l'Enfance - IHE/Haiti & ICF, 2018).

The AHRI and Manicaland studies are multi-round open, general population cohort studies designed to measure the dynamics of HIV, sexual risk behaviour, and demographic change in sub-Saharan African settings. We used rounds 1 through 6 of the Manicaland study, collected between 1998 and 2013. The AHRI data we used were collected annually between 2004 and 2018. The 2016-17 Haiti DHS was a large, nationally representative household survey conducted in 2016 and 2017. We did not incorporate the weights associated with the survey into this analysis because our primary interest was in statistical modelling of partner age distribution as a function of respondent age, not producing population representative statistics for the Haiti population.

These data sets consisted of individuals' reports of their own age and sex and the ages of each of their sexual partners from the last year. Let $i \in (1, ..., N)$ index reported partnerships, $a_i \in [15, 64]$ and $s_i \in \{0, 1\}$ be the age and sex of the respondent in partnership i with $s = 1$ indicating female, and p_i be the age of non-respondent partner in partnership i. These questionnaires do not ask specifically about partner sex, but self-reporting of non-heterosexual partnerships in these populations is thought to be low (Arias Garcia et al., 2020; World Health Organization & UNAIDS, 2020).

Respondents in each of these data sets are disproportionately likely to report that their partners' ages are multiples of five or multiples of five away from their own age, leading to distinct "heaping" in the empirical partner age (or age difference) distributions at multiples of five. We tested the sensitivity of our results to heaping by developing a simple "deheaping" algorithm, applying it to the AHRI data, and running each analysis on the deheaped AHRI data. We present these results in Section B.1.

7.2.2 Probability distribution comparison

To identify the best probability distribution for modelling sexual partner age distributions, we split each data set into 12 subsets by sex and five-year age bin ranging from 20 to 50, resulting in 36 subsets, and fit a number of distribution-dependent variable combinations to each subset.

Distribution	Parameters	Domain	PDF
Normal	μ (location) $\sigma > 0$ (scale)	$\mathbb{R}$	$\frac{1}{\sigma\sqrt{2\pi}} \exp\left[\frac{-x_z}{2}\right]$
Skew normal	μ (location) $\sigma > 0$ (scale) ϵ (skewness)	$\mathbb{R}$	$\frac{2}{\sigma} p(x_z)\Phi(\epsilon x_z)$
Gamma	$k > 0$ (shape) $\theta > 0$ (scale)	$\mathbb{R}^+$	$\frac{1}{\Gamma(k)\theta^k} x^{k-1} \exp\left[\frac{-x}{\theta}\right]$
Beta	$\alpha > 0$ (left) $\beta > 0$ (right)	$\mathbb{R}^{(0,1)}$	$\frac{x^{\alpha-1}(1-x)^{\beta-1}}{B(\alpha,\beta)}$
Sinh-arcinh	μ (location) $\sigma > 0$ (scale) ϵ (skewness) $\delta > 0$ (tail weight)	$\mathbb{R}$	$\frac{1}{\sigma\sqrt{2\pi}} \frac{\delta C_{\epsilon,\delta}(x_z)}{\sqrt{1+x_z^2}} \exp\left[-\frac{S_{\epsilon,\delta}(x_z)^2}{2}\right]$

Table 7.1: Details of the five distributions tested in this analysis. We define $x_z = (x - \mu)/\sigma$, $p(x)$ to be the standard normal PDF, $\Phi(x)$ to be the standard normal cumulative density function, $S_{\epsilon,\delta}(x) = \sinh(\epsilon + \delta \operatorname{asinh}(x))$, and $C_{\epsilon,\delta} = \cosh(\epsilon + \delta \operatorname{asinh}(x))$.

Distributions

We tested five candidate probability distributions: normal, skew normal, beta, gamma, and sinh-arcsinh. Table 7.1 summarises the domains, parameters, and probability density functions (PDFs) of these distributions. Because the gamma distribution is always right-skewed and men typically partner with women who are younger than them, we transformed data among male respondents to be right-skewed when using the gamma distribution. Specifically, we multiplied the men's partners' ages by -1 to reflect the distribution horizontally across the y-axis, and added 150 to the reflected ages to ensure that all resulting values were positive. Similarly, the beta distribution is only defined on the interval $(0, 1)$, so, only when using a beta distribution, we scaled all partner ages to be between zero and one using upper and lower bounds of 0 and 150. The sinh-arcsinh distribution, presented by Jones & Pewsey (2009), is an extension of Johnson's S_U distribution (Johnson, 1949). It has four parameters: location, scale, skewness, and tail weight (denoted, μ, σ, ϵ, and δ respectively), and it can deviate substantially from the normal distribution. Figure 7.1 plots the density of this distribution with $\mu = 0$ and $\sigma = 1$ for a variety of values of skewness and tail weight.

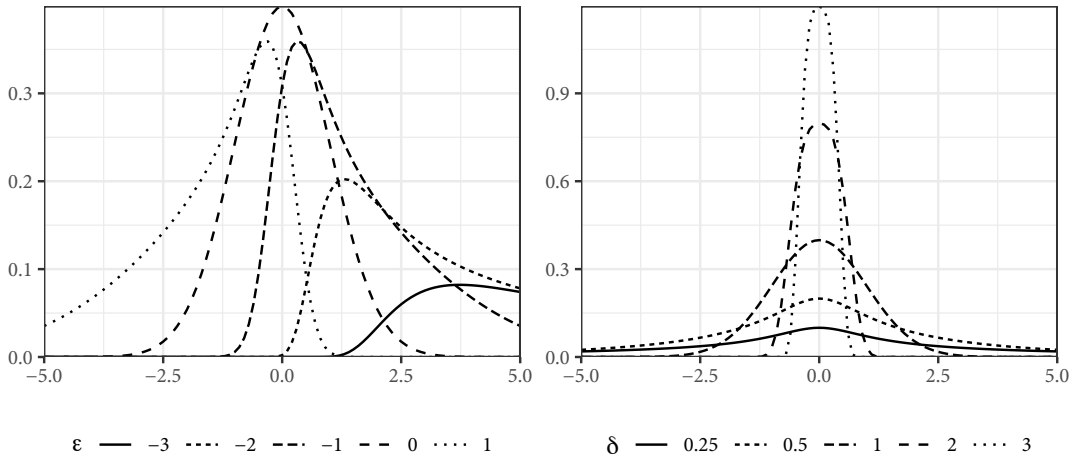

Figure 7.1: The sinh-arcsinh density with $\mu = 0$, $\sigma = 1$, and a variety of assumptions about ϵ and δ

Dependent variable transformations

We considered the possibility that certain distributions could interact well with particular transformations of the dependent variable (partner age) by testing a set of four potential outcome parametrisations. For example, if X is a positive-valued, right-skewed random variable, then assuming $\log X$ is normally distributed might be more effective than assuming that X itself is normal.

Let y_i be the dependent variable value for partnership i, and let a_i and p_i be the respondent age and partner age of partnership i, respectively. We tested the following dependent variables:

1. **Linear age**: $y_i = p_i$. This is untransformed partner age, included as a baseline. It has the undesirable quality of being able to predict negative ages.
2. **Age difference**: $y_i = p_i - a_i$. If changes in expected partner age are consistent across respondent age then this variable would be more consistent across respondent age than the linear age. This parametrisation also allows for negative partner age predictions.
3. **Log-age**: $y_i = \log p_i$. We can use a log link function to ensure that our predictions will be positive-valued.
4. **Log-ratio**: $y_i = \log(p_i/a_i)$. Finally, we can combine the link function and differencing approaches by modelling the log of the ratio of partner to respondent age. This variable will only produce positive predictions and should exhibit the same relative

stability as the age difference variable.

Because the gamma and beta distributions are not defined on the entire real line, we only fit them with the linear age dependent variable with the previously discussed transformations.

To identify which distribution-dependent variable combination best modelled the characteristics of sexual partner age distributions, we stratified each of our three data sets by sex and five-year age bin from 20-24 through 45-49. We fit every viable distribution-dependent variable combination to all 36 subsets independently. Given that we fit only the linear age dependent variable to the gamma and beta distributions, comprising a total of 504 models (14 per data set). We fit each model using the `brms` R package (Bürkner, 2018), defining custom families as necessary.

7.2.3 Distributional regression evaluation

Given a probability distribution that accurately replicated the non-Gaussian characteristics of partner age distributions, we tested whether or not distributional regression would allow us to pool data across age and sex without sacrificing fit. This is a method in which we make all of our distributional parameters, not just the mean, functions of data (Kneib & Umlauf, 2017). Taking conventional Bayesian regression as an example, we have

$$\begin{aligned} y_i &\sim \mathrm{N}(\mu_i, \sigma) \\ \mu_i &= \beta \mathbf{X}_i, \end{aligned} \tag{7.1}$$

where β and $\log \sigma$ are free parameters. There is an explicit assumption in this model that the standard deviation of the generating distribution is constant across all observations. We can use distributional regression to relax this assumption, making σ a function of data:

$$\begin{aligned} y_i &\sim \mathrm{N}(\mu_i, \sigma_i) \\ \mu_i &= \beta^{\mu} \mathbf{X}_i^{\mu} \\ \log \sigma_i &= \beta^{\sigma} \mathbf{X}_i^{\sigma}, \end{aligned} \tag{7.2}$$

where β^{μ} and β^{σ} are now our free parameters. Note that we have not assumed that $\mathbf{X}^{\mu} = \mathbf{X}^{\sigma}$. If $\mathbf{X}^{\sigma}$ is a column of ones, this model is identical to the conventional case. This

Model	Distributional?	Location	Other parameters
Conventional	No	Age-sex interaction	Constant
Distributional 1	Yes	Age-sex interaction	Age and sex effects
Distributional 2	Yes	Age-sex interaction	Age-sex interaction
Distributional 3	Yes	Sex-specific splines	Age-sex interaction
Distributional 4	Yes	Sex-specific splines	Sex-specific splines

Table 7.2: Summary of five models fit in this analysis.

approach increases the complexity of the model and requires more data, but, based on previously described characteristics of how the distribution of partnership age distribution changes with age, even a simple model for our distributional parameters could yield large improvements.

In this case, we used a sinh-arcsinh distribution and specified a model for each of its four parameters. We fit a series of increasingly complex distributional regression specifications to the three data sets using `brms` (Bürkner, 2018), which has deep support for distributional regression.

1. **Conventional**: linear age-sex interaction for location and constants for all three higher-order parameters
2. **Distributional 1**: linear age-sex interaction for location and independent age and sex effects for all other parameters
3. **Distributional 2**: linear age-sex interactions for all four parameters
4. **Distributional 3**: sex-specific spline with respect to age for location and linear age-sex interactions for all other parameters
5. **Distributional 4**: sex-specific splines with respect to age for all four parameters

Table 7.2 describes all five models. By fitting a wide set of specifications, we hoped to assess whether the additional complexity incurred by distributional regression was valuable. More detailed descriptions of each model are available in Section B.2.

7.2.4 Model comparison

Across both analyses, we used two metrics to measure model fit. First, we calculated the expected log posterior density (ELPD), which estimates the density of the model at a new,

unobserved data point (Vehtari, Gelman, & Gabry, 2017). In cases where we wanted to compare across dependent variables, we multiplied the posterior densities of any variables resulting from non-linear transformations of observed partner ages by the Jacobians of the transformations. For example, if our observation model was defined on the log-age dependent variable $y_i = \log p_i$, we divided the posterior density by p_i. We used the `loo` R package (Vehtari et al., 2020) to calculate ELPD values.

To measure the ability of our models to replicate partner age distributions in an objective and interpretable way, we found the root mean squared error (RMSE) between the observed and posterior predictive quantiles. We calculated quantiles from 10 to 90 in increments of 10 by age bin and sex in the data and in the posterior predictions, and found the error in model prediction of each quantile. This measure tells how well our model predicts the entire distribution in the same units as our predictions. It is equivalent to finding the mean squared or median absolute distance from the line of equality in a quantile-quantile (QQ) plot.

7.2.5 Software

We conducted all of these analysis using the R programming language (R Core Team, 2020) and the `brms` library (Bürkner, 2018). We used the `loo` library to estimate all ELPDs (Vehtari et al., 2020), and produced all plots in this paper with the `ggplot2` library (Wickham, 2016). We provide code and data for an example case on Github.

7.3 Results

The AHRI data included 77,619 partnerships, Manicaland had 58,676, and the Haiti DHS had 12,447. As an illustrative example of the distribution of partner ages, Figure 7.2 presents histograms of reported partner ages among women aged 35-39 for each of our three data sets. Figure 7.3 shows the sex- and age bin-specific empirical moments for the three datasets. Mean partner age increased with respondent age consistently for both sexes across all three datasets: among women, mean partner age increased by 26.0, 22.7, and 23.7 years in the AHRI data, Haiti DHS data, and Manicaland data, respectively, between age bins 20-24

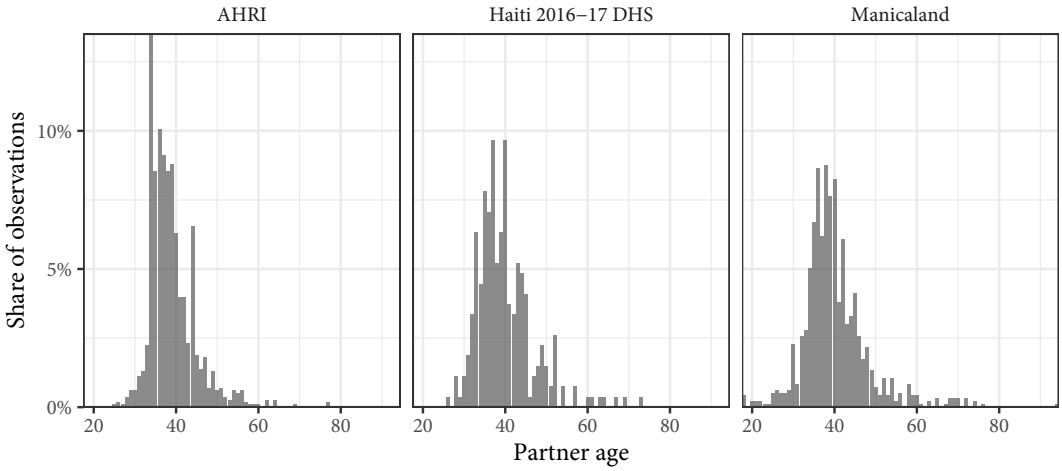

Figure 7.2: Observed partner age distributions among women aged 34 years in all three data sets.

and 45-49. However, higher order moments were less consistent: the standard deviation of women's partners' ages changed by 2.3, 0.5, and 3.5 years in the AHRI data, Haiti DHS data, and Manicaland data, respectively.

Within each dataset, there is systematic variation across sex. For example, the standard deviation of partner ages in the Haiti DHS increased by 2.5 years among men and only by 0.5 years among women. These summary statistics illustrate the heterogeneity of partner age distributions across age and sex.

7.3.1 Probability distribution comparison

To identify the probability distribution that most accurately described the variation in sexual partner age distributions, we identified the dependent variable with the highest ELPD for each distribution-dependent variable combination. Figure 7.4 illustrates each probability distribution's best fit to AHRI data among women aged 35-39 with each of the best distribution-specific dependent variables. Results for all 36 data subsets and the 12 deheaped subsets are in Section B.3. The best dependent variable varied across data subset and probability distribution. Table 7.3 provides the share of data sets for which each dependent variable has the highest ELPD given each distribution. The log-ratio dependent variable was best in 50.0% of subsets with a normal distribution, but it was best in only 27.8% of subsets with a skew normal distribution. The dependent variable that was best in a plurality of subsets in each probability distribution (i.e. the variable with the

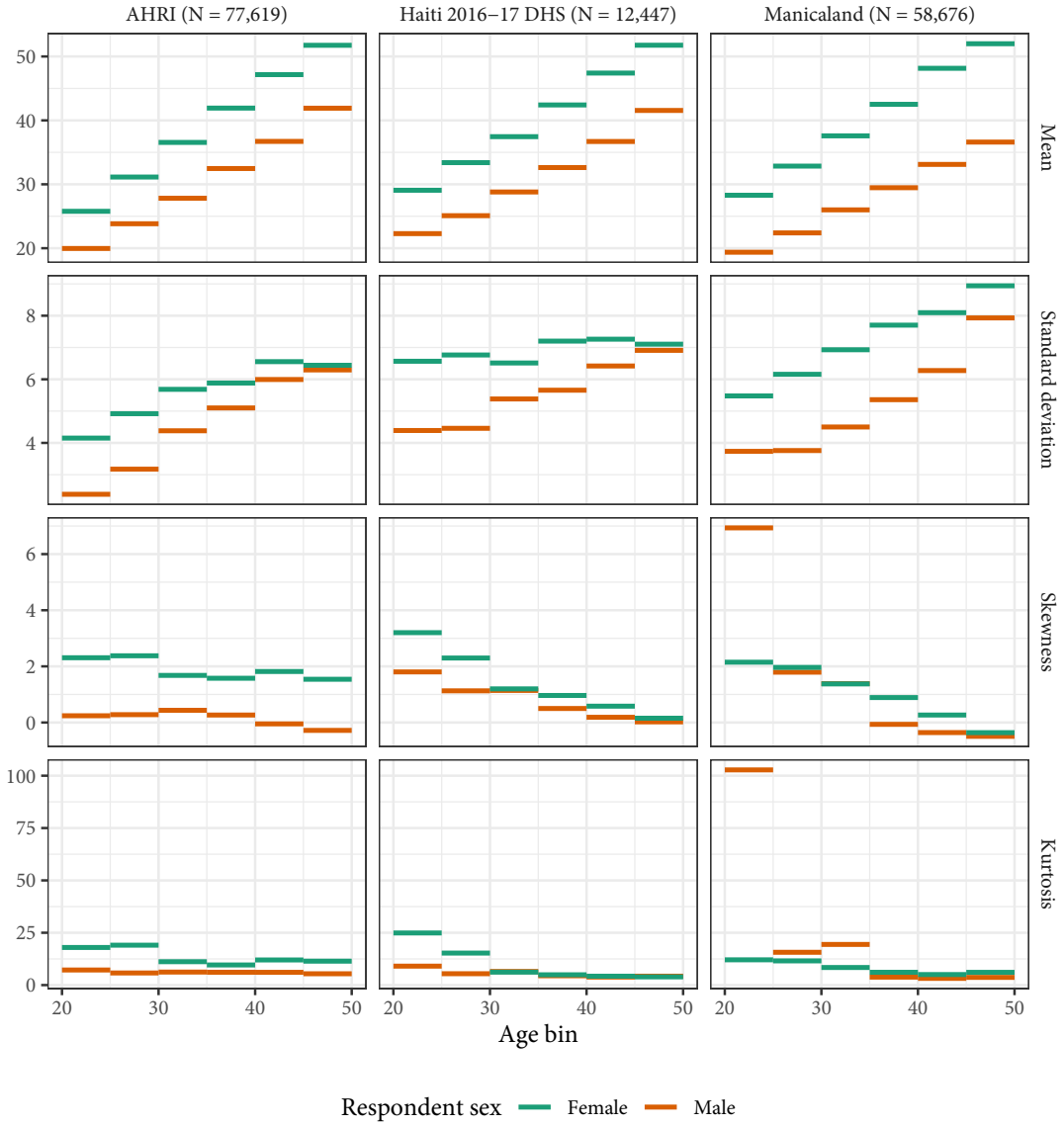

Figure 7.3: Observed means, variances, skewnesses, and kurtoses of partner age by five-year age bin and sex in all three datasets

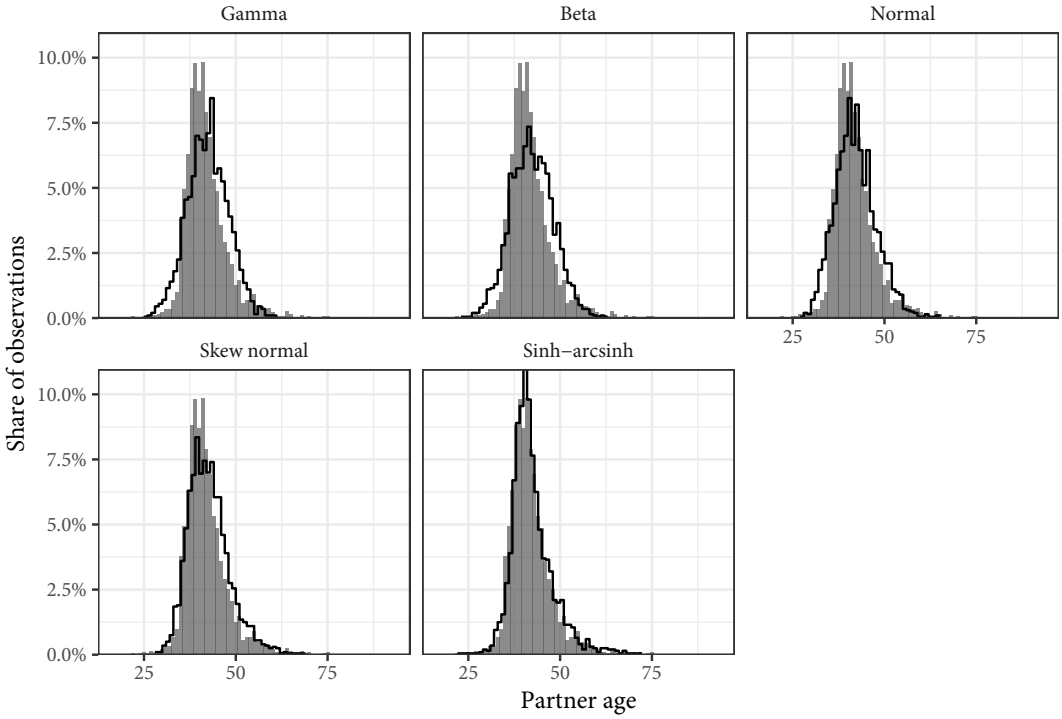

Figure 7.4: Observed partner age distributions (grey bars) and posterior predictive partner age distributions (lines) for each probability distribution among women aged 35-39 in the AHRI data set. Posterior predictive distributions come from fitting each age bin/sex combination independently.

Variable	Normal	Skew normal	Sinh-arcsinh
Age difference	22.2%	25.0%	16.7%
Linear age	8.3%	5.6%	16.7%
Log-age	19.4%	41.7%	30.6%
Log-ratio	50.0%	27.8%	36.1%

Table 7.3: Share of subsets in which each dependent variable yields the highest ELPD given each probability distribution (excluding deheaped AHRI data).

highest percentage in each column of Table 7.3) used a log link function. We restricted all remaining comparisons to each distribution-subset combination's best dependent variable. The sinh-arcsinh distribution had the highest ELPD in 35 of 36 data subsets (98%). In 29 of the 35 (83%) cases in which the sinh-arcsinh provided the highest ELPD, the absolute value of the ratio of the difference between the two best ELPDs and the estimated standard error of the difference was greater than 2, indicating that the sinh-arcsinh distribution was significantly better than the alternatives in the majority of cases. In one case, men aged 20-24 in the Haiti DHS, the skew normal distribution resulted in a slightly higher ELPD than the sinh-arcsinh distribution, but the standard error of the difference was greater than the difference. These results were not affected by deheaping the data first (Section B.1).

To summarise each distribution's performance, we calculated the average ELPD and QQ RMSE across the three datasets (Table 7.4). The sinh-arcsinh distribution had the highest average ELPD and lowest average QQ RMSE in all three datasets. The sinh-arcsinh distribution was, on average, able to predict the empirical quantiles of each data set within half a year of accuracy (0.36, 0.37, and 0.44 years for the AHRI, Haiti DHS, and Manicaland data, respectively). Non-aggregated tables of ELPD differences and QQ RMSEs are presented in the Section B.3.2. Figure 7.5 overlays each data subset-/probability distribution-specific QQ plot within study and distribution choice. Greater deviation from the line of equality indicates less accurate replication of empirical quantiles. We present these plots for both sexes and all age groups in the Section B.3.1.

7.3.2 Distributional regression evaluation

We fit all five distributional regression specifications to all three of our datasets with sinh-arcsinh distributions and log-ratio dependent variables and compared the ELPDs and

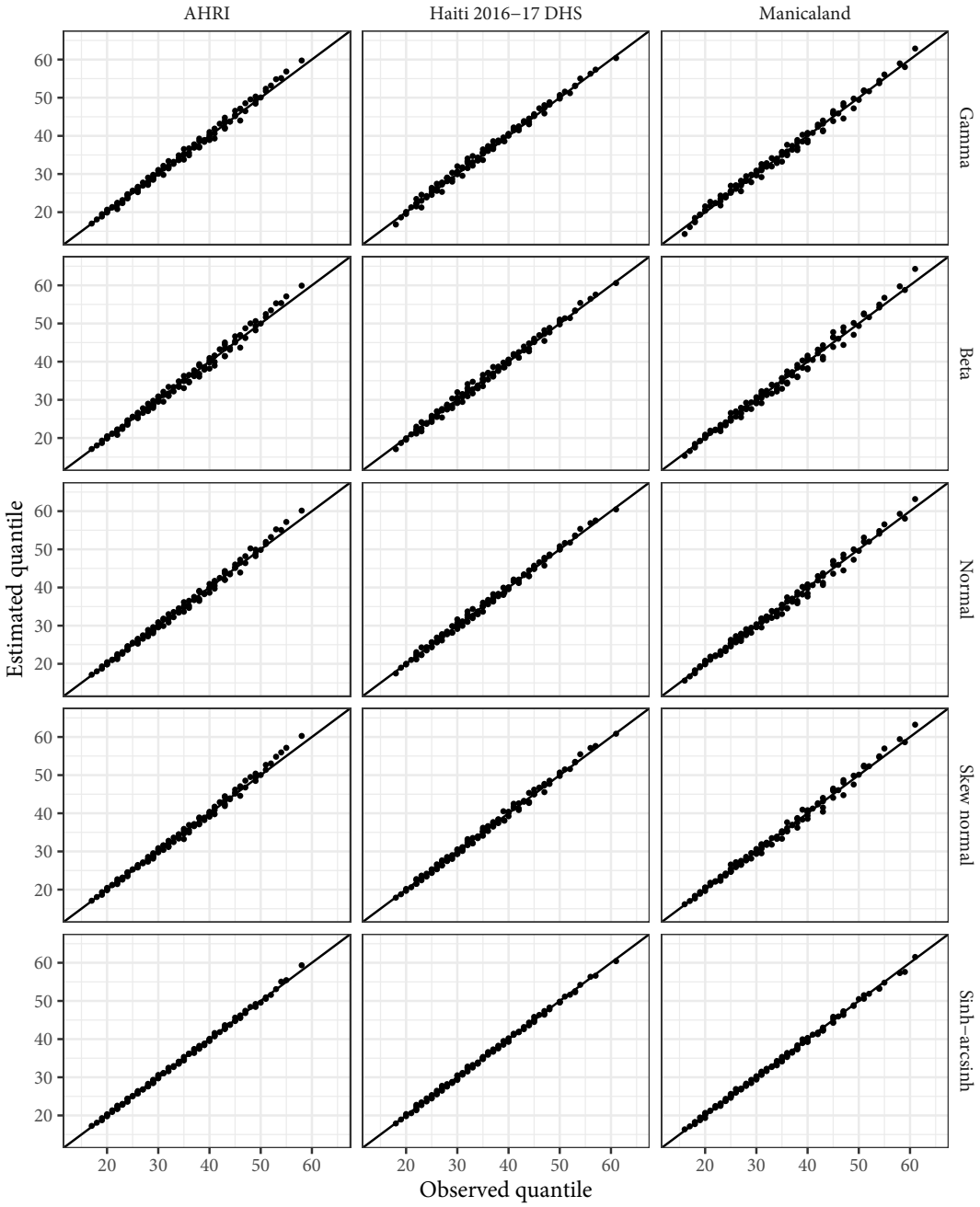

Figure 7.5: Overlaid quantile-quantile (QQ) plots for each probability distribution's best fit to data in all three main datasets. Presented quantiles range from 10th to 90th in increments of 10. Lines closer to the line of equality indicate better fit to empirical quantiles

Distribution	AHRI	Haiti 2016-17 DHS	Manicaland
ELPD			
Gamma	-14847.2	-2917.9	-13152.8
Beta	-14748.0	-2896.5	-13003.5
Normal	-14593.7	-2868.4	-12856.8
Skew normal	-14505.1	-2854.0	-12778.5
Sinh-arcsinh	**-14312.5**	**-2839.5**	**-12625.8**
QQ RMSE			
Gamma	0.83	0.82	0.95
Beta	0.99	0.82	1.11
Normal	0.82	0.68	0.97
Skew normal	0.77	0.65	0.85
Sinh-arcsinh	**0.36**	**0.37**	**0.44**

Table 7.4: Model comparison metrics averaged across all data subsets for all three datasets. Higher ELPD values indicate better fit. Lower QQ RMSE values indicate more accurate prediction of empirical quantiles. Bolded rows are best across all three data sets.

QQ RMSEs as before (provided in Table 7.5). Across all three datasets, the most complex distributional model (Distributional 4) had the highest ELPD and lowest QQ RMSE. When fit to the AHRI and Manicaland data sets (but not for the Haiti DHS), the most complex distributional model was a least two standard errors better than the next best model. Notably, the largest ELPD improvements came from moving from conventional regression (Conventional) to the simplest distributional model (1646.0 units, 361.0 units, and 2181.2 units in the AHRI, Haiti DHS, and Manicaland data, respectively). Full tables are available in Section B.3.2. Figure 7.6 shows the posterior predictive distributions from the conventional regression model and the most complex distributional model among men aged 16 years, 24 years, and 37 years in the AHRI data to illustrate the effect of distributional regression. Not only does the distributional model capture the high peak in the youngest age more accurately, but it also allows the variance of the distributions to change appropriately (beyond the change that naturally results from the log link function). Figure 7.7 illustrates posterior summaries among men and women in the AHRI data for all four distributional parameters for the conventional regression model, the simplest distributional model, and the most complex distributional model. The red estimates (Conventional Regression) of the three higher order parameters were constant across age and sex, whereas the blue estimates (Distributional Model 1) included independent, linear age and sex effects. The orange estimates (Distributional Model 4) were generated sex-specific splines with respect to age,

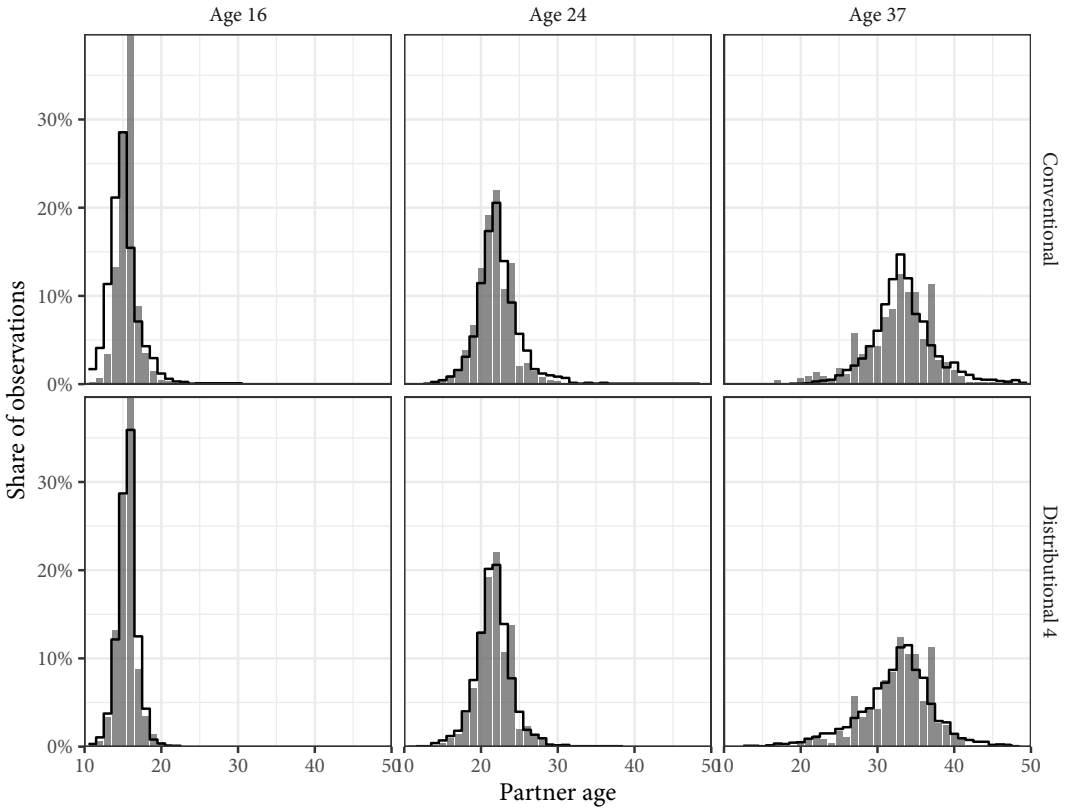

Figure 7.6: Observed partner age distributions (grey bars) and posterior predictive partner age distributions (lines) for conventional regression and the most complex distributional model among men aged 16, 24, and 37 years in the AHRI dataset. Posterior predictive distributions come from regression models fit to the entire AHRI dataset.

Model	AHRI	Haiti 2016-17 DHS	Manicaland
ELPD			
Conventional	52689.2	4777.8	21011.3
Distributional 1	54335.2	5140.8	23192.5
Distributional 2	54794.8	5138.7	23472.1
Distributional 3	55534.2	5196.7	24313.7
Distributional 4	**55841.9**	5207.6	**24516.1**
QQ RMSE			
Conventional	1.30	1.33	2.05
Distributional 1	1.15	0.98	1.89
Distributional 2	1.21	0.99	1.80
Distributional 3	0.93	0.91	1.34
Distributional 4	**0.66**	**0.84**	**1.04**

Table 7.5: ELPD and QQ RMSE values for all five distributional regression models fit to each dataset. The models increase in complexity from Conventional Regression to Distributional Model 4. Bolded ELPD values are more than two standard errors higher than the next best value in the column. Bolded QQ RMSE values are lowest in their column.

allowing for flexible variation across age and sex.

The third row of plots in Figure 7.7, which corresponds to the skewness parameter, illustrates the impact of incorporating sex and age effects into the model. The conventional regression model estimated that neither the distrbution for men nor women exhibited much skewness; the estimated parameter value was -0.05 (95% UI: -0.06 to -0.05) regardless of age, with 0.0 corresponding to perfect symmetry. However, when we allowed independent age and sex effects in Distributional Model 1, we estimated that at age 15, women's skewness was -0.26 (95% UI: -0.27 to -0.25) and men's was 0.11 (95% UI: 0.10 to 0.12).

The most complex model (Distributional Model 4) inferred sex-specific, non-linear variation with respect to age in all four distributional parameters. The non-linearity was particularly dramatic in the scale parameter among men. The scale value began at 0.05 (95% UI: 0.05 to 0.06) among 15-year-olds, peaked among 37-year-olds at 0.11 (95% UI: 0.10 to 0.11), and decreased back down to 0.05 (95% UI: 0.04 to 0.06) at age 64. Finally, Figure 7.8 presents inferred distributional parameters from Distributional Model 4 for both men and women for all three datasets. Based on those plots, the flexible model was justified for most distributional parameters in all three datasets. Were we to continue developing these models, this plot suggests that skewness might only need linear, sex-specific effects with

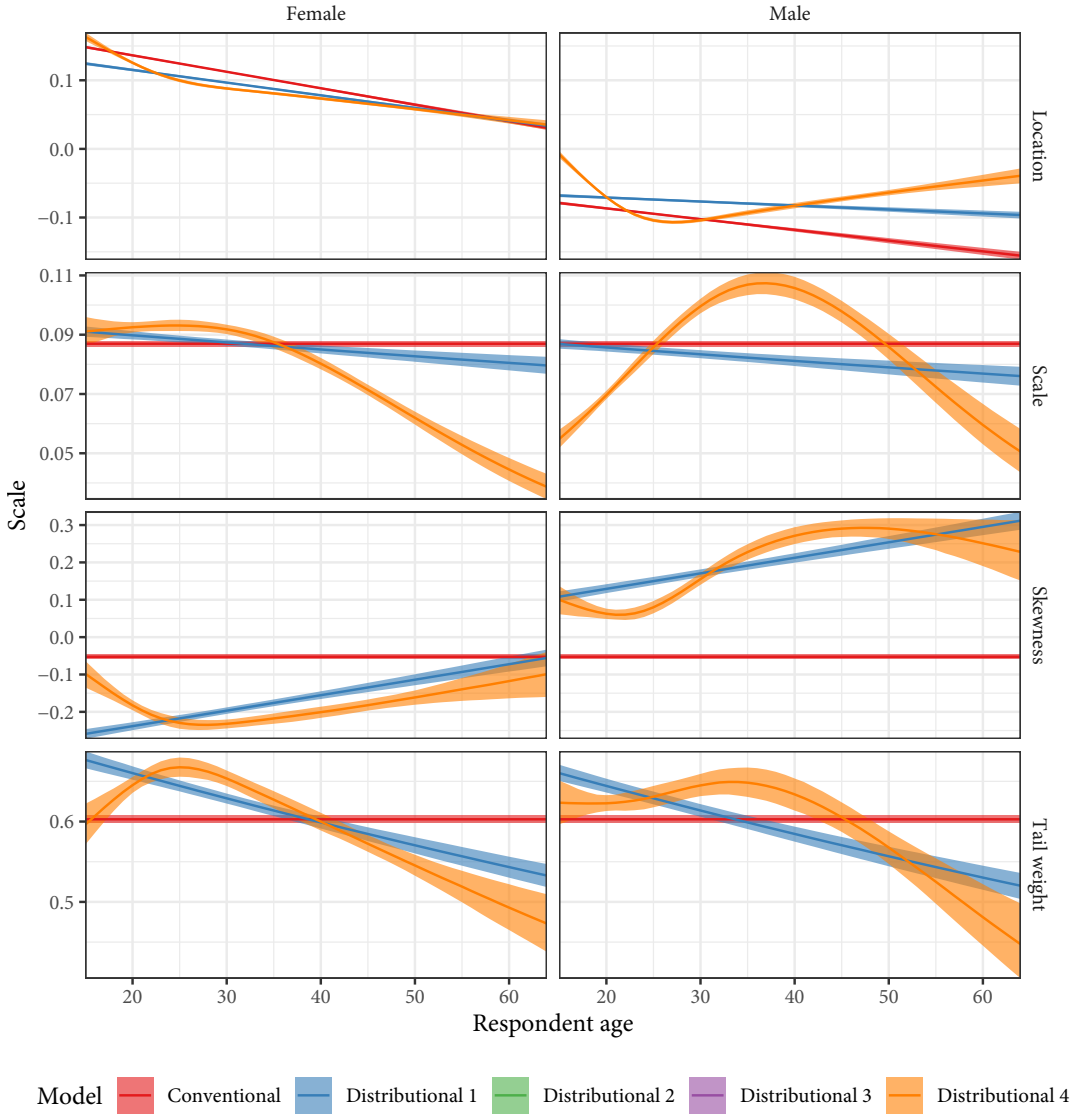

Figure 7.7: Estimated sinh-arcsinh distributional parameters from the conventional regression model, and distributional models 1 and 4 fit to the AHRI data. “Conventional" assumes no variation across age and sex, “Distributional 1" allows for independent age and sex effects, and “Distributional 4" includes sex-specific splines with respect to age.

respect to age. Interestingly, the 2016-2017 Haiti DHS and Manicaland estimates exhibit similar patterns across all four parameters, despite the different socio-cultural contexts surrounding partnerships in the two populations. We also note that the DHS does not collect data on adults aged 50 years and older, so our estimates in Haiti from age 50 to age 64 are purely extrapolated.

7.4 Discussion

We found that the sinh-arcsinh distribution reproduced observed sexual partner age distributions better than a number of other possible distributional assumptions across age and sex in three distinct data sets. We integrated this finding into a distributional regression framework using existing statistical modelling software. Even the simplest distributional regression in our set of candidate models far outperformed conventional regression, in which all moments except the first are estimated as constants. Our most complex distributional model fit better than all other models in all three data sets, suggesting that modelling these data benefits from the additional complexity.

These results indicate that distributional regression models with sinh-arcsinh distributions can accurately replicate age-/sex-specific sexual partner age distributions. This approach presents a number of advantages over previous methods. First, like Smid et al., it allows a unique distribution for every age-sex combination. As Figure 7.3 illustrates, partner age distributions can exhibit substantial, systematic variation across age and sex in any of the first four moments, so we must consider modelling strategies that allow for such variation. Second, distributional regression offers a principled method to propagate uncertainty through this estimation process.

Finally, distributional regression implemented through `brms` provides access to a deep set of hierarchical modelling tools that could enable estimation in a variety of low-data settings. We evaluated a small set of relatively simple distributional models in this work, but, theoretically, each distributional parameter could have its own, arbitrarily complex hierarchical regression model. Using these tools, one could estimate unique partner age distributions across levels of stratification that are substantively interesting but do not provide sufficient sample size for independent estimation (e.g. study sites or geographic

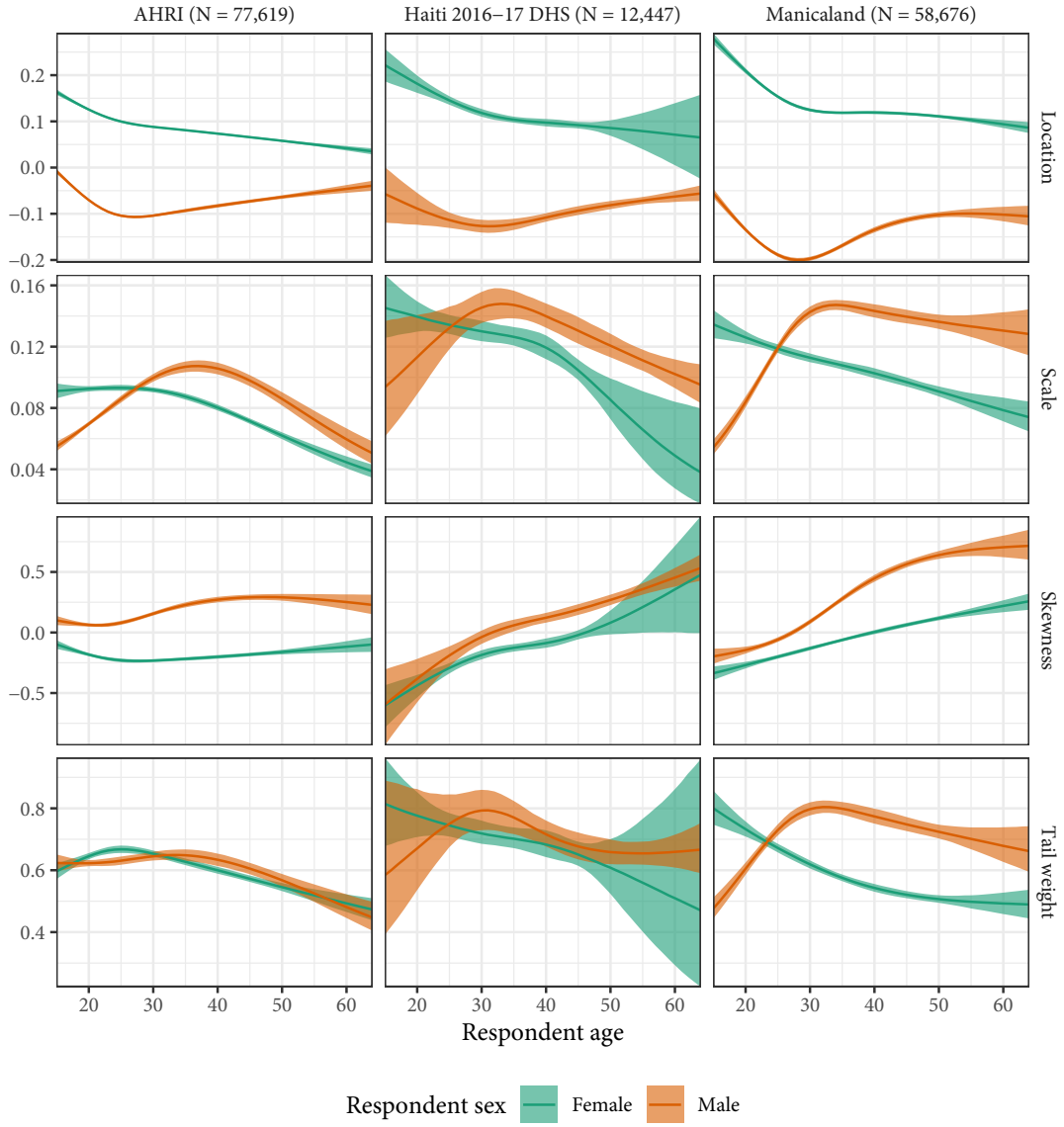

Figure 7.8: Estimated sinh-arcsinh distributional parameters for Distributional Model 4 fit to the three main data sets.

areas).

We have identified several limitations in this approach. First, the amount of data required to produce usefully precise estimates is not tested. Each additional distributional parameter introduces model parameters, so this method is more complex than conventional regression. The sinh-arcsinh distribution did fail to produce the highest ELPD in our smallest data subset (N = 170), but it was not significantly worse than the best distribution. More importantly, by integrating these data into a distributional modelling framework, we gain the ability to impose structure on these parameters, which could easily offset the cost of any additional model parameters.

Interpreting the inferred model parameters in sinh-arcsinh regression can also be difficult. Whereas conventional regression estimates the effects of covariates on expected values, the sinh-arcsinh distribution is parametrised in terms of a location parameter. This parameter correlates closely with the central tendency of the distribution, but it is not strictly the mean We can reparametrise the distribution so that we estimate a mean (and therefore effects of covariates on the expected value), but it is not currently possible in the probabilistic programming software that underlies `brms`.

Third, our analysis assumed that we were operating at a level of stratification at which partnerships are basically comparable, but any number of factors could lead to fundamentally different partner age distributions. For example, we did not control for whether the partnership was same-sex or the type of the partnership (married, casual, etc.). That said, our distributional framework would allow us to incorporate data on any of those factors directly into the model.

Despite these limitations, we believe that the strategy we present will work well in future projects that require estimates of partner age distributions. We plan to use these methods to produce age-mixing matrices to inform epidemic models of HIV, but there are a number of directions that could be explored. We are specifically interested in leveraging the spatio-temporal structure of the survey data used here. Hierarchical mapping exercises with household survey data are increasingly common in epidemiology, but estimating spatially varying partner age distributions would require an evaluation of how best to model higher order moments over space. For example, we would need to consider how the variance of partner age distributions varies by urbanicity.

Similarly, population-based studies typically collect far more detailed information on partnerships than we took advantage of here. Relationship type is a key confounder of the association between respondent age and partner age (that we ignored for the purposes of our experiments). We might, for example, expect the age distributions of casual partners to vary substantially from those of long-term cohabiting partners. Because we have built our model in a pre-existing regression framework, incorporating new covariates into any of the distributional regression specifications should be simple.

We believe that our framework offers a flexible, accurate, and robust method for smoothing and interpolating sexual partner age distributions, but these methods are not specific to partner age distributions. The sinh-arcsinh distribution is relatively easy to implement and does not incur high computational cost, so it could be applied in many settings. Even without the distributional regression framework we have used here, allowing the third and fourth moments of the distribution to vary from the "default" normal values could be valuable across a variety of applications.

Distributional regression is also underutilised in social science applications. We often work with large surveys that would comfortably support models for higher order parameters. Data requirements will vary by application and model, but, as we have shown here, even a simple distributional model can improve fit and avoid biasing estimates.

Chapter 8

Discussion

In the previous chapters, I have developed and tested a spatio-temporal epidemic model of HIV. I have discussed the strengths and weaknesses of each experiment in its own chapter. In this chapter, I discuss the body of work as a whole and outline a number of directions for future research.

8.1 Summary of work

In this work, I have developed a hierarchical Bayesian epidemic model of HIV that infers prevalence, incidence, and ART coverage simultaneously over space and time. This model incorporates all available biomarker data from household surveys and a wider set of ANC facility testing data than previous models. Unlike previous work, the model incorporates regional ART patient counts into the likelihood, rather than treating them as constraints for the epidemic model.

The model connects the epidemic model to a suite of generalised additive models for underlying dynamics, the parameters of which are inferred. Although EPP has always inferred time-varying HIV transmission rates, the model I proposed here allows transmission rates to vary over both space and time and additionally infers the rate of ART initiation and the initial state of the epidemic model. Inferring the ART initiation rate fits more naturally into a model defined in terms of partial derivatives than directly modelling ART coverage. I attached onto this model a version of the ART attendance model from Eaton et al. (2021),

which reconciles household-based survey data with facility-based ART programme data.

Specifying the model requires numerous design decisions related to both the underlying linear models and the likelihood. To empirically guide these decisions, I conducted a large model specification experiment with data from Malawi using a cross-validation strategy designed to promote accuracy in recent estimates. Consistent with Eaton & Bao (2017), I found that ANC testing data were more effectively modelled by the overdispersed beta-binomial distribution than by a binomial distribution. Given the beta-binomial ANC observation model, the cross-validation strategy identified a number of model features as better than others but did not conclusively identify one specific configuration as best.

For computational efficiency, the model comparison study omitted spatial transmission, assuming instead that all infections among residents of a region were endogenous to the region. I conducted a separate experiment examining the impact of spatial transmission on inferred incidence by applying five different assumptions about the degree of spatial mixing to one of the better configurations from the model comparison study. I found that incorporating spatial mixing into the model increased computation time by approximately a factor of six. Spatial mixing had spatially heterogeneous effects on inferred incidence, but the cross-validation strategy weakly suggested that the best model was still the one without spatial mixing.

With the results of the model specification tests, I fit the model to to the full dataset from Malawi to produce estimates of district-level HIV incidence rates, prevalence, and ART coverage among men and women aged 15-49 years. The model fit well to prevalence data and ART patient counts, but smoothed ART coverage across sex. There was substantial variation in incidence across districts, highlighting the importance of spatially resolved modelling. The model estimated that recent trends in the epidemic in Malawi have been dominated by rapid increases in ART coverage across district and sex, resulting commensurate decreases in incidence.

I then explored the potential of integrating this model into digital data collection mechanisms and updating estimates with new data as they become available. Rather than update estimates once per year in a process requiring substantial human intervention, we could automatically update estimates once every quarter. By supplementing real data with simulated future observations, I found that adding new data every quarter improved the precision of

estimates even far in the past and that the model was able to recover true incidence at a fixed date with a reasonable degree of precision after less than two years of data collection after the date. Because current data on HIV indicators reflect past incidence, the model was not able to forecast incidence reliably.

Finally, I presented a statistical model designed to smooth and interpolate self-reported sexual partner age data. Although my epidemic model has no age structure, sexual age-mixing data would need to be smoothed before inclusion into an age-structured version. The statistical model of partner age data integrates the four-parameter sinh-arcsinh distribution into a distributional regression framework. It fit well to observed partner age distributions from a variety of settings and could be extended to include any of the familiar hierarchical modelling tools available in **brms**.

8.1.1 Assessment of original aims

In Section 1.1, I outlined the specific aims of my thesis. In brief, they were to develop and validate a spatio-temporal epidemic model capable of estimating HIV incidence at a subnational level in data-sparse settings. Acknowledging that true model validation was impossible in the absence of direct measurement of incidence, I outlined a cross-validation strategy based on observable indicators and proposed comparisons to other estimates of HIV burden. I also specified that the model would be "sufficiently precise" if it estimated that the posterior probability of a 50% or greater decrease in incidence between 2010 and 2021 was less than 20% or greater than 80% in every district of Malawi.

In general, the proposed model met these aims. It fit well to most of the incorporated data and, in recent years, aligned well with other estimates of HIV burden in Malawi. Although these results do not indicate if predicted incidence is accurate, they suggest that the model is consistent with the data that are available and with other methods of HIV burden estimation. Notably, relatively diverse sets of model specifications led to similar fit to out-of-sample data, indicating that the same dataset could be generated by substantially varying incidence series. This means that the cross-validation strategy cannot be interpreted as cross-validating with respect to incidence. In Chapter 6, the model recovered true incidence after approximately 1.5 years of new data. Finally, the model achieved the desired level of precision: posterior probabilities of 50% or greater decreases in incidence between 2010 and 2021 of less than

20% or greater than 80%. In fact, the estimates of the changes in incidence over this period were considerably more precise than the target level of precision.

8.2 Strengths

The model proposed here incorporates a number of methodological improvements over its predecessors.

8.2.1 Multivariate modelling of HIV

First, I have focused on incorporating more population-level data sources than previous models in as robust a way as possible. Following Eaton et al. (2021), I fit simultaneously to HIV testing, ART adherence, and recency data from large household surveys. Independent measurement of ART coverage provides a particularly important point of leverage for reconciling prevalence data and patient counts. Given total population, prevalence implies a number of PLHIV that should, in theory, be consistent with the observed patient counts; direct measurement of ART coverage gives us an alternate view of this system, improving inference.

Unlike Naomi, this model was designed to infer changes in the epidemic over time. Naomi fits to a single survey at an initial point in time, and projects up to two time points in the future by linearising the epidemic. In a typical use case, the user inputs a survey from 2015 and produce estimates for 2015, 2018, and 2020. Because Naomi ignores progression and mortality dynamics and takes such large time steps, it is not well-suited for estimation of trends in incidence. Without considering disease progression dynamics, a single survey provides little information about changes in incidence. As discussed in Chapter 1, reliable estimates of how HIV incidence varies over both space and time are critical for effective policy-making.

8.2.2 ANC data and observation model

I have also included a broader set of ANC data than previous models (Bao, 2012; Eaton et al., 2019). With access to granular data from Malawi, I re-aggregated facilities that are already included in EPP by quarter, as opposed to year, to align them with the modelled time step. I also produced regional measurements of ANC prevalence by aggregating over the hundreds of facilities that are not currently including the UNAIDS estimation process. Fitting to all facilities individually might be preferable, but even including a regional aggregate should reduce the selection bias induced by using historical sentinel sites. I have also built the insight from Eaton & Bao (2017) into a more natural observation model, using the overdispersed beta-binomial distribution instead of the normal approximation proposed by Eaton et al. The cross-validation exercise in Chapter 3 suggested that the beta-binomial observation model captured the variation across facility and time better than a binomial model.

8.2.3 Improved model of treatment seeking

Perhaps the most significant change this model makes in terms of data is its handling of ART patient counts. Forcing an epidemic model to match exogenous patient counts exactly at every time point requires us to account for parameter sets that do not produce enough PLHIV without treatment under the current to reach the target number of patients. The model either needs to generate new PLHIV, distorting the meaning of the parameters with respect to incidence, or carry the difference forward and fail to match the target counts. This problem becomes substantially more complex when the epidemic model uses a more granular level of stratification than the ART patient counts. In this work, I fit the predicted number of PLHIV on treatment to ART patient counts, obviating the need for such constraints, while still matching the observed patient counts nearly perfectly.

This approach acknowledges that patient count series must contain a small amount of error and gives the user more flexibility in how they prepare their ART data. Existing models need complete series of counts, requiring the user to remove outliers and interpolate over time. My model requires no interpolation and can handle outliers as well as any other statistical model. The user can drop any data points they believe to be erroneous, and the

model will smooth over the gap.

I also note that if treatment coverage continues to increase and times until initiation decrease, the change in the number of ART patients will become an increasingly accurate measurement of true new infections. If every new person LHIV began treatment the moment they were infected and no one on treatment died, then the change in patient counts would be exactly the number of new infections. That case does not reflect the real world, but the thought experiment highlights the potential value of ART programme data in models of HIV incidence.

The ART observation model is facilitated by the novel model of ART initiation I have developed. Rather than model ART coverage as in Eaton et al. (2021), I model ART initiation directly. In a temporally structured setting, modelling coverage over time and imposing it on an ODE-based epidemic model creates a numerical problem similar to the one in EPP: it introduces a constraint that the epidemic model might not be able to meet in an internally consistent way. Modelling initiation directly introduces no such constraint and offers inference that might be interesting on its own.

8.2.4 General model developments

In a similar vein, this model infers the initial state of the projection. The earliest data points in sub-Saharan Africa were collected in the mid-1990s, so beginning projection in 1970 is, in one sense, computationally wasteful. On the other hand, the differences between my estimates and the UNAIDS estimates in earlier time periods suggest that this choice might come with certain disadvantages. A user could begin projection of my model much earlier, or we could consider methods to align initial transmission rates with external estimates.

I have also considered, but not tested, the hypothesis that beginning projection after the peak of the epidemic makes the approximate inference strategy more appropriate. If the overall trajectory of the epidemic is more closely determined by early transmission rates than later transmission rates, then earlier transmission rate parameters will be correlated with much later parameters in potentially complex ways that make the Laplace approximation inappropriate. I have found this to be true in an informal simulation study and in an application to a model of COVID-19 (Flaxman et al., 2020), but I have not tested the

intuition formally.

I have also proposed a novel parametrisation for transmission rate of HIV that extends existing approaches to space and integrates them into a broader family of stochastic time series models. The transmission rate model proposed in Section 2.4.2 could be viewed as a conceptual generalisation of the "r-spline" model with respect to time (Hogan & Salomon, 2012) and an extension with respect to space, although those authors modelled force of infection, not log-transmission rates. ARIMA models are well understood and offer desirable statistical properties in forecasting applications.

8.2.5 Robust cross-validation

Because we are most interested in estimates of recent and near-future incidence, I have used a cross-validation scheme that focuses on short-term forecasting of programmatic data. This strategy is designed to mimic new data acquisition and therefore does not include any degree of spatial cross-validation. Because these types of models are designed to be used on fixed sets of known administrative regions, I argue that spatial cross-validation is inappropriate. The strategy used here also acknowledges the fact that the bulk of future data on HIV indicators in SSA will likely come from programmatic sources, not large household surveys.

The model specification experiment conducted in Chapter 3 gives us some reassurance that the decisions used later in the work were, if not better than others, then at least comparably good. I did not find that one model was best, although certain individual choices did lead to better out-of-sample fit. Such model specification tests are often relegated to sensitivity analyses to verify that intuitively made choices have not had too large of an effect on results, but, given the sheer number of decisions needed to specify the model in Chapter 2, a more systematic approach was needed. The model is designed to run on a high-performance computing (HPC) cluster, so fitting hundreds of specifications is as easy as fitting one.

8.2.6 Comparison to Sartorius et al. (2021)

There are fundamental similarities between this work and Sartorius et al. (2021), so a detailed delineation of the differences between the two methods is valuable. First, Sartorius

et al. fit to posterior prevalence predicted by the geostatistical model from Dwyer-Lindgren et al. (2019), not to the data that the geostatistical model was originally fit to. This approach encouraged consistency between the two sets of estimates, but fitting directly to the data used by Dwyer-Lindgren et al. is intuitively and technically preferable. The authors approximated the district-level posterior distribution of temporally varying prevalence from the geostatistical with a multivariate normal distribution designed to capture the correlation in prevalence over time within each district. However, it is unclear how the uncertainty in predicted incidence that resulted from fitting to these multivariate normal approximations compared to the uncertainty that would arise from fitting the incidence model to the original data. In particular, by fitting to posterior prevalence in years as early as 2000, their model could be unduly confident in times when there were effectively no data. Because the model proposed here is fit directly to data, it should not raise such concerns.

Although the two models function similarly, their respective treatments of spatial dynamics are substantively different. Sartorius et al. fit EPP independently for every district, meaning that any spatial correlation in their estimates was an artefact of the inherent spatial correlation in prevalence from the geostatistical model. In the model proposed here, statistical strength is pooled across districts in the models for initial prevalence, ART initiation, and HIV transmission rates, resulting in more reliable estimation in districts with less data. More complex spatial priors for each any underlying linear model could also be considered. Although I did not include spatial in the final model presented here, the model from Sartorius et al. cannot accommodate explicit spatial mixing without substantial modifications.

Finally, the two models varied dramatically in their treatments of ART programme data. First, Sartorius et al. did not have access to the spatio-temporally granular ART programme data used in this analysis and in Eaton et al. (2021). Instead, they conducted a literature review that found estimates of subnational ART coverage in 29 of 44 countries. In the 15 countries with no subnational ART coverage data, they assumed that ART coverage in every district was exactly national-level ART coverage. Treatment coverage is a key determinant of population-level incidence, so assuming that it was constant over space fundamentally limited their ability to detect the spatial disparities a model of this kind is designed to detect. It is also unclear how this approach exacerbated or resolved the discrepancies described by

Eaton et al. and in Section 1.3.3 between household-based coverage data and facility-based patient count data.

8.3 Limitations

Taking my PhD work as a whole, I have noted several limitations. Some of these are fundamental, and some can be addressed in future research. First, the compartmental model lacks certain structure across which outcomes vary substantially. I have used four CD4 compartments in alignment with Thembisa, whereas Spectrum uses seven. Assuming estimates of mortality and progression are sufficiently reliable for each of the seven CD4 groups, a more detailed model of disease progression could offer more accurate estimates of mortality and incidence. I have also omitted progression through ART duration in order to minimise the number of compartments.

More importantly, I have omitted age structure. Initial conceptualisations of this model included single-year ages, but the additional computational cost made inference impractical. Model fits expired after three days on the Imperial College HPC cluster without converging even using the approximate inference strategy. I have attempted to account for the lack of explicit age structure by age-weighting mortality and progression rates over time using estimates from Spectrum and EPP-ASM. This problem is not unique my model; EPP-ASM separates its age-structured demographic model from its age-aggregated disease progression model. Further work is needed to identify computational strategies that are efficient enough to fit a combined demographic-epidemic model to data.

These omissions point to a broader limitation of this analysis: due to time constraints, I was only able to cross-validate a relatively small subset of the design decisions needed to specify the model. I did not interrogate the effects of the structural decisions made in the specification of the compartmental model: specifically, the choices of CD4 categories, the omission of age, and the progression of people on treatment. These decisions could have material effects on inferred incidence. For example, a model that incorporated age would naturally capture the effects of an ageing population of PLHIV on age-specific incidence, as well as the differential effects by age and sex (Akullian et al., 2017). Critically for this exercise, the age-/sex-structures of populations can vary substantially over small areas, so

any bias in incidence induced by omitting age dynamics likely affects the spatial distribution of incidence. As it stands, my model must absorb the net effects of these dynamics into the spatio-temporally varying transmission rate parameters.

Similarly, the choice of CD4 categories could substantially impact these estimates. If reliable estimates of progression and mortality are available for finer CD4 categories than the ones used here, a model that stratifies by those categories will produce more accurate estimates of mortality than this model. This effect will be particularly pronounced in cases where the distribution of PLHIV across the finer categories is not uniform and the changes in mortality are nonlinear, because the approach used here is effectively a linear approximation. More accurate estimation of mortality would result in a more accurate decomposition of changes in prevalence and therefore allow the model to identify incidence more precisely.

Conversely, I did not consider simpler model structures. It is possible that in settings with high ART coverage, a model stratifying by fewer CD4 categories or not stratifying at all could offer sufficiently accurate estimates of mortality. Such a strategy would reduce the computational cost of each projection, potentially facilitating inference in larger cases than the one presented here. However, an approach that separates PLHIV only by treatment status might be most effective in settings with reliable cause-specific mortality data, in which mortality can be inferred directly. Relying on fixed assumptions about AIDS mortality with and without treatment in a model with no CD4 stratification would likely compound the issues raised in the previous paragraph.

The model of ART patient counts represents an improvement over many existing approaches to incorporating ART programme data, but it could be substantially improved. In particular, it assumes that attendance at time t is independent from time $t-1$; that is, that individuals are newly allocated to a treatment region at every time point. I expect that this results in overestimation of variance in patient counts. Relaxing this assumption presents a substantial challenge. We could track PLHIV on treatment by district-of-treatment in addition to district-of-residence, but this would multiply the size of the compartmental model by slightly less than the number of regions. As I will discuss in Section 8.5, a more productive approach might be to focus on this problem at the national first, by integrating model of ART initiation into existing models of testing and treatment behaviour.

Further, we only know how many people are receiving treatment at each facility, not where

those people typically reside. The ART attendance model reconciles the counts of PLHIV on treatment implied by ART coverage and prevalence with observed patient counts but does not fit directly to data measuring where people from each region seek treatment. If the model sees far more or far fewer people on treatment in a given district than it would expect given population, prevalence, and ART coverage, then it adjusts the number of people receiving treatment in that district and its neighbours. In many cases, this model is not well-identified by data.

A separate approximation is that I aggregated data from hundreds ANC facilities to produce regional aggregates, while still treating historical sentinel sites fundamentally different than the other sites. Aggregating over facilities masks meaningful variation, and separating out former sentinel sites likely gives them undue influence. However, testing a model that fits to every site was out of scope for this analysis. Future models should consider how best to model the full set of ANC testing data.

I accounted for systematic differences between prevalence data from ANC facilities and household surveys with a model that takes logit-transformed general-population prevalence as the intercept for a logit-linear model for facility-specific prevalence. This method is directly inspired by the one proposed by Bao (2012), but in this case, I included facility-specific random effects for both the level and trend with respect to time in logit-prevalence. The facility-level temporal trends allow the representativeness of each facility to change with the ageing population of PLHIV, but if surveys are not conducted regularly in the future, the estimates of these trends will become increasingly unreliable.

I have also ignored all routine HIV testing and diagnosis data from the general population. We can reasonably assume that testing behaviour is correlated with HIV infection risk, so routine testing data by themselves are unlikely to offer accurate measurement of HIV prevalence in the general population. Re-testing among PLHIV is also thought to be common (Maheu-Giroux et al., 2019), introducing another source of bias. This problem is closely related to the inaccuracies in the ART attendance model and could similarly be resolved with a more accurate model of testing and treatment seeking.

Finally, most of the results in this work were the result of fitting with approximate inference strategy that assumes that all hyperparameters are fixed and that all other parameters comprise a multivariate normal distribution. Even with this strategy, the model still takes

one to two hours to fit in Malawi. There are likely to be improvements made in the implementation, but alternate computational strategies are worth exploring.

8.4 Implications

This work addresses several notable methodological gaps, some specific to HIV modelling and others applicable to infectious disease modelling more generally. As has been highlighted repeatedly during the ongoing SARS-CoV-2 pandemic, statistical and computational concerns about how to connect epidemic models to data are far from resolved (Chatzilena et al., 2019; Flaxman et al., 2020; Grinsztajn et al., 2021). Certain lessons from this work apply specifically to HIV, while others might be instructive in other applications.

Setting aside the utility of the estimates themselves, much of the methodological work from my thesis can be applied directly to existing models of HIV. The beta-binomial ANC observation model handles ANC testing data in a more natural way than the one proposed by Eaton & Bao (2017) and does not require artificially inflating variance. The beta-binomial likelihood could be added to any existing models of ANC testing data.

The model of treatment seeking, although imperfect, could also be incorporated directly into existing models. The generalised negative binomial distribution I have used here enabled comparison across time and across geographic units of dramatically varying sizes. More fundamentally, treating ART patient counts as data, not as constraints, allowed us to infer ART initiation.

More broadly, this work demonstrates a largely successful fusion of Bayesian hierarchical statistical models with an epidemic model. Epidemic models take advantage of the fundamental dynamics of infections disease to predict burden with a minimum of parameters, but they are rarely underpinned by hierarchical statistical models. The simple transformation at the beginning of Section 2.4.2 illustrates that standard epidemic models are, in a manner of speaking, compatible with hierarchical modelling. Equation (2.14) would be just as valid a model of incidence if it included a more complex regression specification. If we had data on risk factors that varied across observation units, we could add a linear regression component to the expression for $\log \lambda$:

$$\log \lambda = \log \beta + \gamma \cdot \mathbf{X}_t + \log \frac{I}{N}. \tag{8.1}$$

This framework would lessen my model's dependence on random effects and could even incorporate classical causal inference regression designs.

This work also illustrates the value of multivariate statistical models in infectious disease epidemiology. I have modelled underlying dynamics using the epidemic model, which provided all required indicators to specify likelihoods for many different sources of data. Fitting to prevalence, ART coverage, and ART patient counts simultaneously identified an epidemic model that would otherwise likely be non-identifiable. Estimating the underlying, true epidemic with a model that leveraged the principles of infectious disease epidemiology provided a coherent, self-consistent way to reconcile varied data sources.

8.5 Future work

This work could be improved upon or extended in any number ways. First and foremost, future work must examine methods for incorporating age into the epidemic model. It is possible a different implementation of the model would facilitate efficient computation. Regardless, by omitting age, the dynamics of mortality and disease progression used here are likely to be biased.

Between Chapters 3, 4 and 6, I made extensive use of the proposed cross-validation strategy to assess the identifiability and precise specification of the epidemic model. However, each of these exercises was limited in scope due to time and computational constraints. Chapter 3 focused primarily on the transmission rate model because it is largest unknown determinant of incidence, while Chapter 6 combined real and synthetic data to maximise short-term utility for applications in Malawi. In restricting the scope of these analyses, I have limited their abilities to answer important, broad questions about identifying incidence and model specification, but I two new analyses extending their methodologies.

First, as discussed briefly in Section 6.4, the partial simulation study presented in Chapter 6 could be replaced with a larger-scale study on broader set of fully simulated datasets. In this analysis, we would simulate hundreds or thousands of datasets from randomly generated

epidemics resembling the epidemics of SSA. We could then assess the identifiability of incidence under the correct model specification by fitting to each dataset without holding any data out and comparing each inferred incidence series to the true incidence series originally used to generate the data. Because correct specification is impossible in real-world applications, this analysis could be replicated with varying, incorrect specifications to measure impact of specification on how accurately incidence can be inferred.

These simulated datasets could also be used in the same analysis to generalise the results from Chapter 6. Averaging across all simulated datasets, we could more robustly estimate the amount of data needed after a fixed time point to estimate incidence at that time point. Perhaps more usefully, we could estimate the number of years of data needed to recover true incidence for each dataset separately and identify which epidemic characteristics (high or low prevalence, high or low ART coverage, etc.) affect the amount data needed. This would provide countries with a set of heuristics for understanding how reliable their incidence estimates are, even in the absence of direct measurement. With these methods, the first study would measure how well incidence is identified by data under correct and incorrect model specifications and provide countries with guidance on how many years of data are needed to accurately predict incidence at a fixed point in time.

A second study would extend the cross-validation exercise presented in Chapter 3 to include as many of the decisions required to specify this model as possible. For a fixed set of real data, this analysis would test the transmission rate model, the ART initiation model, the model of the initial state of the epidemic, the prior assumptions for each of these models (in particular, the use of i.i.d. random effects), the subdivision of PLHIV into CD4 categories, and the assumptions about mortality and progression. Because of the computational challenges involved in adding age structure to this model, I do not believe it is practical to test in this analysis. Even if each of these seven decisions was restricted to three possible values, this analysis would require fitting 2,187 unique models. This is not an insurmountable computational challenge, but parsing the results of this analysis would be more complex than those of Chapter 3. This analysis would identify the best specification or set of specifications for a single country and provide a template for similar exercises in other countries.

To improve the model of treatment seeking and begin to incorporate general population

routine testing data, the models of ART initiation and ART patient counts could be incorporated into the Shiny90 model, which is a national-level model of testing and treatment (Maheu-Giroux et al., 2019). The Shiny90 model fits simultaneously to self-reported HIV testing data and HIV testing service programme data but does not infer incidence or ART initiation. Modelling ART initiation and HIV testing at the national level would allow us to infer incidence and treatment seeking behaviour from widely available routine testing and treatment data.

Noting that many of the weaknesses I have discussed stem of choices made in the name of computational efficiency, a comprehensive review of software and inference procedures for deterministic inferential epidemic models would be valuable. I have implemented this model in C++ with TMB, but developing these types of models on graphical processing units (GPUs) could lead to significant improvements in computation time.

An alternate approach could be to omit costly ODE simulations and replace prediction with a black-box, learned simulator. Recent work in particle- and mesh-based physics simulation has shown that neural networks can be taught to replicate extremely complex physical systems, dramatically reducing the amount of computation required for simulation (Pfaff et al., 2021; Sanchez-Gonzalez et al., 2020). On the scale of physics simulations, ODE-based epidemic models are quite simple and stable, so I believe that a learned simulation approach is worth investigating. Replacing forward Euler or Runge-Kutta with a GPU-compatible neural network prediction step could dramatically improve computation time, facilitating inference with robust MCMC methods in cases where inference is otherwise not possible.

8.6 Conclusion

In this work, I have developed a model of HIV that allows for flexible inference of many key epidemiological indicators from population-level data. In doing so, I have demonstrated how combining hierarchical linear models, epidemic models, and disparate data sources facilitates inference of otherwise-unobservable epidemic dynamics. I have proposed several directions for future work, many of which could improve inference for epidemic models of other infectious diseases.

Appendix A

Supplementary information

District	ISO-3166-2 code
Northern	
Chitipa	CT
Karonga	KR
Likoma	LK
Mzimba	MZ
Nkhata Bay	NB
Rumphi	RU
Central	
Dedza	DE
Dowa	DO
Kasungu	KS
Lilongwe	LI
Mchinji	MC
Nkhotakota	NK
Ntcheu	NU
Ntchisi	NI
Salima	SA
Southern	
Balaka	BA
Blantyre	BL
Chikwawa	CK
Chiradzulu	CR
Machinga	MH
Mangochi	MG
Mulanje	MU
Mwanza	MW
Neno	NE
Nsanje	NS
Phalombe	PH
Thyolo	TH
Zomba	ZO

Table A.1: ISO-3166-2 codes for every district in Malawi with "MW" prefix removed.

Figure A.29 compares district-level incidence between 2000 and 2018 from this study and Sartorius et al. (2021). The correlation between the two sets of point estimates was 0.94. The estimates from this study were within the LBD credible interval in 73.7% of cases, while CIs from this study covered the LBD estimates in 53.4% of cases.

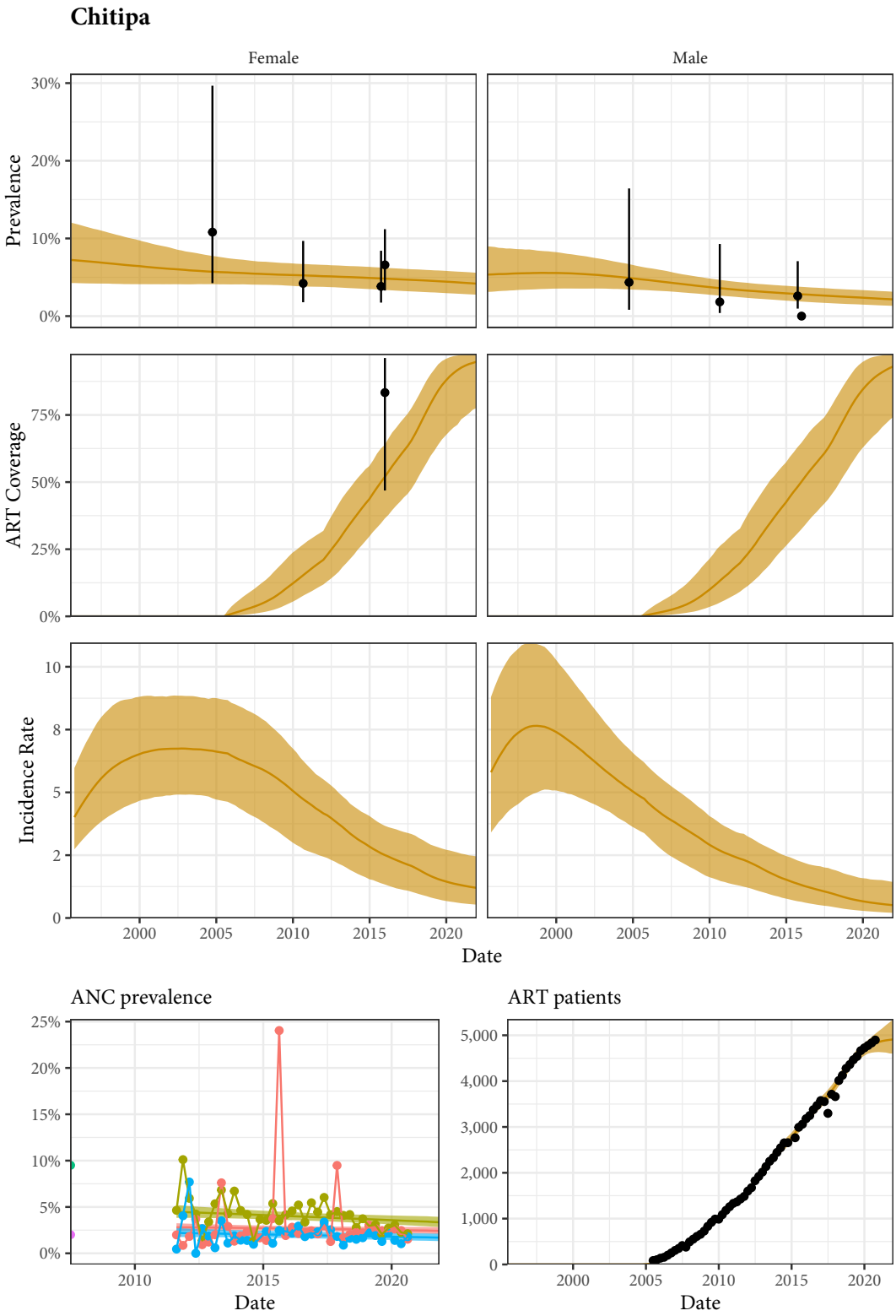

Figure A.1: Estimated prevalence, ART coverage, incidence, ANC prevalence, and ART patient counts in the Chitipa district of Malawi with household survey data, ANC facility data, and programmatic reporting data (points). Different colours on panel "ANC prevalence" indicate different ANC facilities.

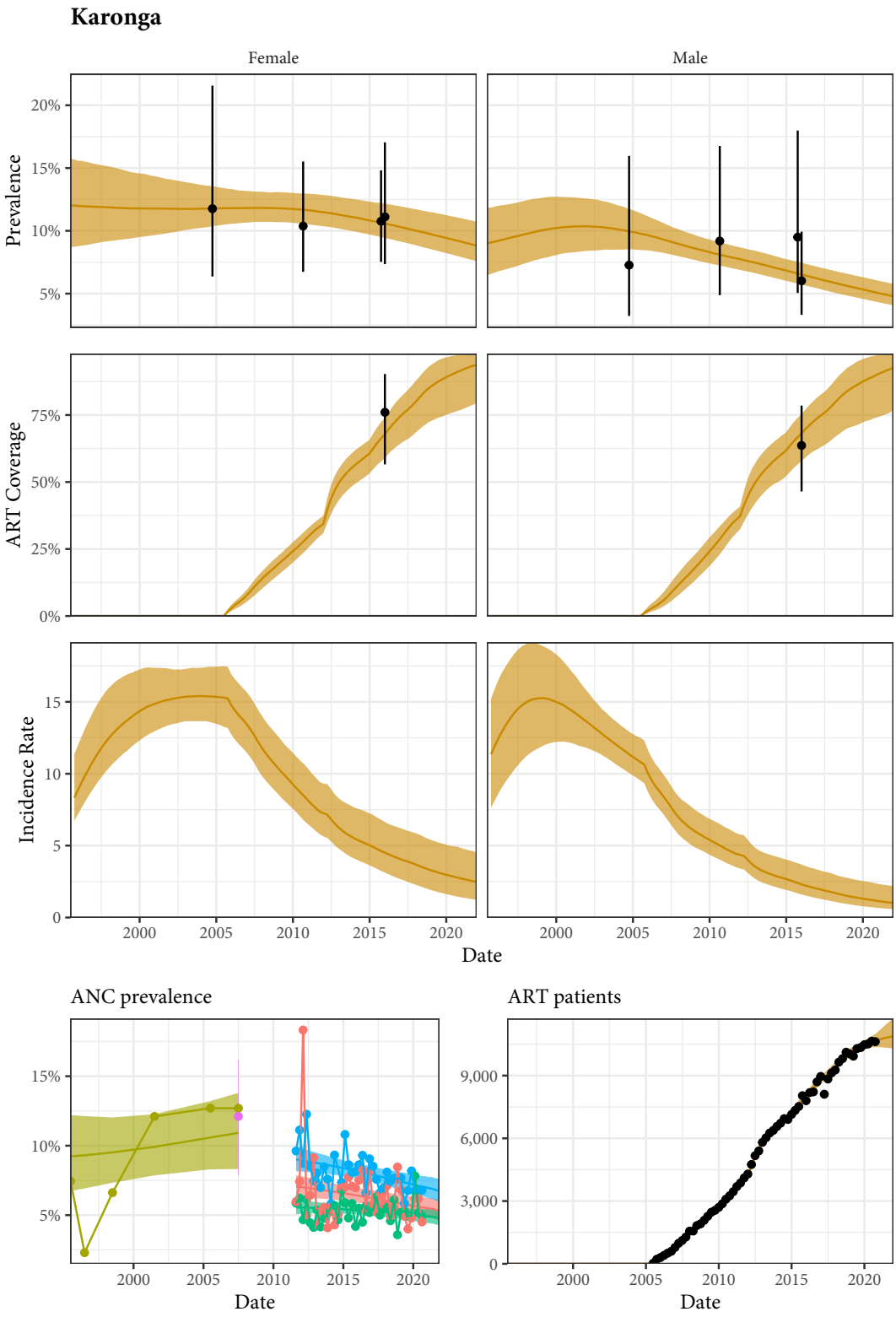

Figure A.2: Estimated prevalence, ART coverage, incidence, ANC prevalence, and ART patient counts in the Karonga district of Malawi with household survey data, ANC facility data, and programmatic reporting data (points). Different colours on panel "ANC prevalence" indicate different ANC facilities.

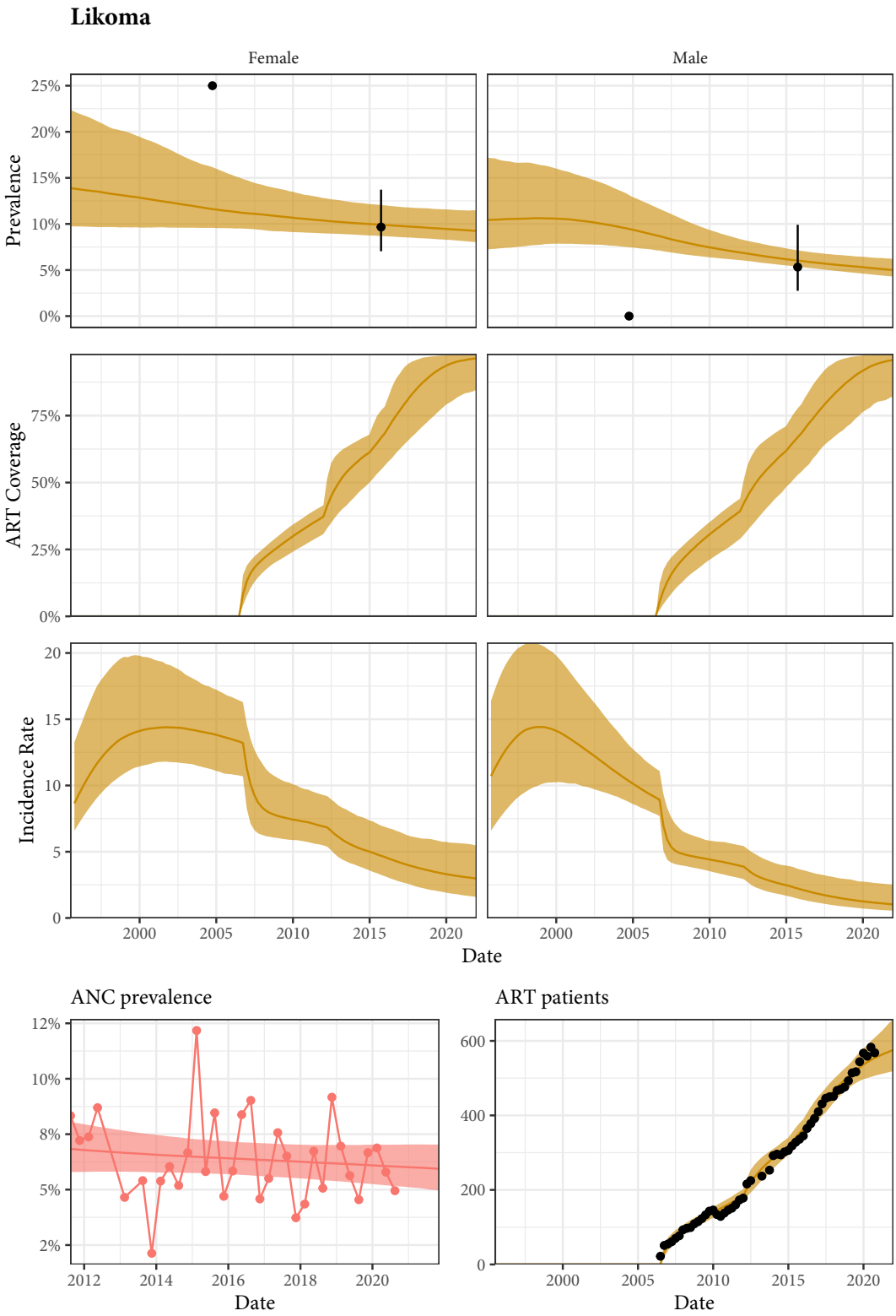

Figure A.3: Estimated prevalence, ART coverage, incidence, ANC prevalence, and ART patient counts in the Likoma district of Malawi with household survey data, ANC facility data, and programmatic reporting data (points). Different colours on panel "ANC prevalence" indicate different ANC facilities.

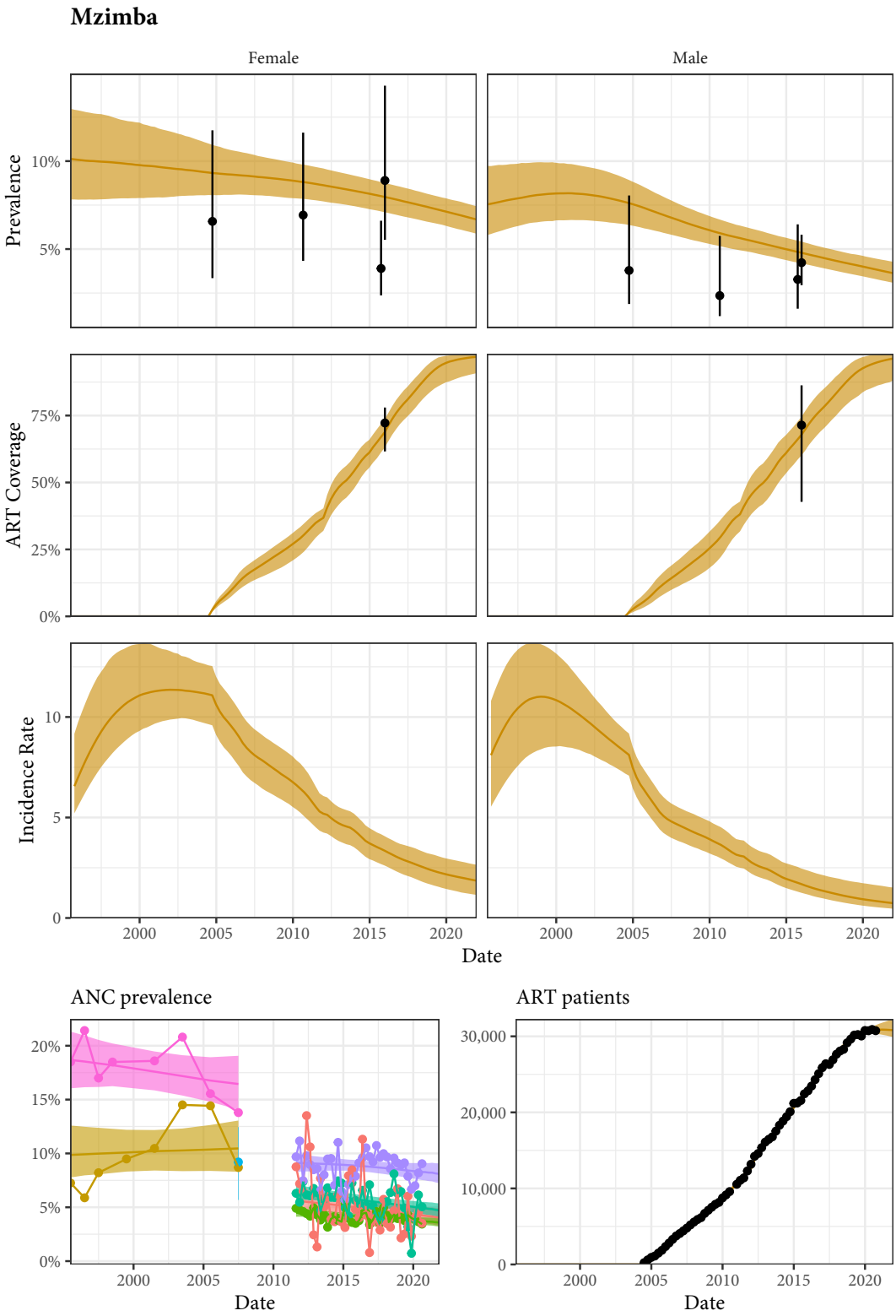

Figure A.4: Estimated prevalence, ART coverage, incidence, ANC prevalence, and ART patient counts in the Mzimba district of Malawi with household survey data, ANC facility data, and programmatic reporting data (points). Different colours on panel "ANC prevalence" indicate different ANC facilities.

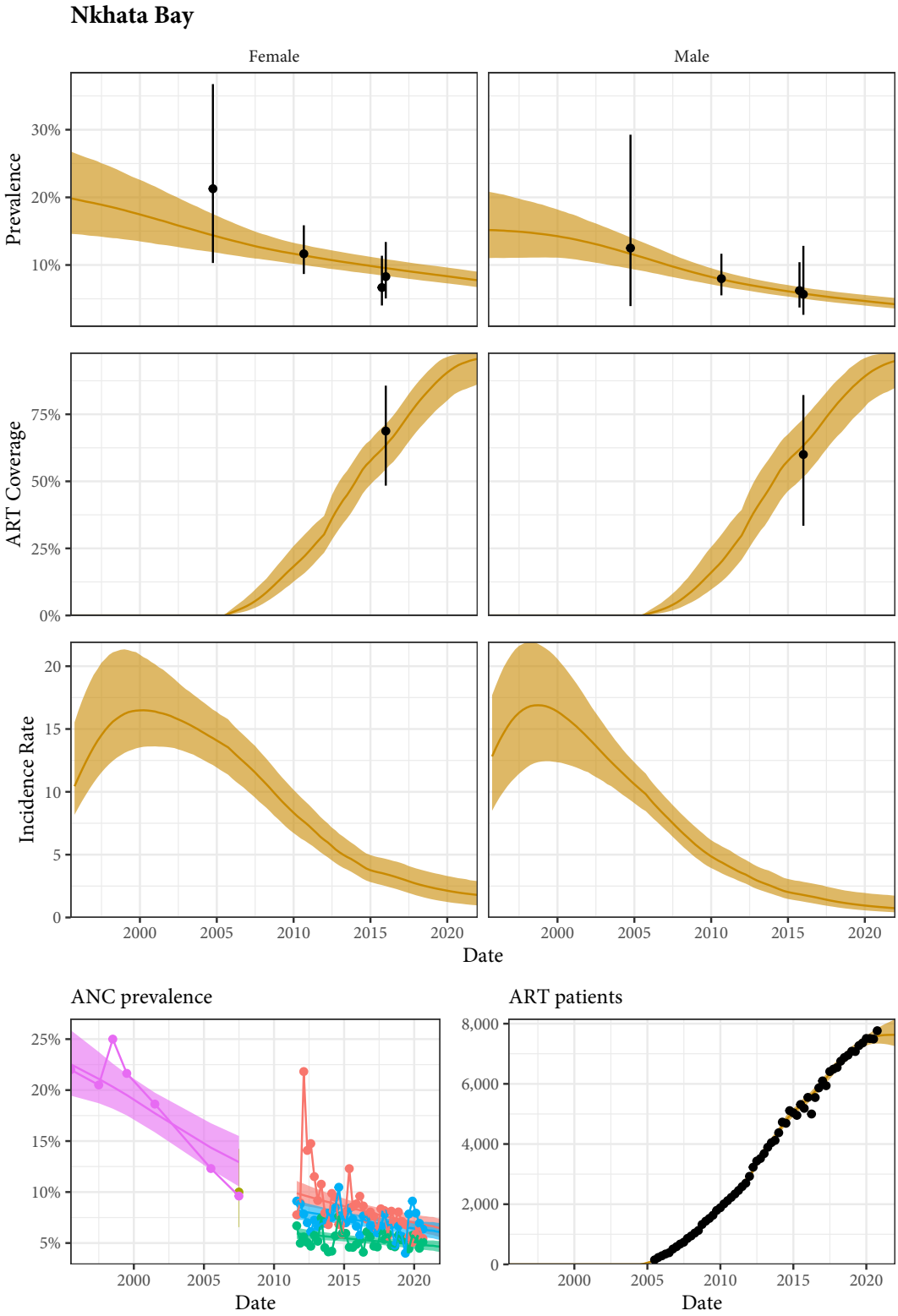

Figure A.5: Estimated prevalence, ART coverage, incidence, ANC prevalence, and ART patient counts in the Nkhata Bay district of Malawi with household survey data, ANC facility data, and programmatic reporting data (points). Different colours on panel "ANC prevalence" indicate different ANC facilities.

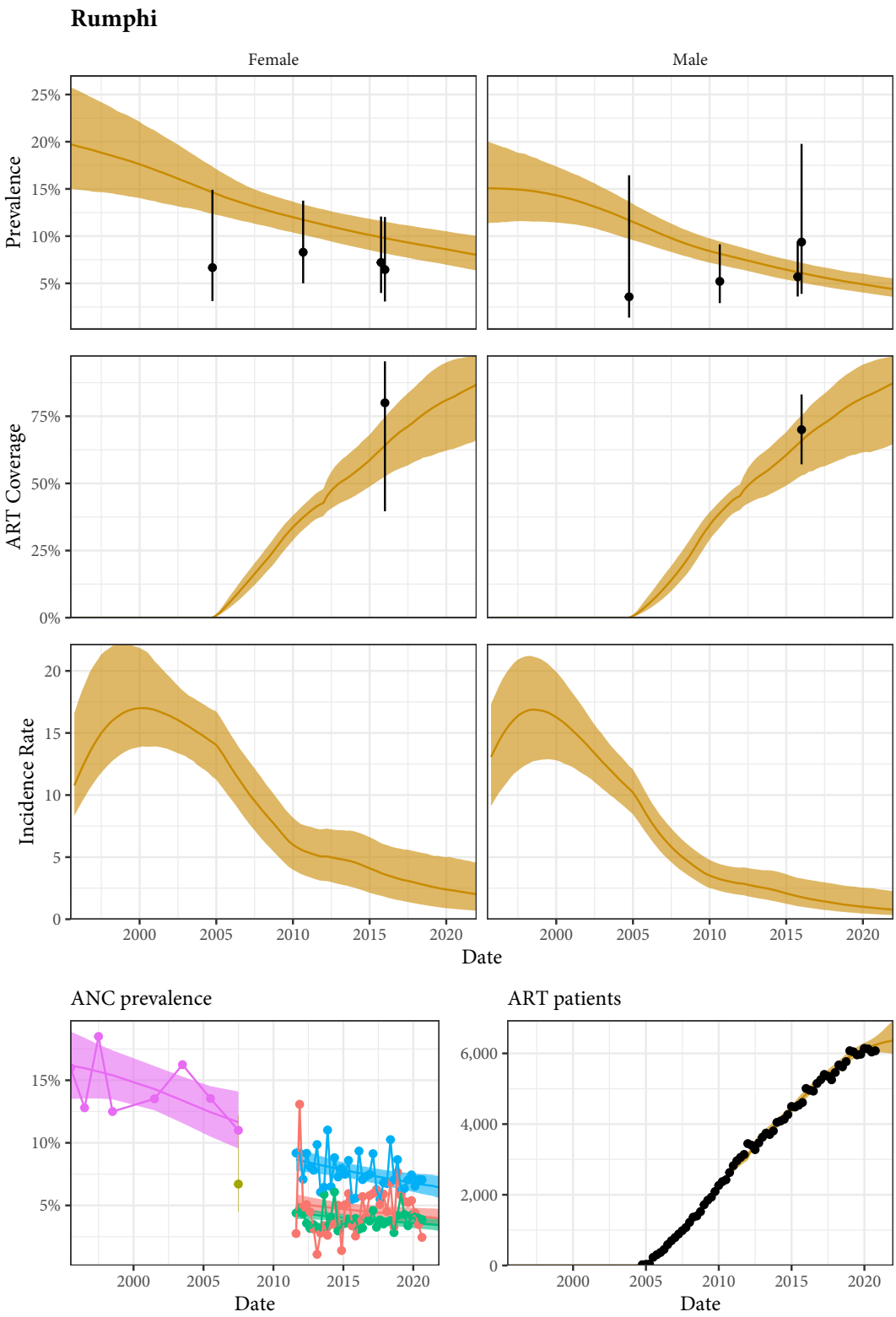

Figure A.6: Estimated prevalence, ART coverage, incidence, ANC prevalence, and ART patient counts in the Rumphi district of Malawi with household survey data, ANC facility data, and programmatic reporting data (points). Different colours on panel "ANC prevalence" indicate different ANC facilities.

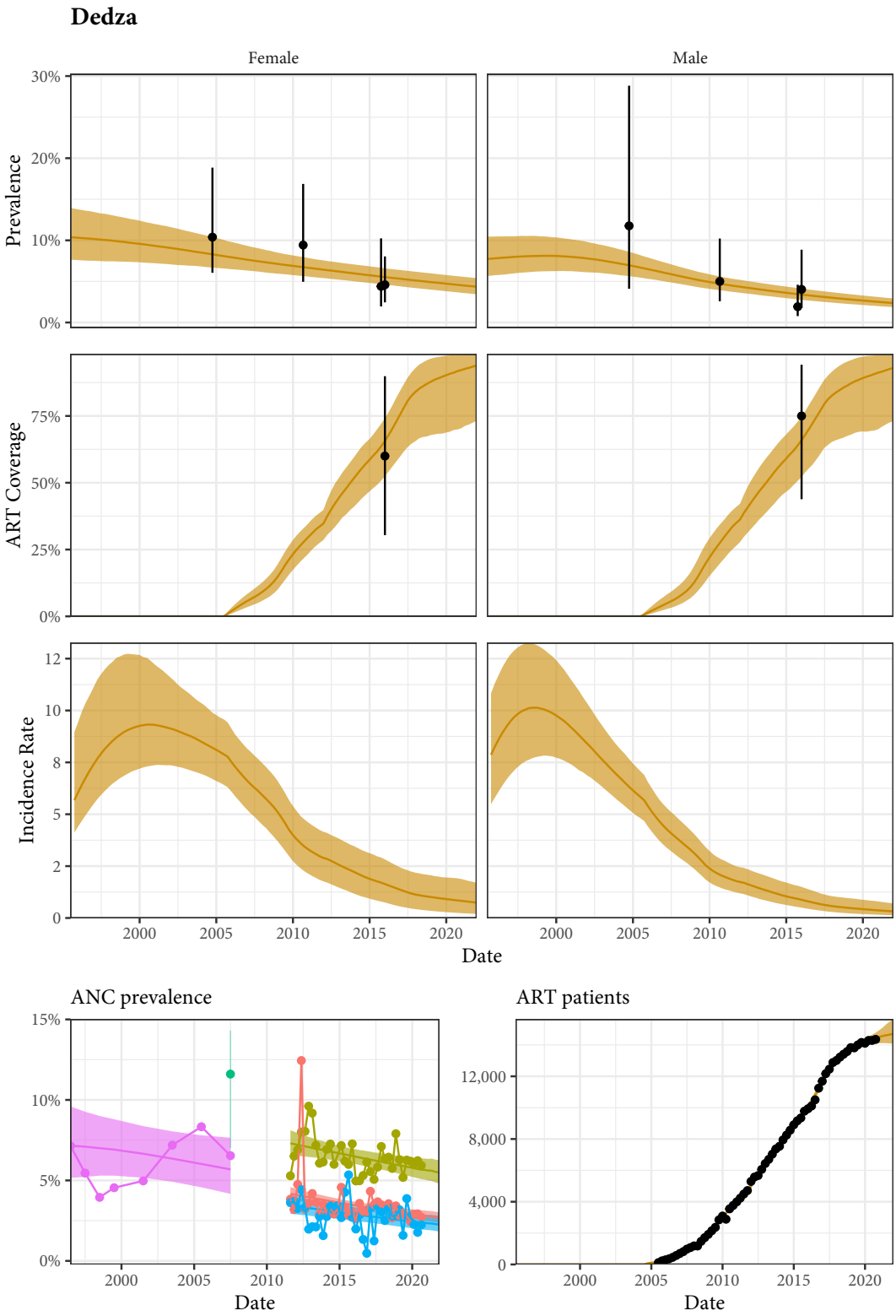

Figure A.7: Estimated prevalence, ART coverage, incidence, ANC prevalence, and ART patient counts in the Dedza district of Malawi with household survey data, ANC facility data, and programmatic reporting data (points). Different colours on panel "ANC prevalence" indicate different ANC facilities.

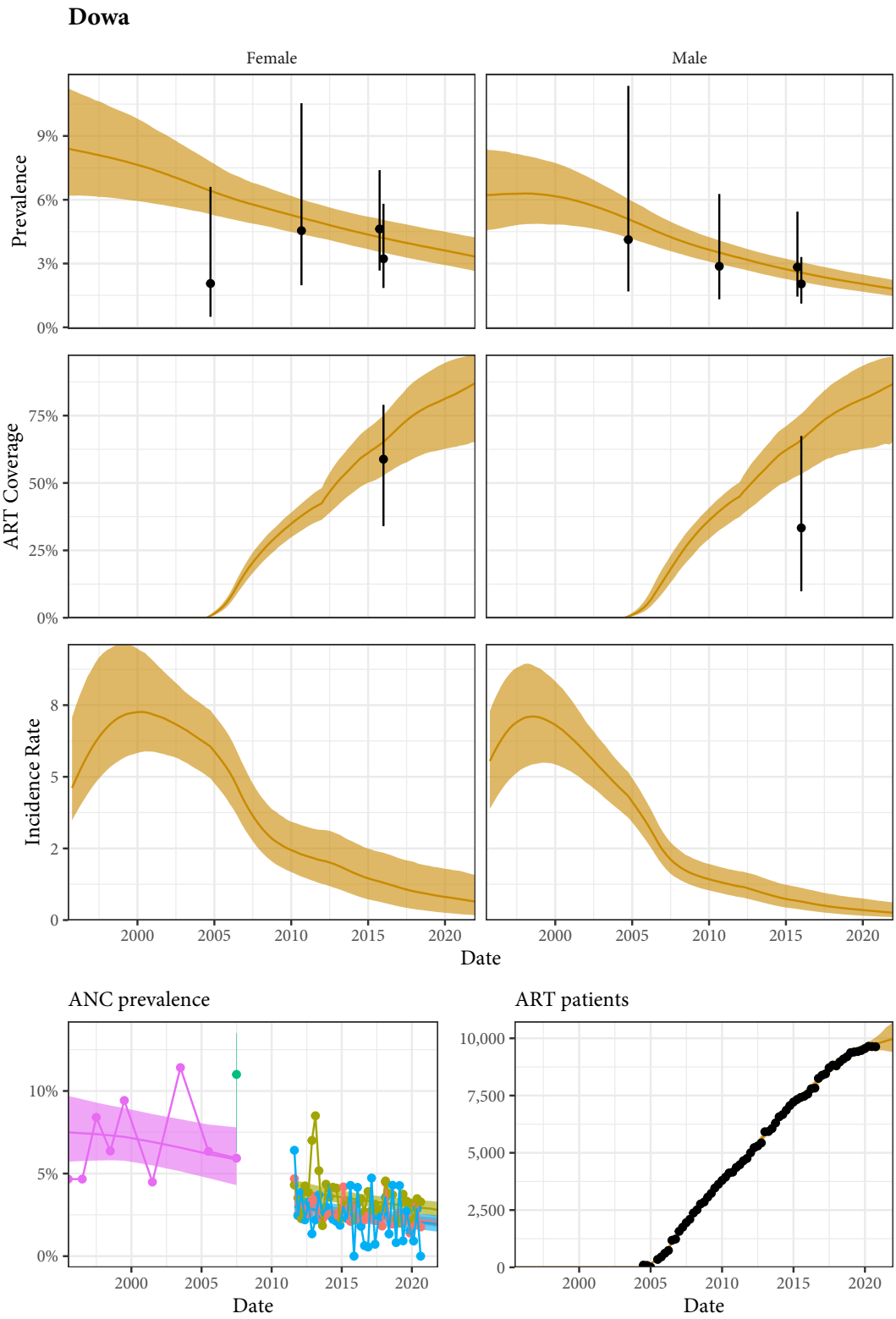

Figure A.8: Estimated prevalence, ART coverage, incidence, ANC prevalence, and ART patient counts in the Dowa district of Malawi with household survey data, ANC facility data, and programmatic reporting data (points). Different colours on panel "ANC prevalence" indicate different ANC facilities.

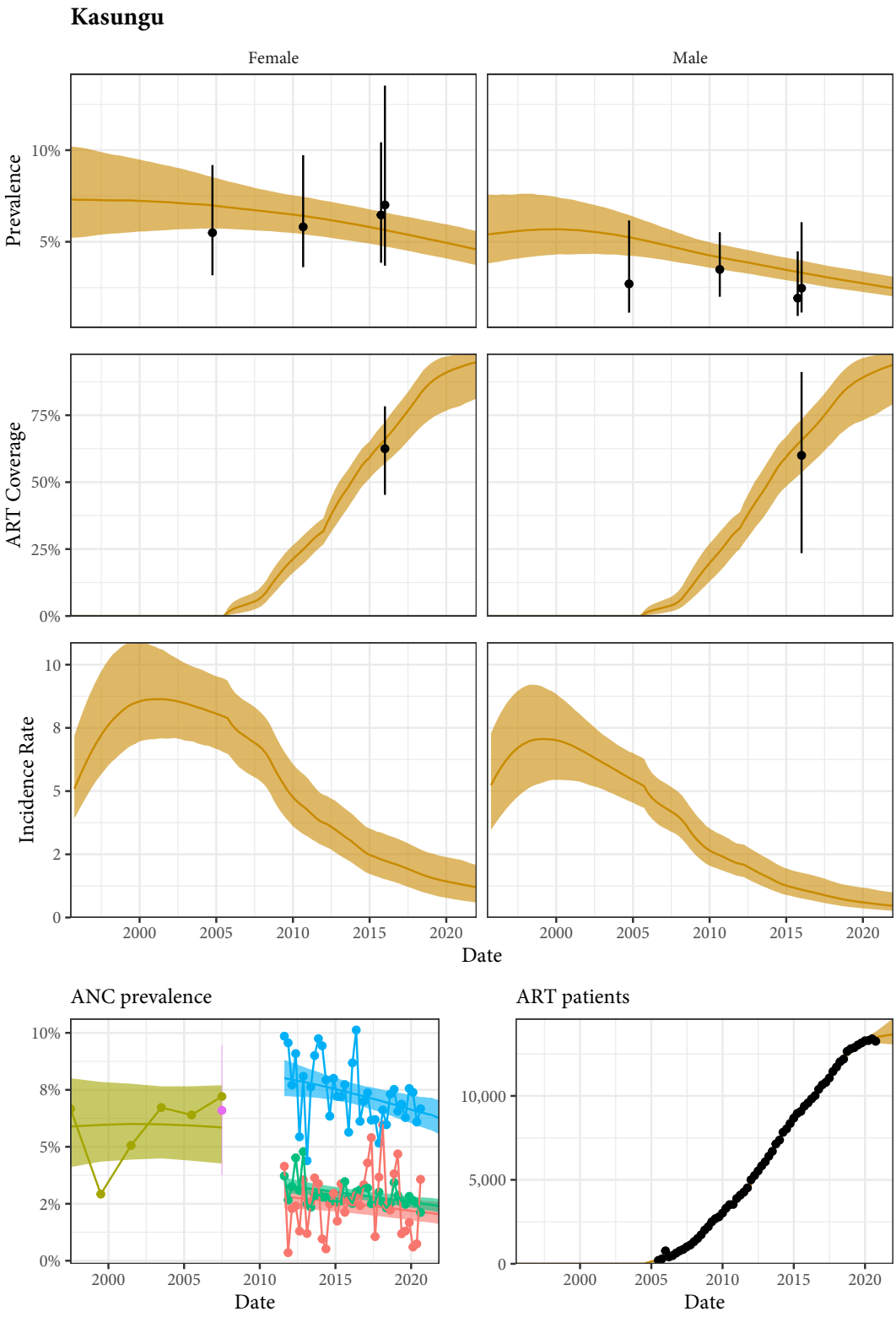

Figure A.9: Estimated prevalence, ART coverage, incidence, ANC prevalence, and ART patient counts in the Kasungu district of Malawi with household survey data, ANC facility data, and programmatic reporting data (points). Different colours on panel "ANC prevalence" indicate different ANC facilities.

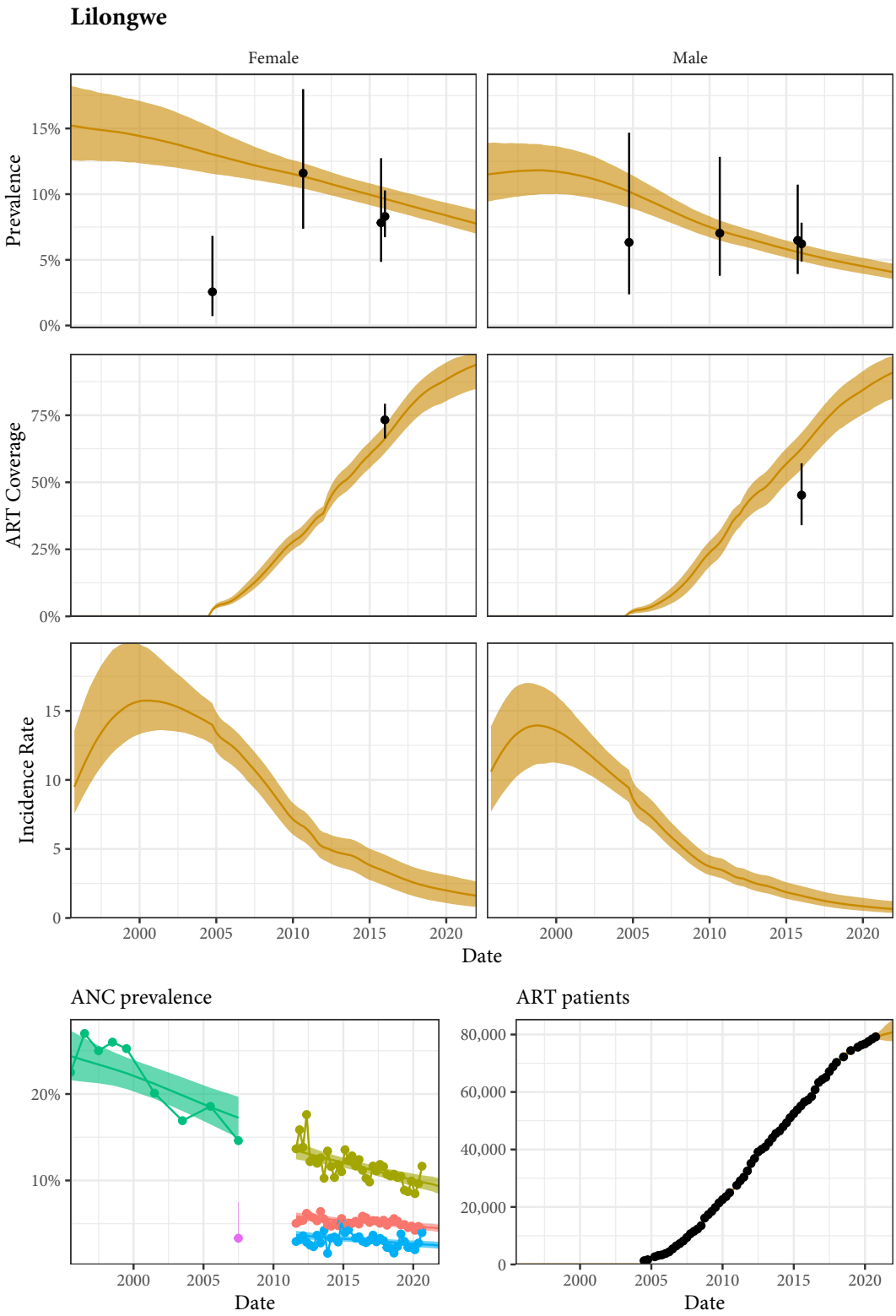

Figure A.10: Estimated prevalence, ART coverage, incidence, ANC prevalence, and ART patient counts in the Lilongwe district of Malawi with household survey data, ANC facility data, and programmatic reporting data (points). Different colours on panel "ANC prevalence" indicate different ANC facilities.

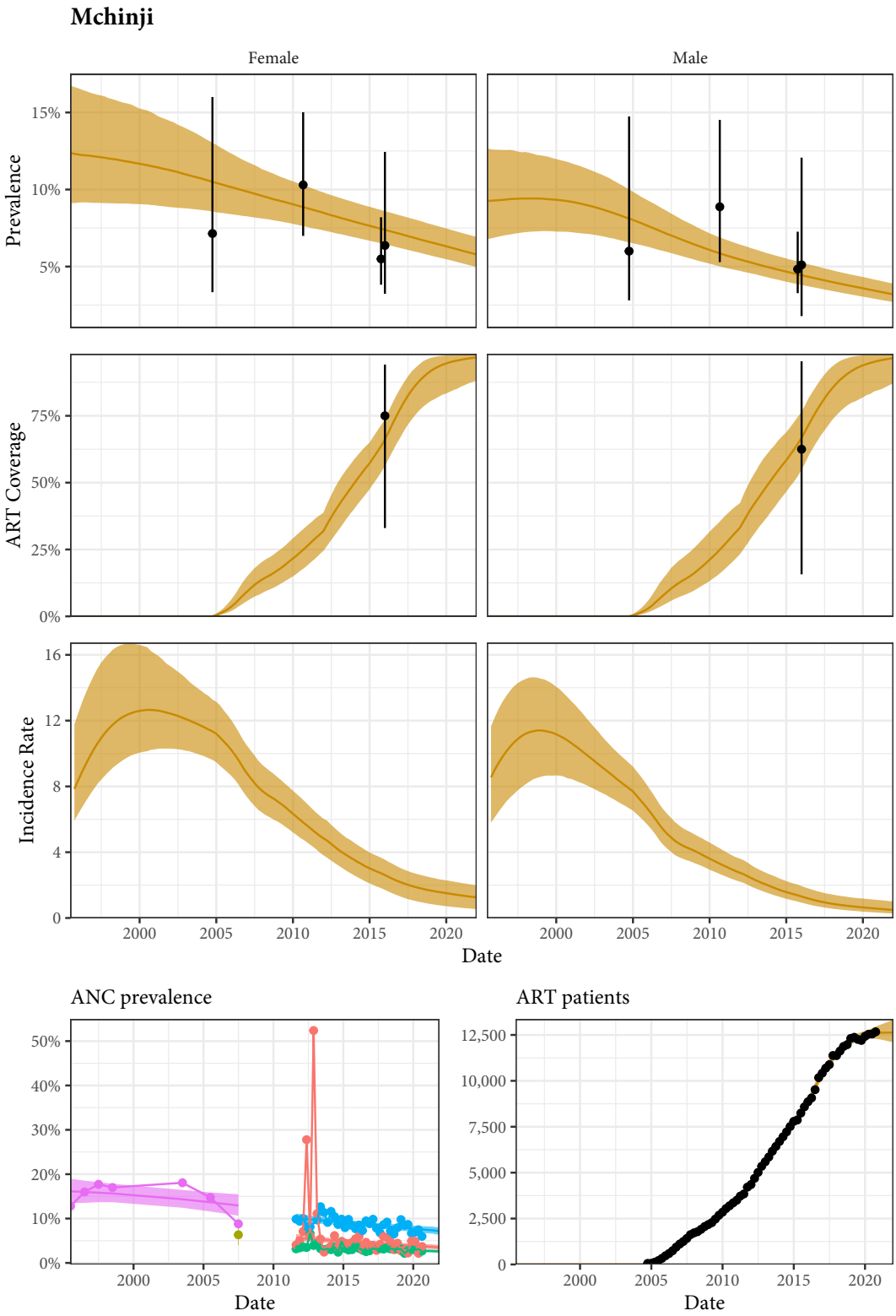

Figure A.11: Estimated prevalence, ART coverage, incidence, ANC prevalence, and ART patient counts in the Mchinji district of Malawi with household survey data, ANC facility data, and programmatic reporting data (points). Different colours on panel "ANC prevalence" indicate different ANC facilities.

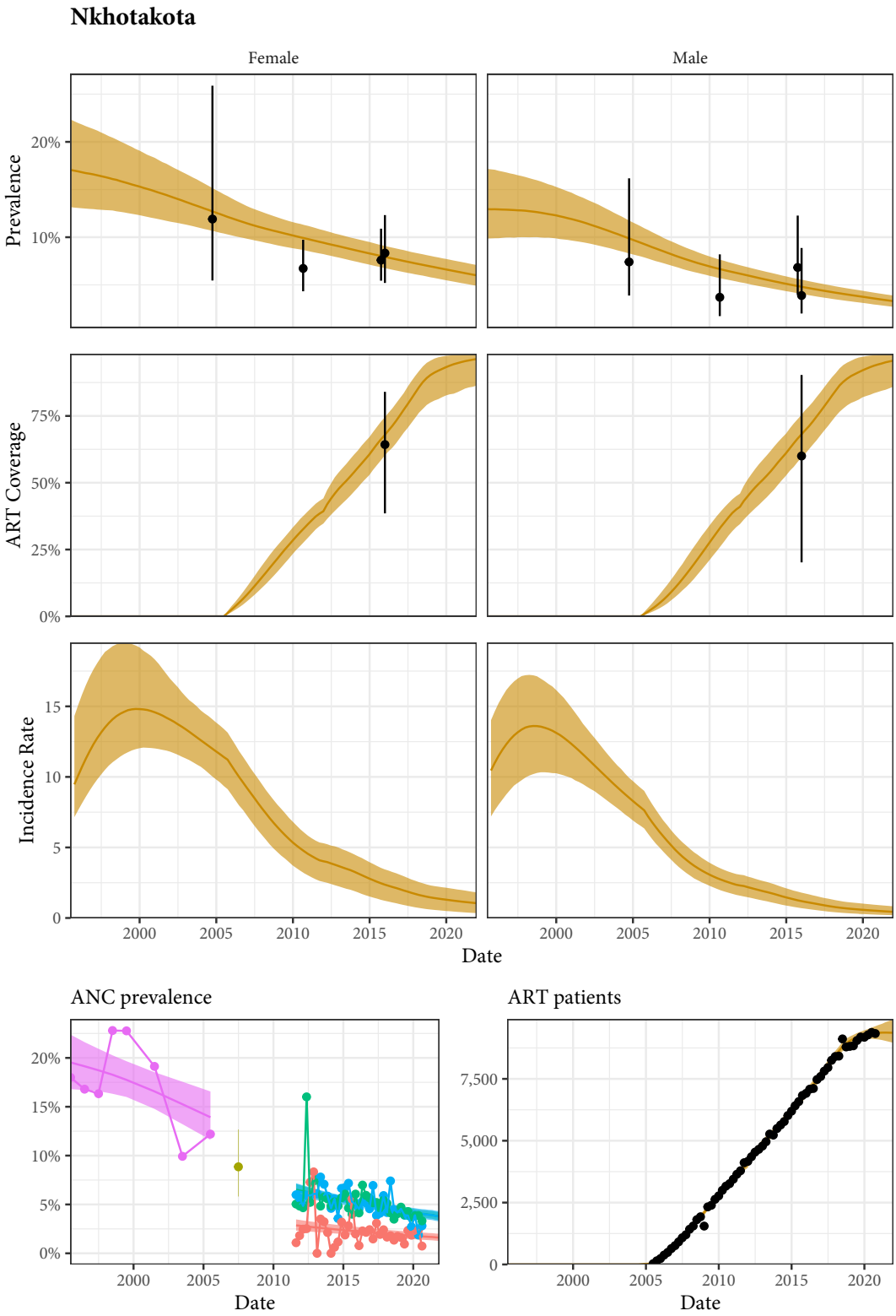

Figure A.12: Estimated prevalence, ART coverage, incidence, ANC prevalence, and ART patient counts in the Nkhotakota district of Malawi with household survey data, ANC facility data, and programmatic reporting data (points). Different colours on panel "ANC prevalence" indicate different ANC facilities.

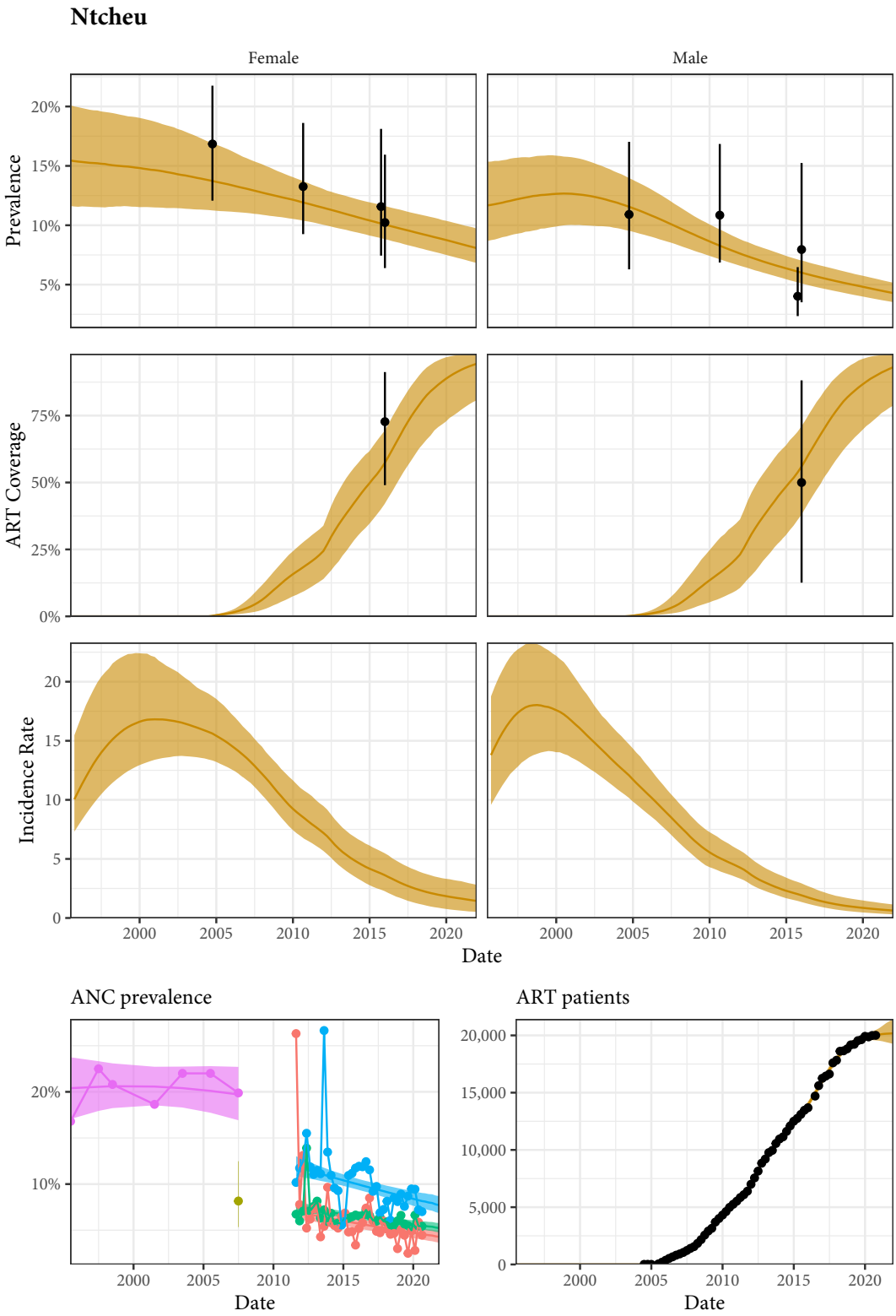

Figure A.13: Estimated prevalence, ART coverage, incidence, ANC prevalence, and ART patient counts in the Ntcheu district of Malawi with household survey data, ANC facility data, and programmatic reporting data (points). Different colours on panel "ANC prevalence" indicate different ANC facilities.

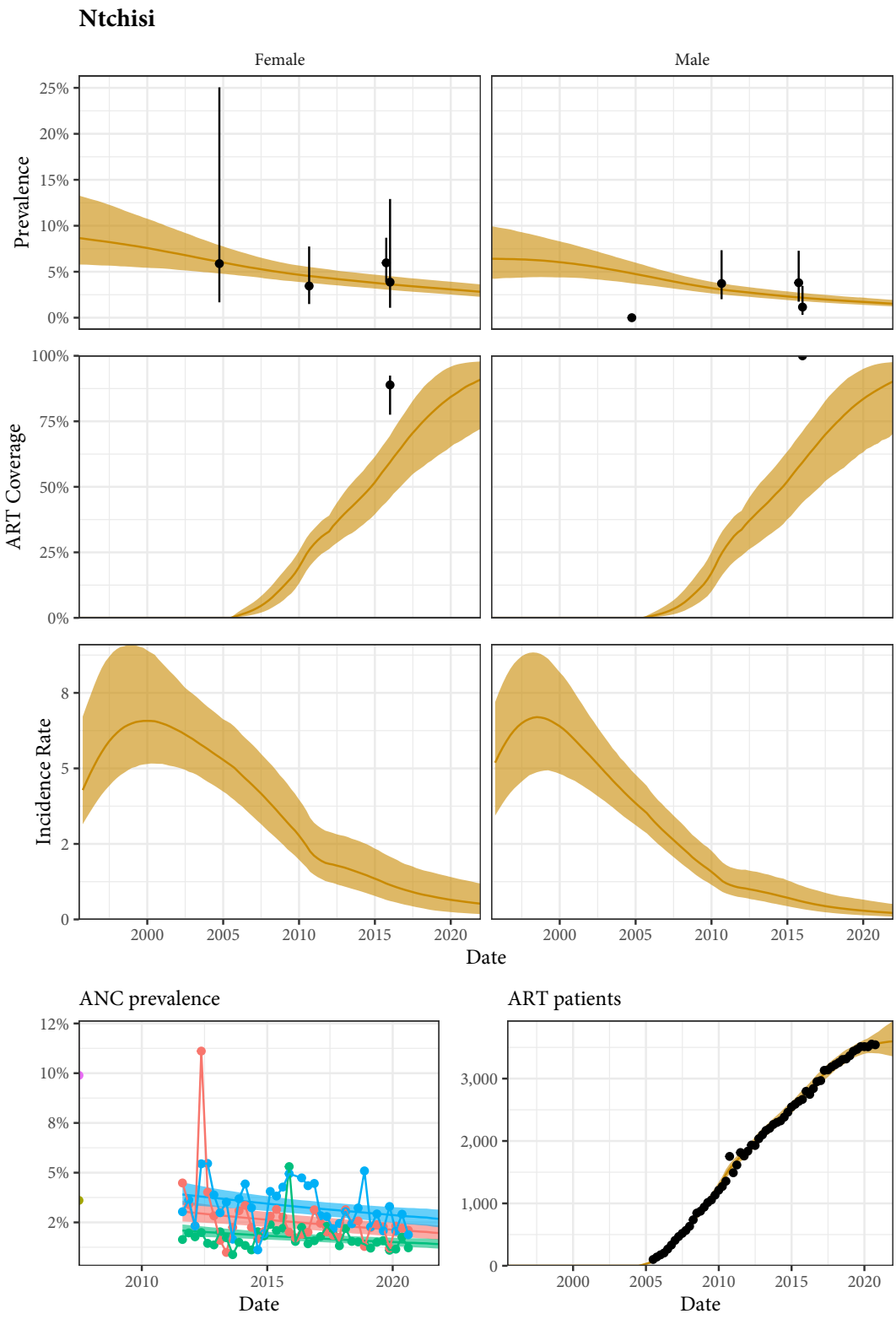

Figure A.14: Estimated prevalence, ART coverage, incidence, ANC prevalence, and ART patient counts in the Ntchisi district of Malawi with household survey data, ANC facility data, and programmatic reporting data (points). Different colours on panel "ANC prevalence" indicate different ANC facilities.

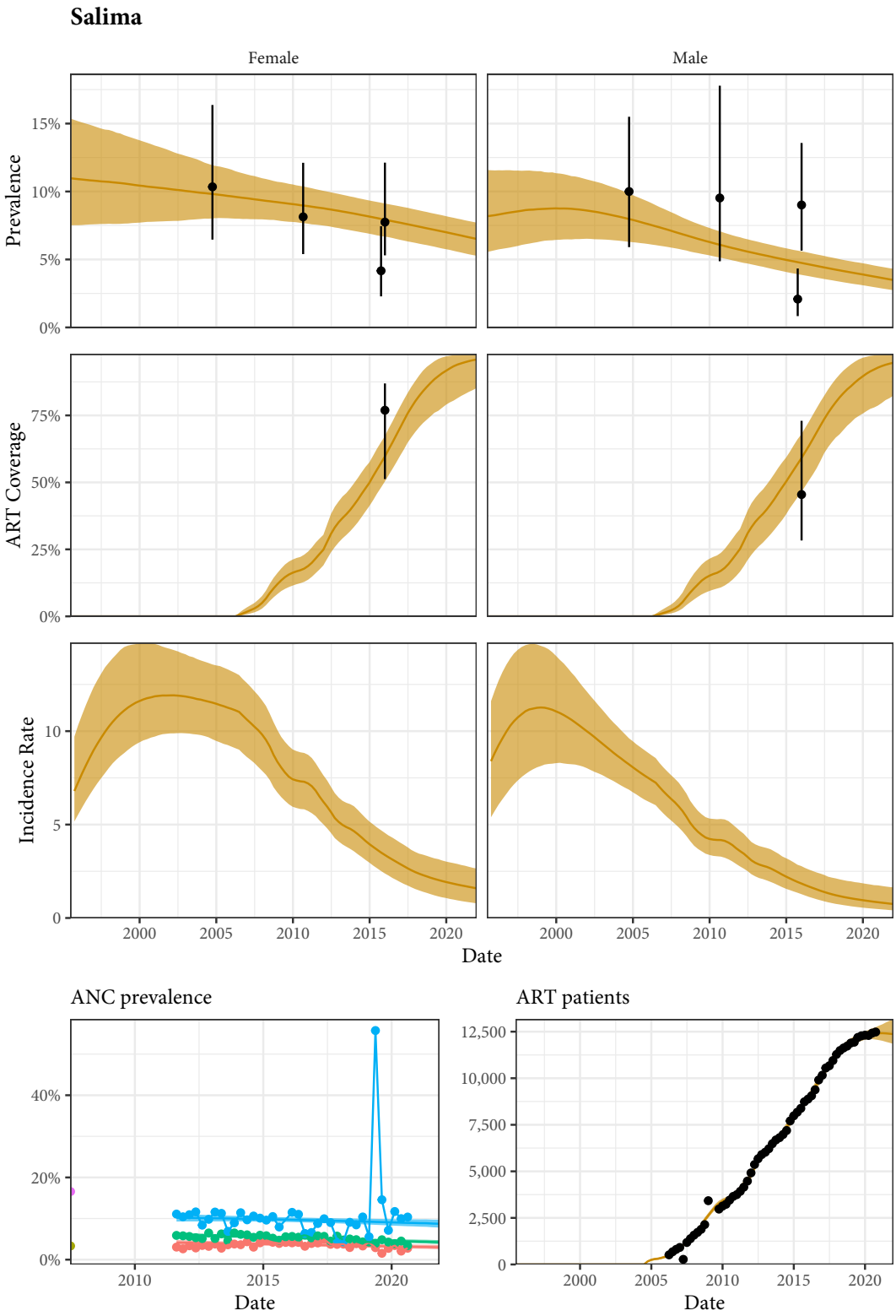

Figure A.15: Estimated prevalence, ART coverage, incidence, ANC prevalence, and ART patient counts in the Salima district of Malawi with household survey data, ANC facility data, and programmatic reporting data (points). Different colours on panel "ANC prevalence" indicate different ANC facilities.

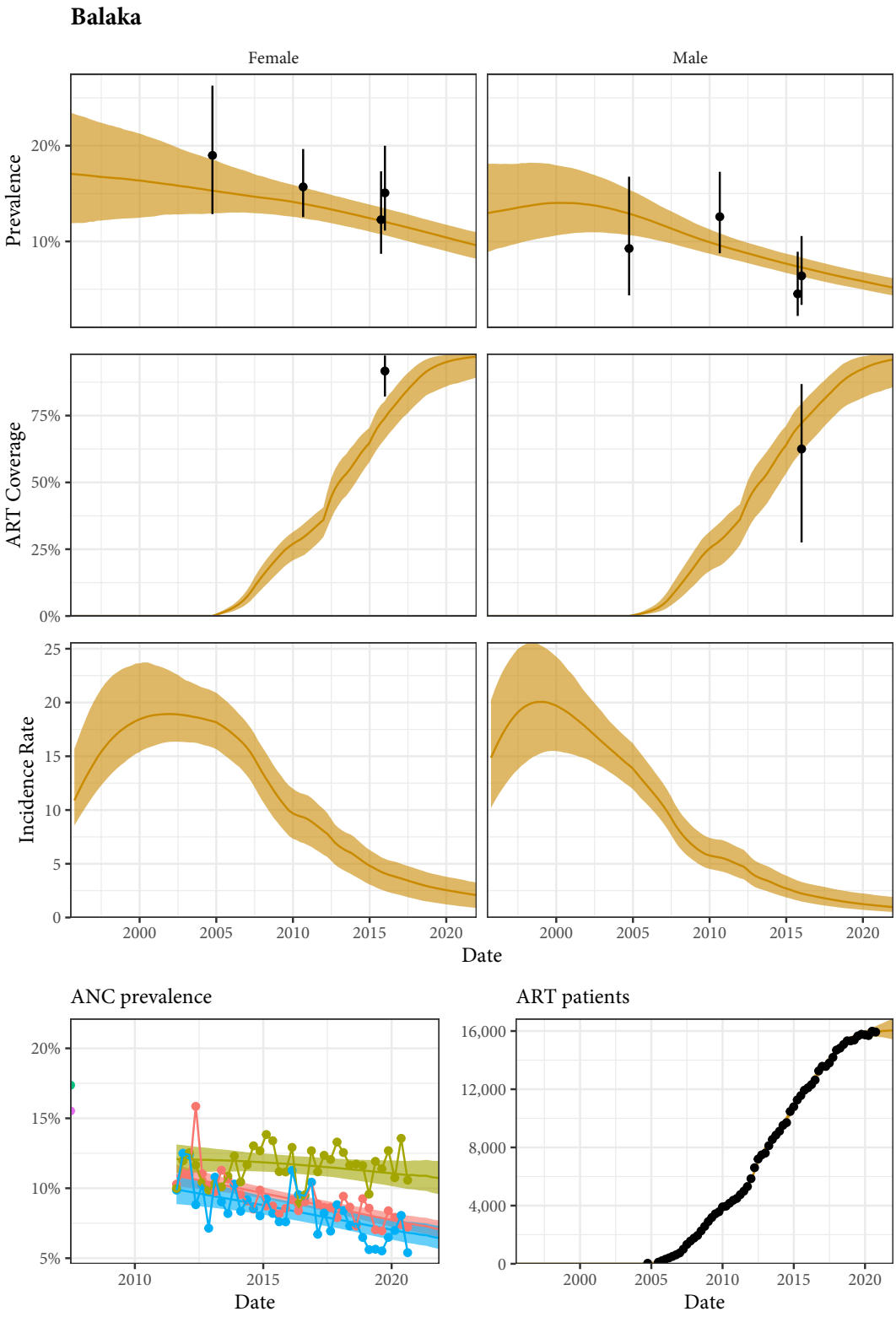

Figure A.16: Estimated prevalence, ART coverage, incidence, ANC prevalence, and ART patient counts in the Balaka district of Malawi with household survey data, ANC facility data, and programmatic reporting data (points). Different colours on panel "ANC prevalence" indicate different ANC facilities.

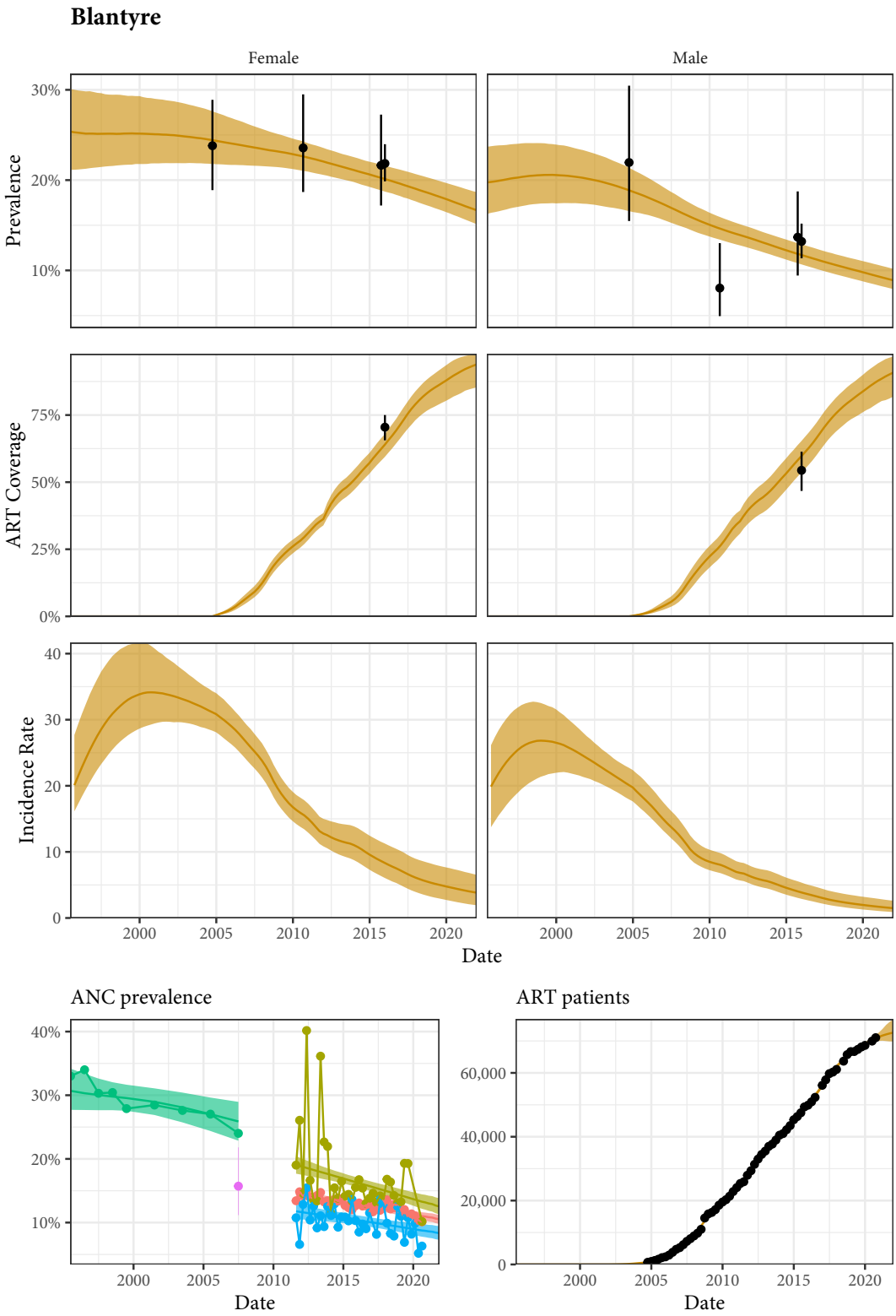

Figure A.17: Estimated prevalence, ART coverage, incidence, ANC prevalence, and ART patient counts in the Blantyre district of Malawi with household survey data, ANC facility data, and programmatic reporting data (points). Different colours on panel "ANC prevalence" indicate different ANC facilities.

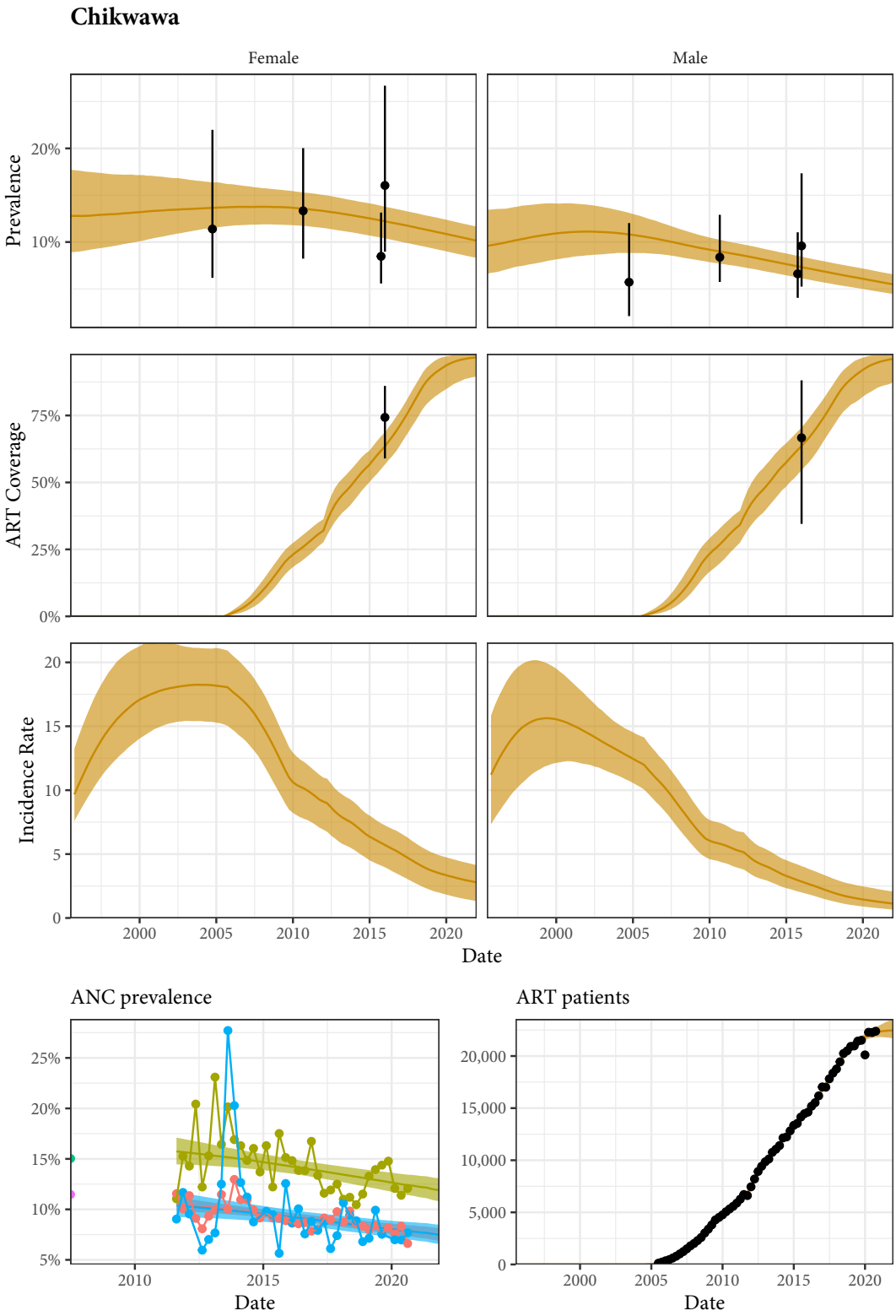

Figure A.18: Estimated prevalence, ART coverage, incidence, ANC prevalence, and ART patient counts in the Chikwawa district of Malawi with household survey data, ANC facility data, and programmatic reporting data (points). Different colours on panel "ANC prevalence" indicate different ANC facilities.

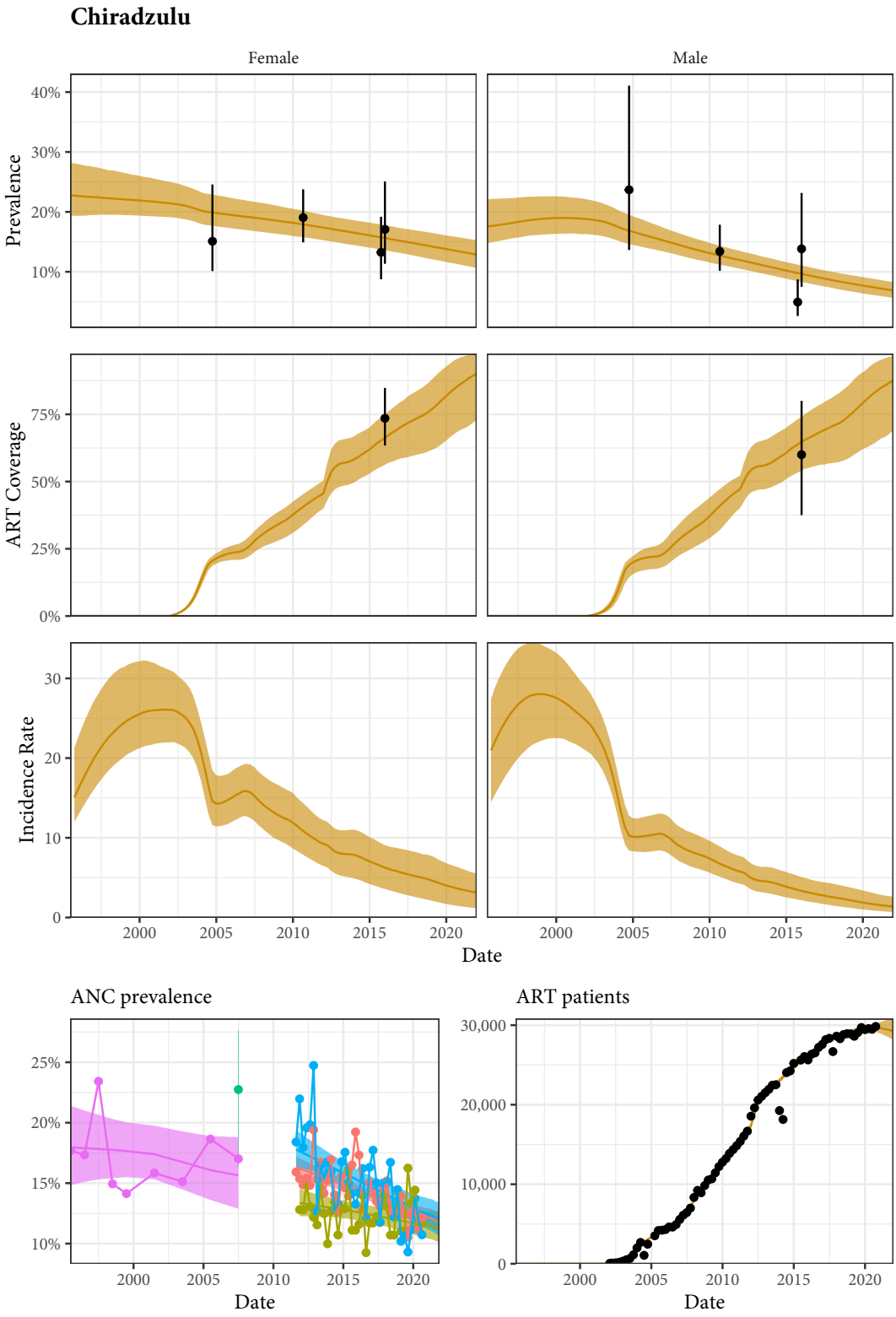

Figure A.19: Estimated prevalence, ART coverage, incidence, ANC prevalence, and ART patient counts in the Chiradzulu district of Malawi with household survey data, ANC facility data, and programmatic reporting data (points). Different colours on panel "ANC prevalence" indicate different ANC facilities.

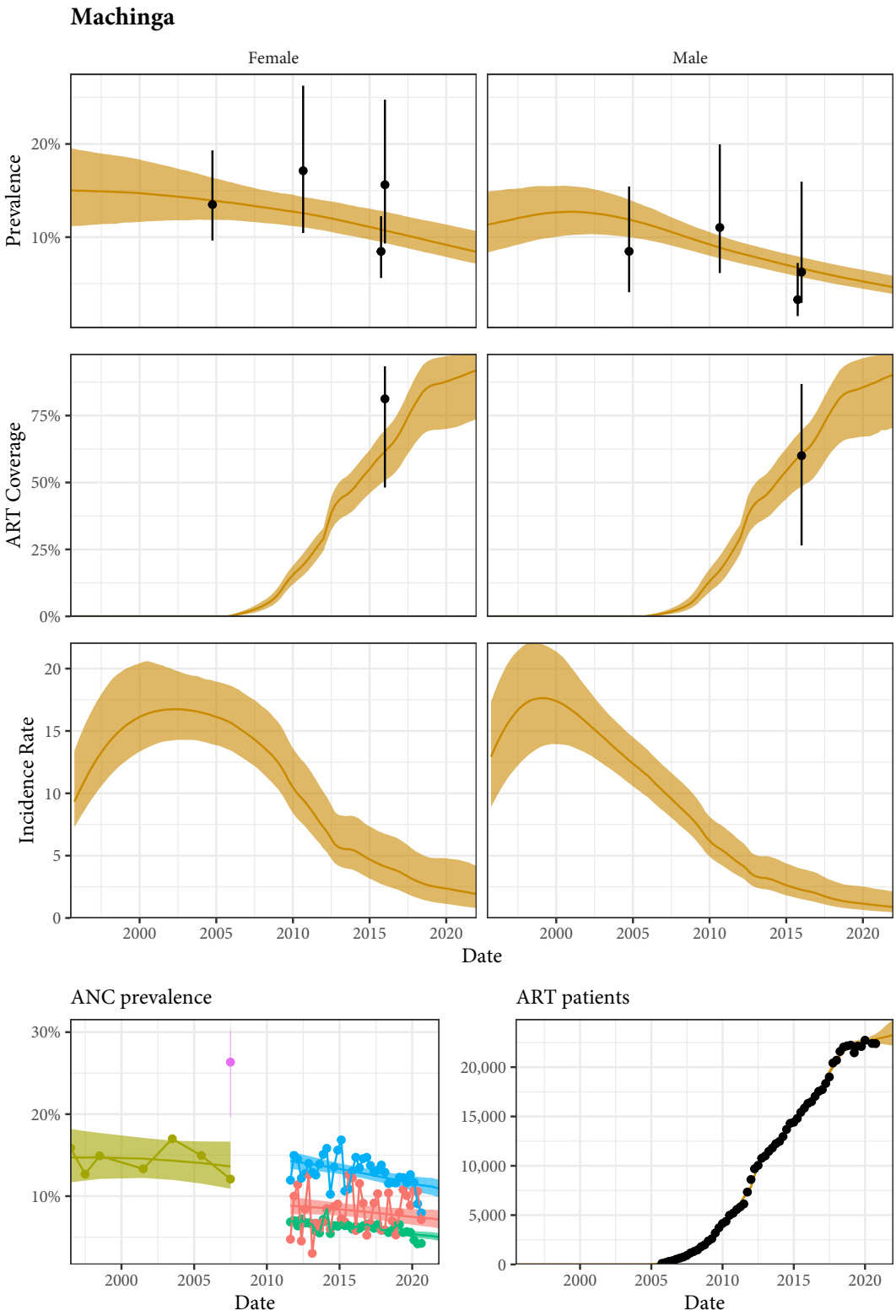

Figure A.20: Estimated prevalence, ART coverage, incidence, ANC prevalence, and ART patient counts in the Machinga district of Malawi with household survey data, ANC facility data, and programmatic reporting data (points). Different colours on panel "ANC prevalence" indicate different ANC facilities.

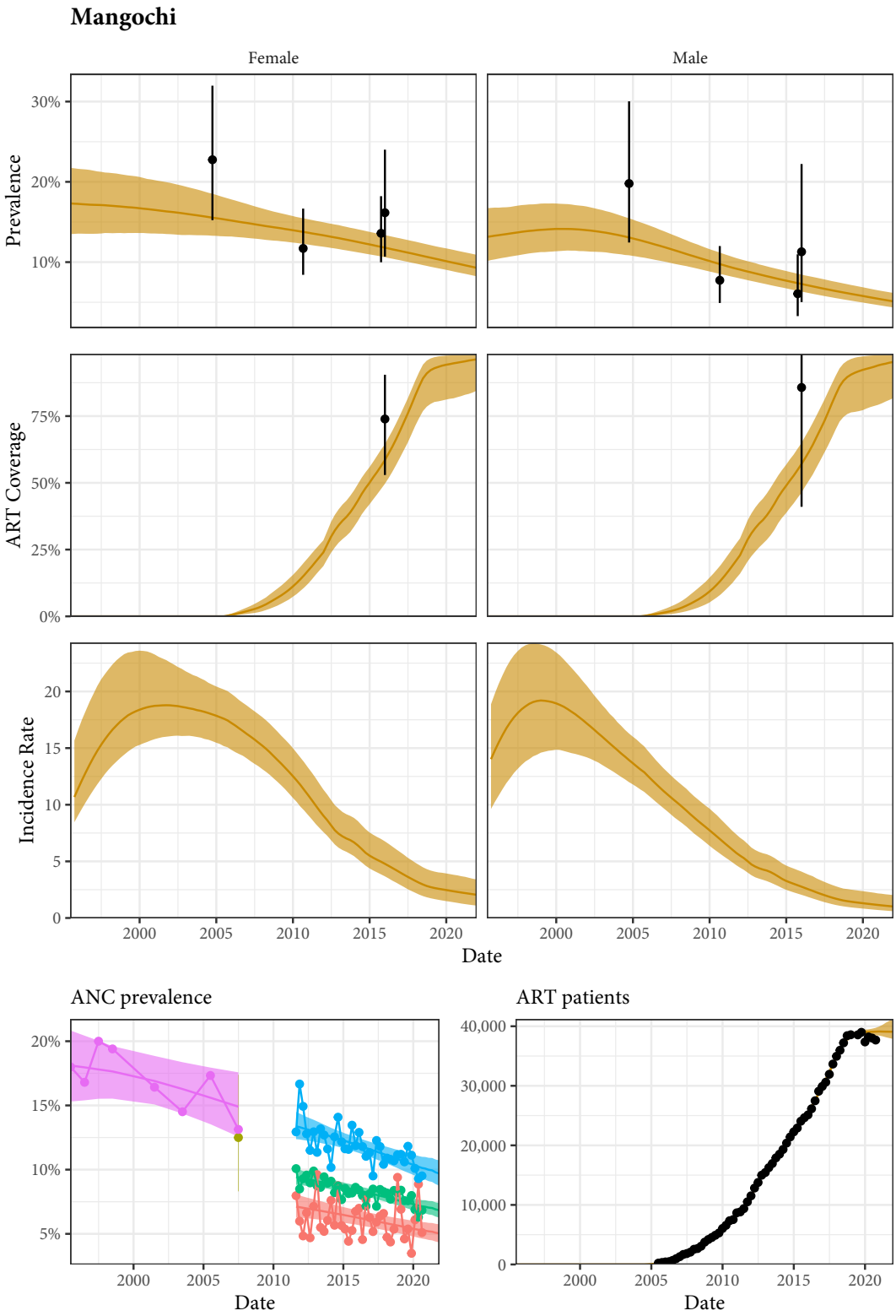

Figure A.21: Estimated prevalence, ART coverage, incidence, ANC prevalence, and ART patient counts in the Mangochi district of Malawi with household survey data, ANC facility data, and programmatic reporting data (points). Different colours on panel "ANC prevalence" indicate different ANC facilities.

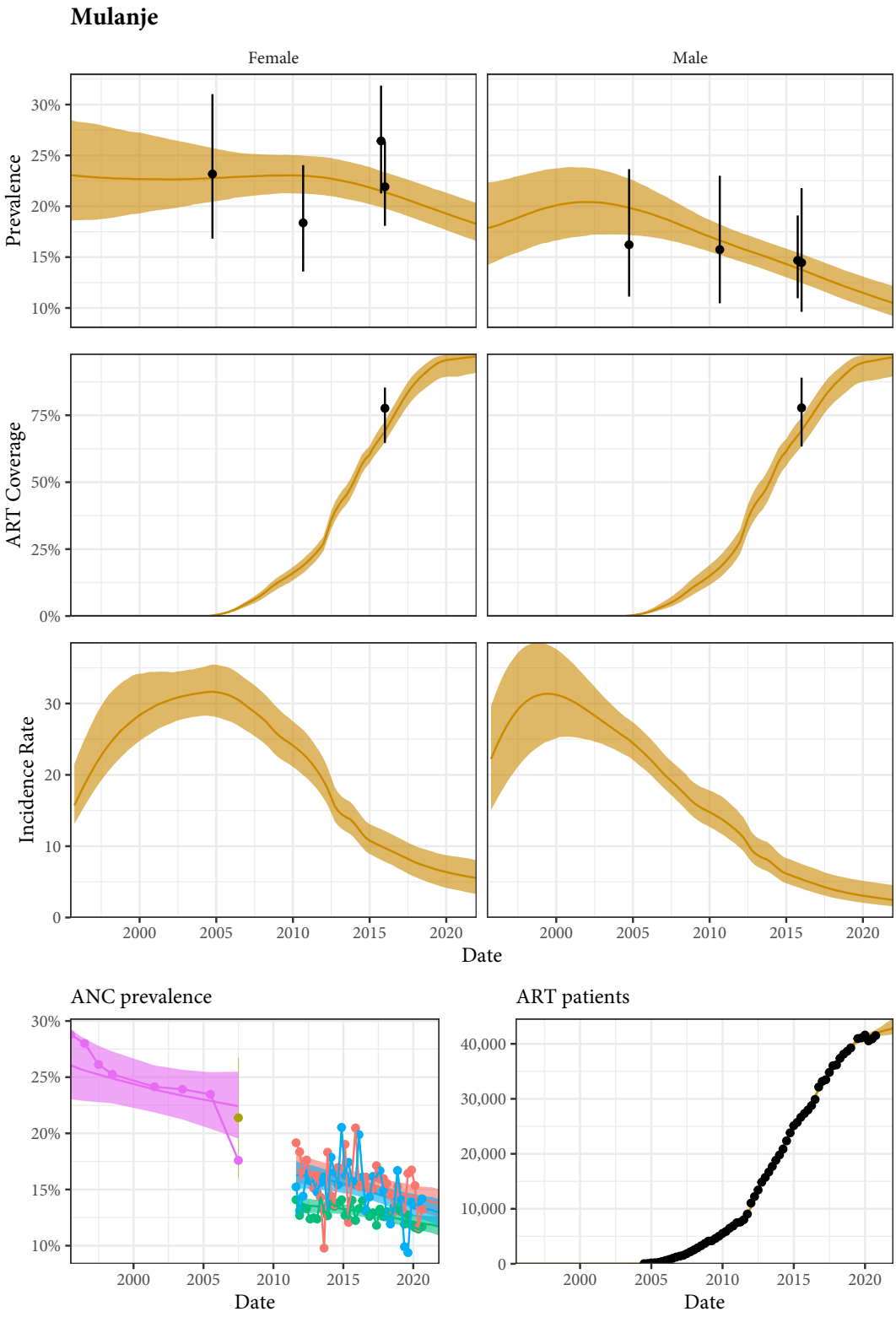

Figure A.22: Estimated prevalence, ART coverage, incidence, ANC prevalence, and ART patient counts in the Mulanje district of Malawi with household survey data, ANC facility data, and programmatic reporting data (points). Different colours on panel "ANC prevalence" indicate different ANC facilities.

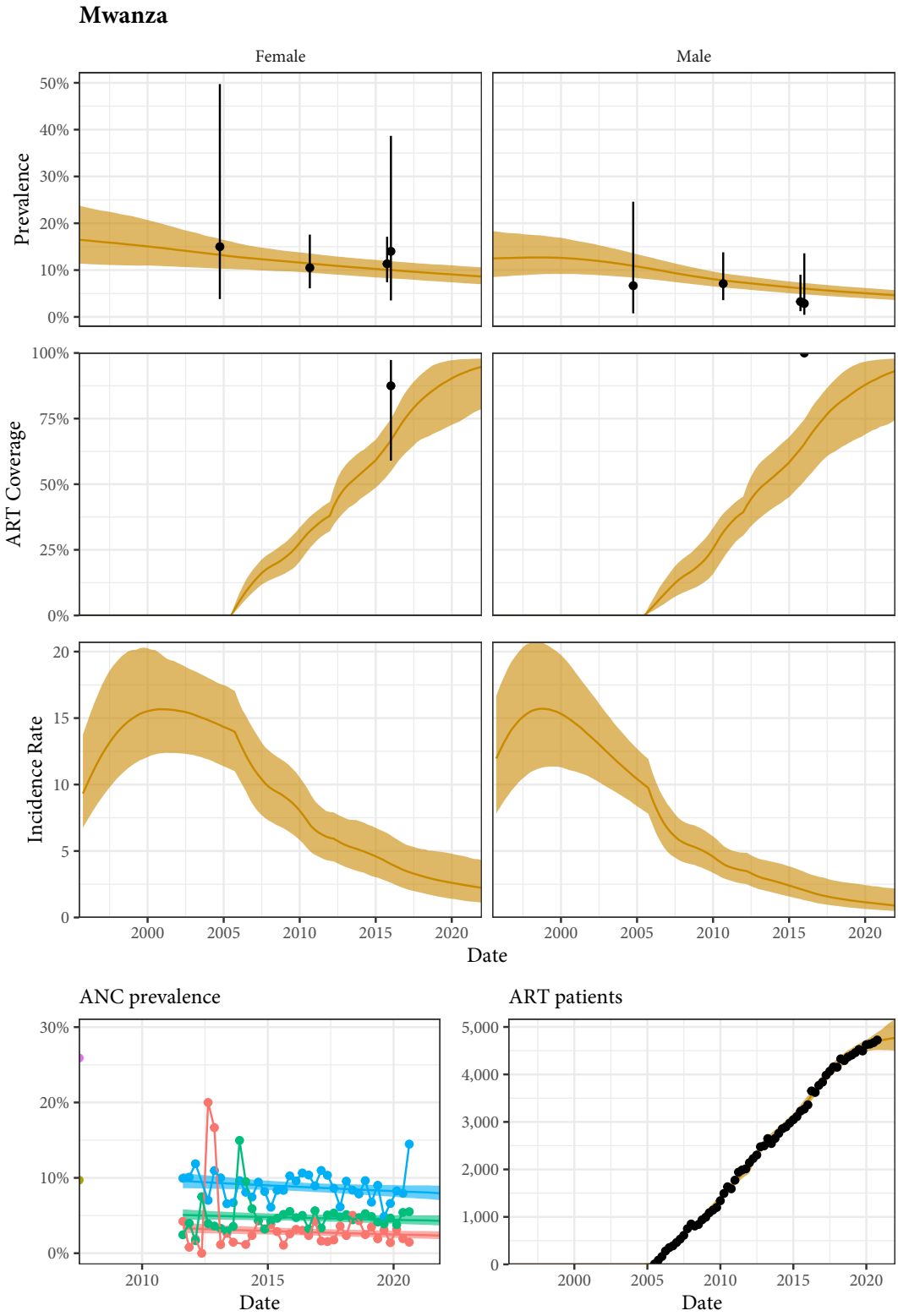

Figure A.23: Estimated prevalence, ART coverage, incidence, ANC prevalence, and ART patient counts in the Mwanza district of Malawi with household survey data, ANC facility data, and programmatic reporting data (points). Different colours on panel "ANC prevalence" indicate different ANC facilities.

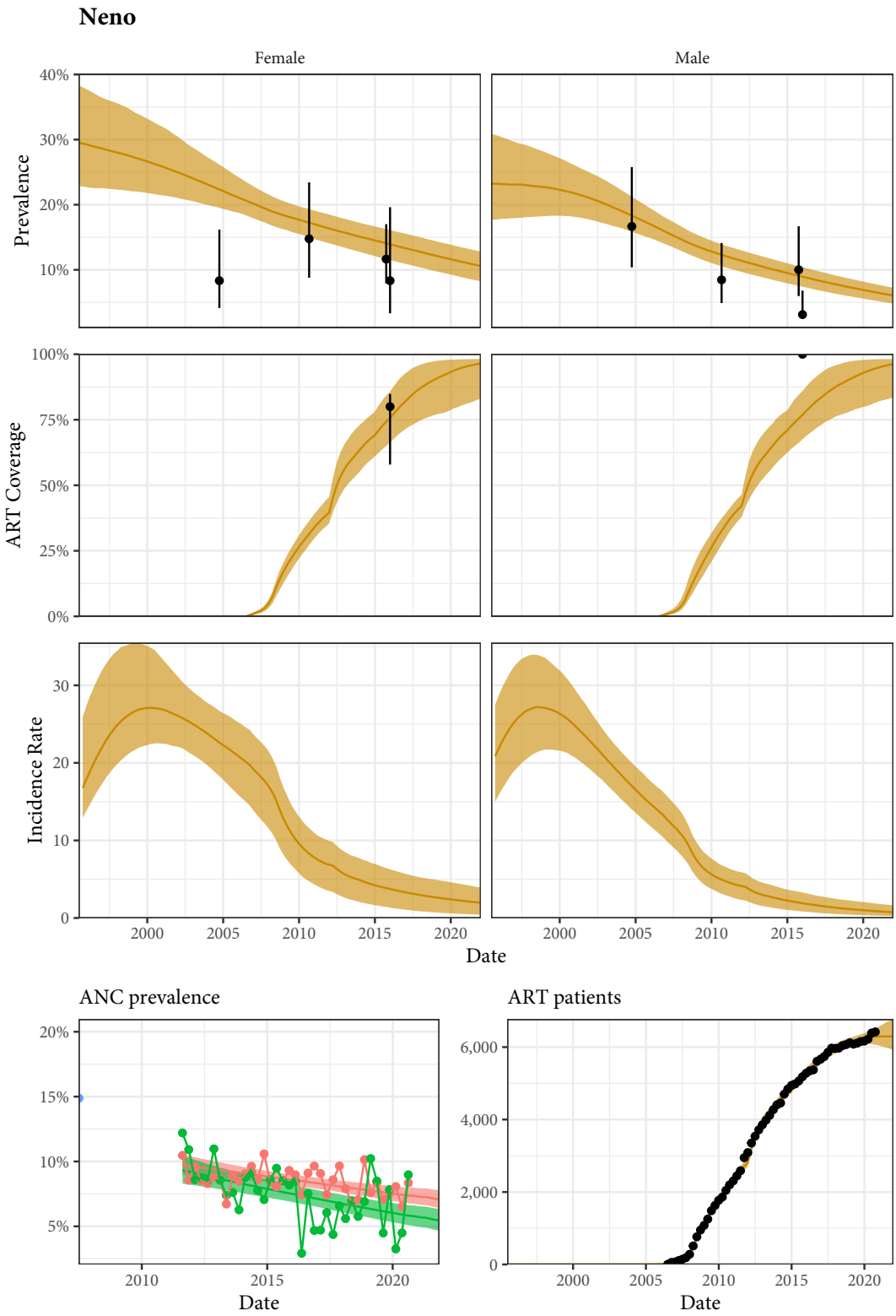

Figure A.24: Estimated prevalence, ART coverage, incidence, ANC prevalence, and ART patient counts in the Neno district of Malawi with household survey data, ANC facility data, and programmatic reporting data (points). Different colours on panel "ANC prevalence" indicate different ANC facilities.

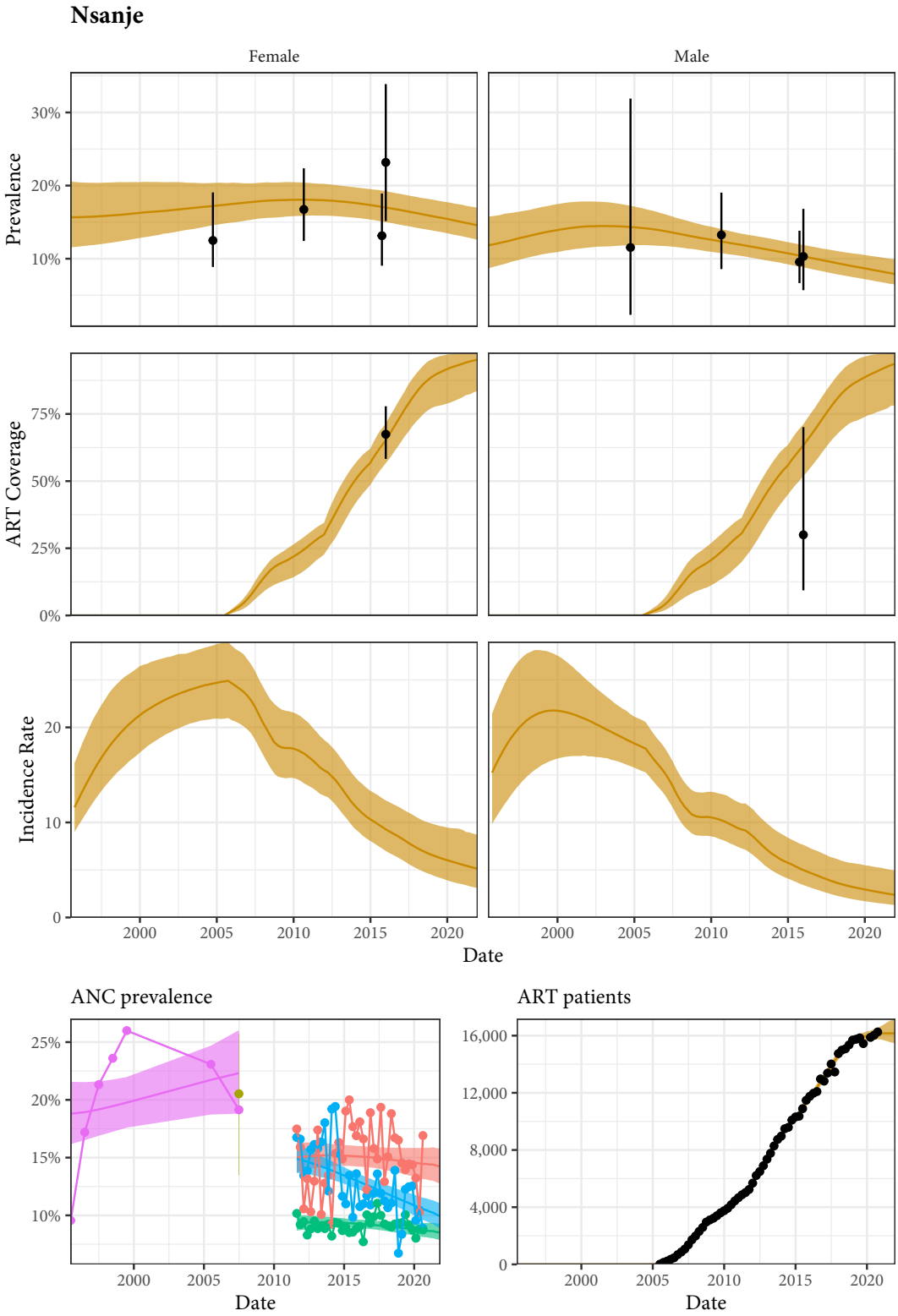

Figure A.25: Estimated prevalence, ART coverage, incidence, ANC prevalence, and ART patient counts in the Nsanje district of Malawi with household survey data, ANC facility data, and programmatic reporting data (points). Different colours on panel "ANC prevalence" indicate different ANC facilities.

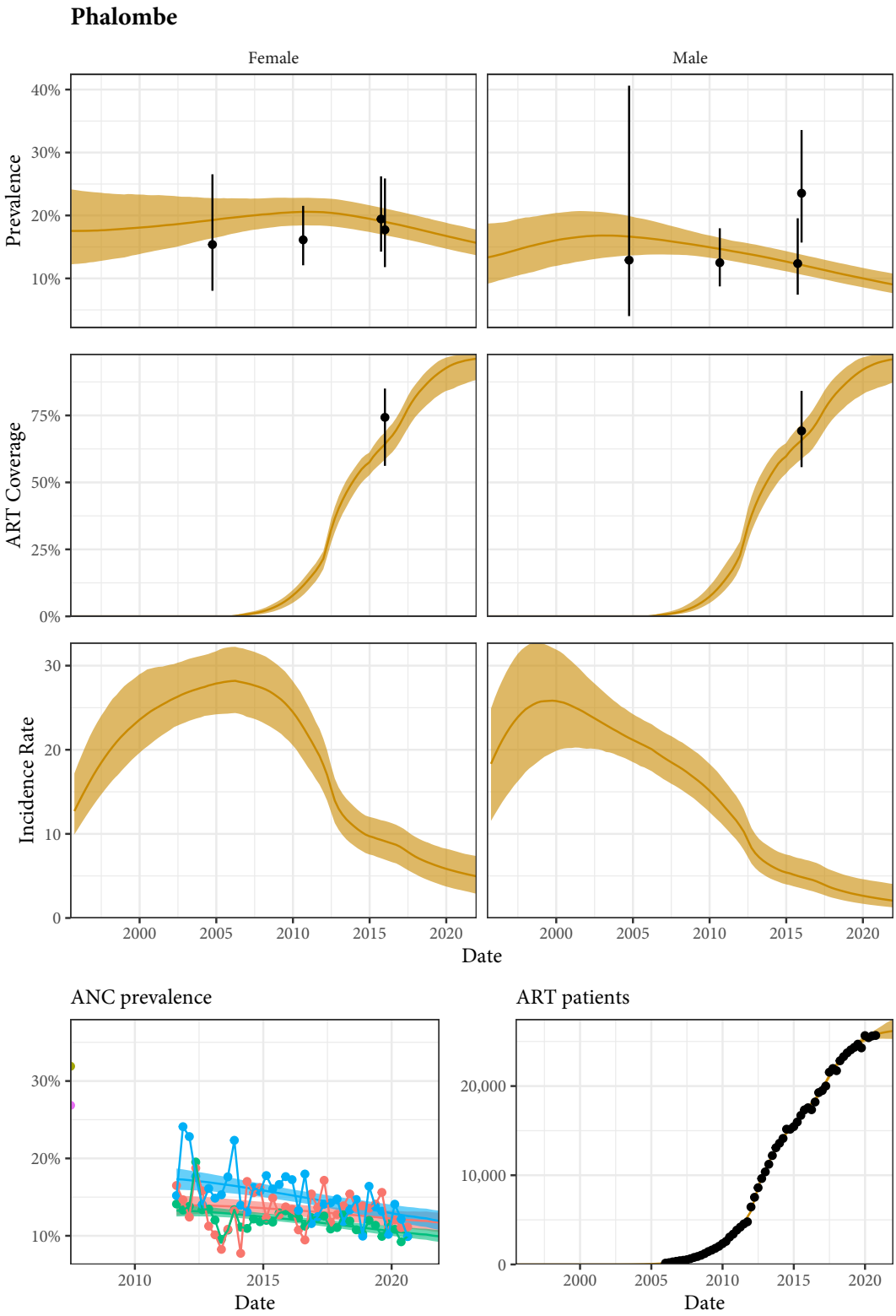

Figure A.26: Estimated prevalence, ART coverage, incidence, ANC prevalence, and ART patient counts in the Phalombe district of Malawi with household survey data, ANC facility data, and programmatic reporting data (points). Different colours on panel "ANC prevalence" indicate different ANC facilities.

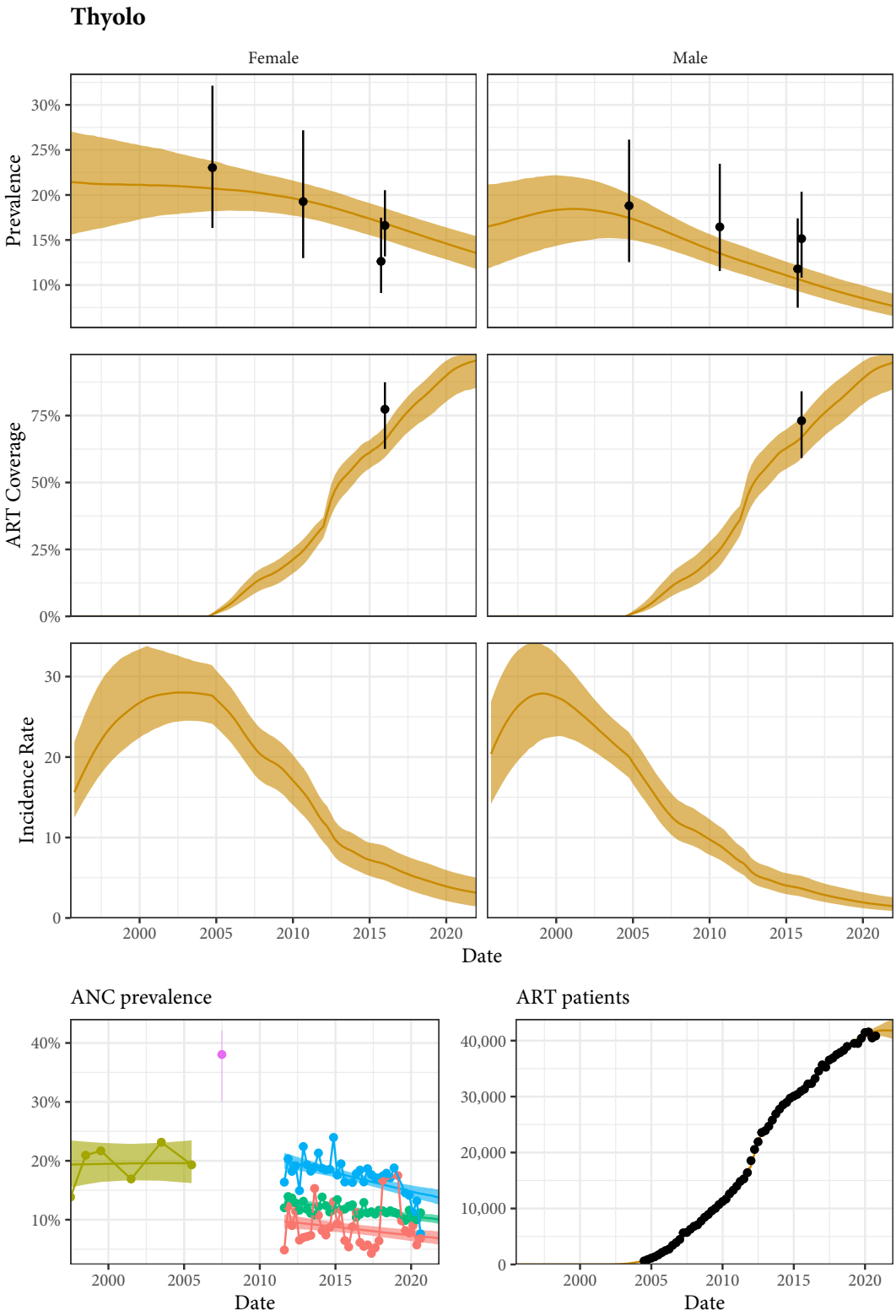

Figure A.27: Estimated prevalence, ART coverage, incidence, ANC prevalence, and ART patient counts in the Thyolo district of Malawi with household survey data, ANC facility data, and programmatic reporting data (points). Different colours on panel "ANC prevalence" indicate different ANC facilities.

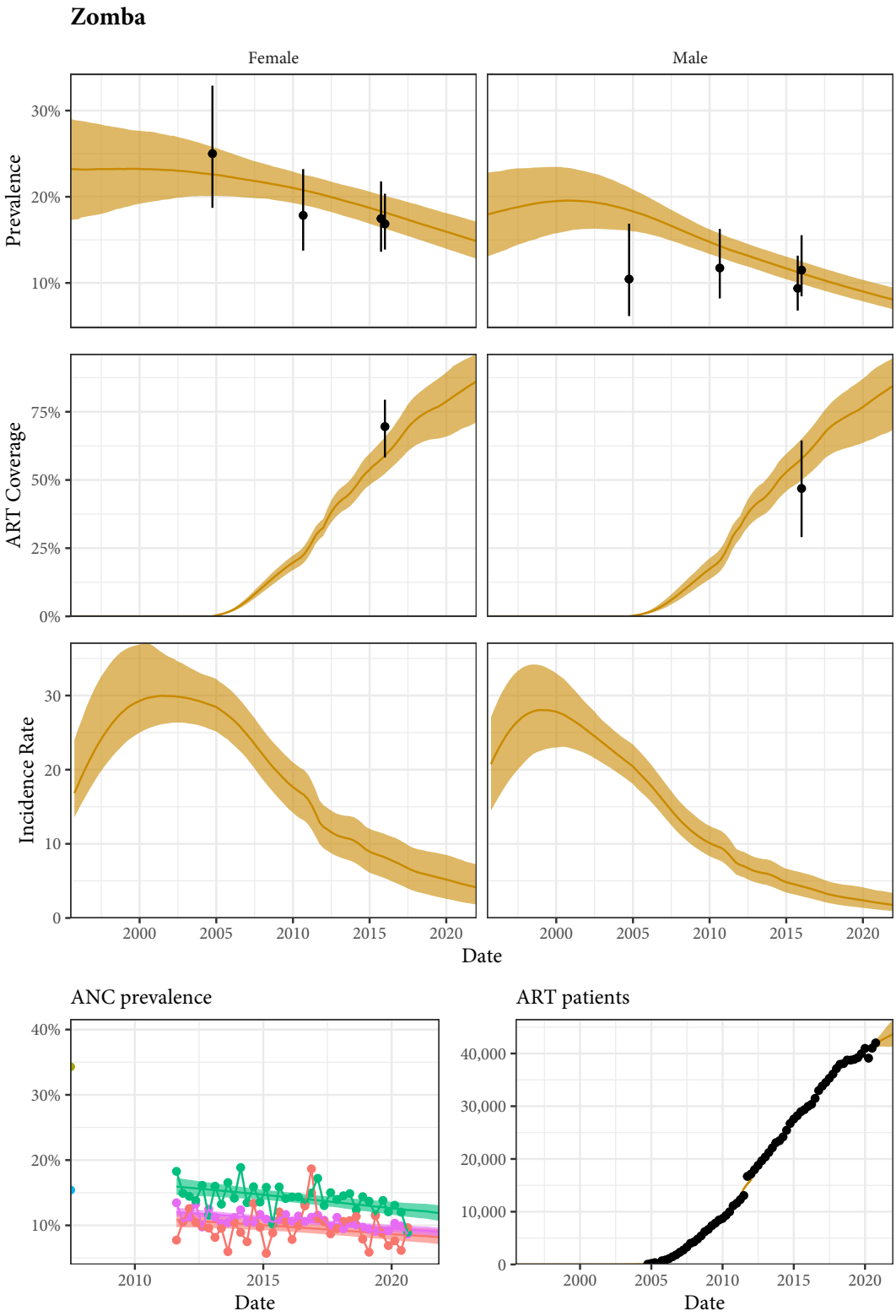

Figure A.28: Estimated prevalence, ART coverage, incidence, ANC prevalence, and ART patient counts in the Zomba district of Malawi with household survey data, ANC facility data, and programmatic reporting data (points). Different colours on panel "ANC prevalence" indicate different ANC facilities.

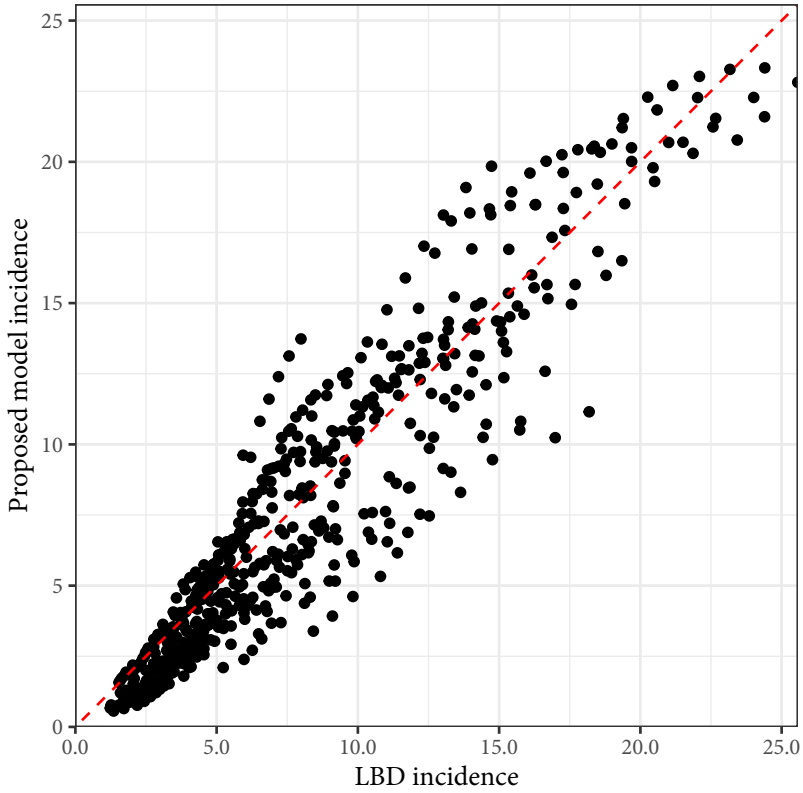

Figure A.29: Comparison of the estimates from Sartorius et al. to the estimates from this study from 2000 through 2018 aggregated over sex.

Appendix B

Appendix of Wolock et al. (2021)

This is a reproduction of the appendix from Wolock et al. (2021). I have updated the section references.

B.1 Age heaping

Respondents in each of these data sets are disproportionately likely to report that their partners' ages are multiples of five or multiples of five away from their own age, leading to distinct "spikes" in the empirical partner age (or age difference) distributions at multiples of five. The left panel of Figure B.1 illustrates this phenomenon among women aged 24 years in the AHRI data. These spikes, widely referred to as "heaping," could bias our results towards certain probability distributions, so we developed a simple deheaping algorithm, applied it to the AHRI data. To account for the possibility that heaping affected the results, we developed a simple deheaping algorithm and treated the deheaped AHRI data as a fourth dataset. Due to the structure of the questionnaire ("how many years older/younger is your partner than you?"), the AHRI partner age data exhibit strong heaping on partner ages that are multiples of five years from the respondent's age. For example, among women aged 24 years, we observe far more partners aged exactly 29 years than expected.

Let $n_{s,a,p}$ be the number of observed partnerships with $s_i = s$, $a_i = a$, and $p_i = p$. Fixing age to be a and sex to be s, we can find the expected count at partner age p, $\hat{n}_{s,a,p}$ by fitting a Nadaraya-Watson estimator to all ordered pairs $(p, n_{s,a,p})$ such that $p - a$ is *not* a multiple

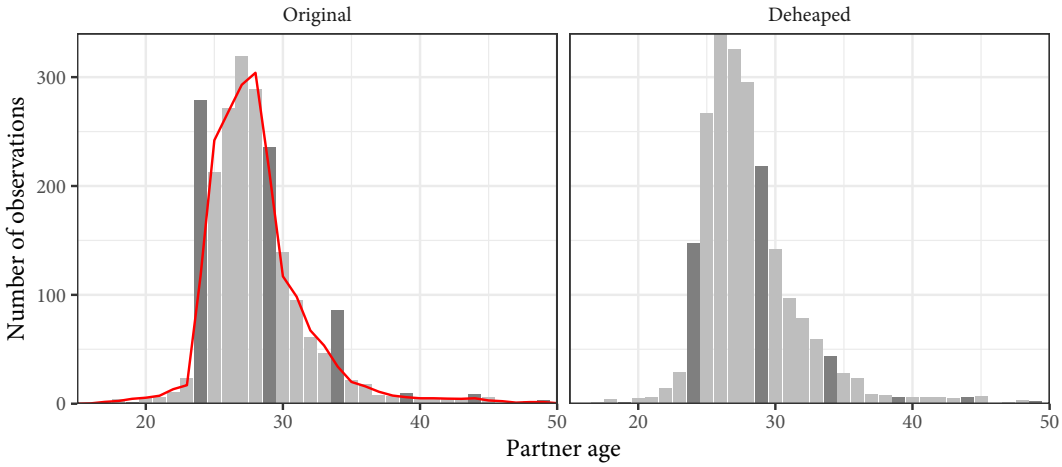

Figure B.1: Illustration of the effect of the deheaping algorithm on women aged exactly 24 years in the AHRI data. Dark grey bars correspond to ages identified as potentially heaped (multiples of five away from 24). The red line is the expected count of observations estimated by excluding any potentially heaped ages.

of five. We can then find the positive-valued excess counts at all p such that $p - a$ *is* a multiple of five:

$$e_{s,a,p} = \max(n_{s,a,p} - \hat{n}_{s,a,p}, 0). \tag{B.1}$$

This quantity, $e_{s,a,p}$, is what the Nadaraya-Watson estimator has identified as number of heaped observations. Fixing $p^\star$ to be a partner age such that $(p^\star - a) \bmod 5 \equiv 0$, we assume that all of the excess mass at $p^\star$ will be allocated to the four partner ages on either side of $p^\star$. We find the share of $e_{s,a,p^\star}$ to be allocated to each of $(p^\star - 2, ..., p^\star + 2)$, denoted $b_{s,a,p}$, as

$$b_{s,a,p} = \frac{n_{s,a,p}}{\sum_{i=-2}^{2} n_{s,a,p^\star+i}}, \tag{B.2}$$

substituting in $\hat{n}_{s,a,p^\star}$ for $n_{s,a,p}$ wherever applicable. Finally, we find the number of individuals to be reassigned from $p^\star$ to each p within two years of $y^\star$ as $d_{s,a,p} = b_{s,a,p} \cdot e_{s,a,p^\star}$. Note that each partner age can only "receive" partnerships from its nearest multiple of five and that each multiple of five can only "send" partnerships to itself and the four partner ages on either side of it. For each y within two years of $y^\star$, we randomly select $\lfloor d_{s,a,p} \rceil$ individuals to move from $p^\star$ to p. We apply this method for both sexes and all respondent ages with at

least two observations separately.

Figure B.1 illustrates the effect of this process on data among women aged 24 in the AHRI data. Despite its simplicity, the deheaping algorithm seems to produce distributions that should be sufficiently plausible for the purposes of this experiment. If our results differed substantially between the original and deheaped AHRI data, we would need to consider the possibility that our results could be an artefact of heaping.

This method is quite simple, but it seems to work reasonably well on the AHRI data. Regardless, we do not need a perfect deheaping algorithm for this application; we just need one that will give us a *plausibly* deheaped version of the AHRI data. If the results differ drastically between the heaped and deheaped datasets (*i.e.* if one probability distribution works perfectly only on the deheaped data), then we will know that our results are sensitive to irregularities in the data.

B.1.1 Results

Figure B.2 shows the presence of age heaping among women in the AHRI data, as well as the effects of our deheaping algorithm. Visible diagonal lines indicate that women were disproportionately likely to report that the difference between their partner's age and their own age was a multiple of five. Heaping to partner ages (not partner age differences) would manifest as horizontal lines. As we can see in the right panel, the deheaping procedure resolves the majority of the heaping. We cannot validate the algorithm, but for the purposes of this experiment, simply producing plausibly deheaped age distributions should be sufficient. Table B.1 provides ELPD and QQ RMSE values for all five regression models fit to the deheaped AHRI data. As with the heaped AHRI data, the most complex distributional model had the highest ELPD (58504.0). From these results, we conclude that the presence of heaping in the three main datasets is unlikely to have substantially altered the results of this analysis.

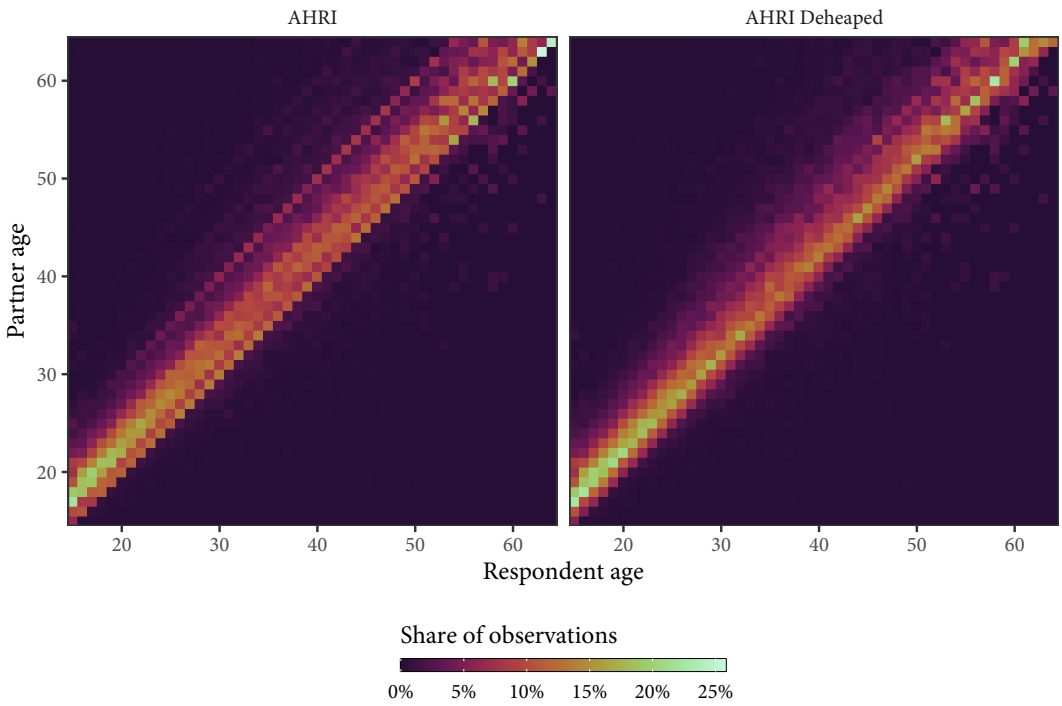

Figure B.2: Observed sexual partner age distributions among women in the AHRI data. The left panel is original data, and the right panel is the same data set after deheaping age differences from multiples of five.

Model	AHRI Deheaped
ELPD	
Conventional	55296.2
Distributional 1	57097.4
Distributional 2	57503.7
Distributional 3	58219.2
Distributional 4	**58504.0**
QQ RMSE	
Conventional	1.26
Distributional 1	1.06
Distributional 2	1.14
Distributional 3	0.92
Distributional 4	**0.62**

Table B.1: ELPD and QQ RMSE values for all five models fit to deheaped AHRI data The models increase in complexity from Conventional Regression to Distributional Model 4. Bolded ELPD values are more than two standard errors higher than the next best value in the column. Bolded QQ RMSE values are lowest in their column.

B.2 Model specification details

We modelled the log-ratio dependent variable using the four-parameter sinh-arcsinh distribution:

$$\begin{aligned} y_i &\sim \sinh(\mu_i, \sigma_i, \epsilon_i, \delta_i) \\ \mu_i &= \beta^{\mu}\mathbf{X}_i^{\mu} \\ \log \sigma_i^{\star} &= \beta^{\sigma}\mathbf{X}_i^{\sigma} \\ \epsilon_i &= \beta^{\epsilon}\mathbf{X}_i^{\epsilon} \\ \log \delta_i &= \beta^{\delta}\mathbf{X}_i^{\delta} \\ \sigma_i &= \sigma_i^{\star}\delta_i, \end{aligned} \tag{B.3}$$

where β^{μ}, β^{σ}, β^{ϵ}, and β^{δ} are free parameters. We placed essentially arbitrary shrinkage priors on all coefficients:

$$\beta^{\mu}, \beta^{\sigma}, \beta^{\epsilon}, \beta^{\delta} \sim \mathrm{N}(0, 5). \tag{B.4}$$

First, we fit a conventional regression, in which only the location parameter, μ, is a function of data. Specifically, we allowed for linear sex and age effects and a linear interaction between respondent sex and age (s_i and a_i, respectively) in the model of μ:

$$\begin{aligned} \mathbf{X}_i^{\mu} &= (1, s_i, a_i, s_i \cdot a_i) \\ \mathbf{X}_i^{\sigma}, \mathbf{X}_i^{\epsilon}, \mathbf{X}_i^{\delta} &= (1). \end{aligned} \tag{B.5}$$

In the second model, we allowed the three higher order distributional parameters to vary by age and sex:

$$\begin{aligned} \mathbf{X}_i^{\mu} &= (1, s_i, a_i, s_i \cdot a_i) \\ \mathbf{X}_i^{\sigma}, \mathbf{X}_i^{\epsilon}, \mathbf{X}_i^{\delta} &= (1, s_i, a_i). \end{aligned} \tag{B.6}$$

In the third model, all four distributional parameters had age, sex, and age-sex interaction effects:

$$\mathbf{X}_i^{\mu}, \mathbf{X}_i^{\sigma}, \mathbf{X}_i^{\epsilon}, \mathbf{X}_i^{\delta} = (1, s_i, a_i, s_i \cdot a_i) \tag{B.7}$$

To allow for the possibility of non-linear variation with respect to age in the fourth model, we modelled the location parameter using sex-specific natural splines on age:

$$\begin{aligned} \mathbf{X}_i^{\mu} &= (1, s_i, \phi_1(a_i), ..., \phi_K(a_i), s_i \cdot \phi_1(a_i), ..., s_i \cdot \phi_K(a_i)) \\ \mathbf{X}_i^{\sigma}, \mathbf{X}_i^{\epsilon}, \mathbf{X}_i^{\delta} &= (1, s_i, a_i, s_i \cdot a_i), \end{aligned} \tag{B.8}$$

where K is the number of columns in the spline design matrix. By including a second set of basis function values that are multiplied by s_i, we are estimating an additional, female-specific trend with respect to age.

Finally, we fit a fifth model, in which all four distributional parameters were modelled as sex-specific splines with respect to age:

$$\mathbf{X}_i^{\mu}, \mathbf{X}_i^{\sigma}, \mathbf{X}_i^{\epsilon}, \mathbf{X}_i^{\delta} = (1, s_i, \phi_1(a_i), ..., \phi_K(a_i), s_i \cdot \phi_1(a_i), ..., s_i \cdot \phi_K(a_i)) \,. \tag{B.9}$$

B.3 Full Results

B.3.1 Supplementary Figures

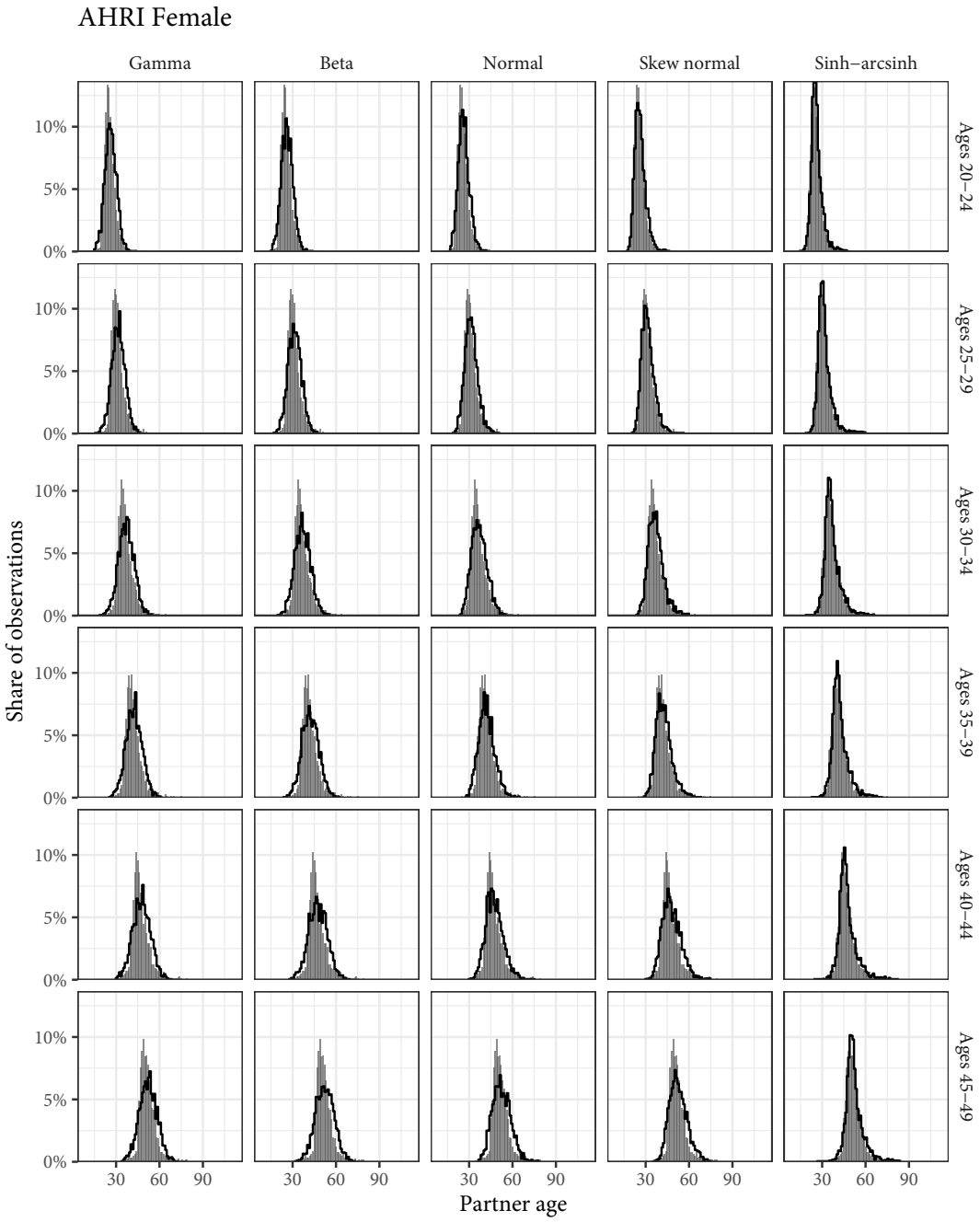

Figure B.3: Observed partner age distributions (grey bars) and posterior predictive partner age distributions (lines) for each probability distribution among women in the AHRI data set. Here, we plot the posterior predicitve distribution associated with each distribution's highest-ELPD dependent variable.

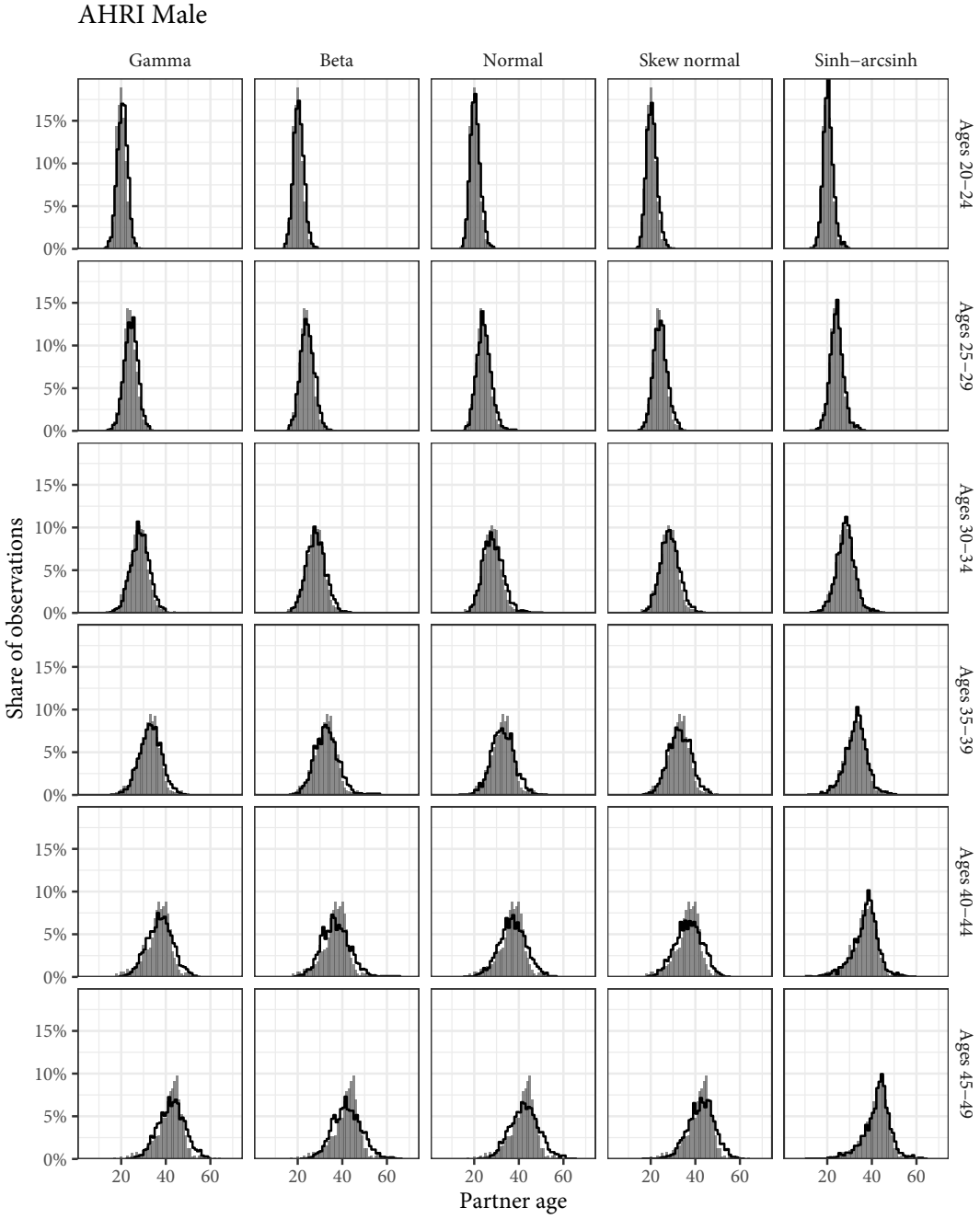

Figure B.4: Observed partner age distributions (grey bars) and posterior predictive partner age distributions (lines) for each probability distribution among men in the AHRI data set. Here, we plot the posterior predicitve distribution associated with each distribution's highest-ELPD dependent variable.

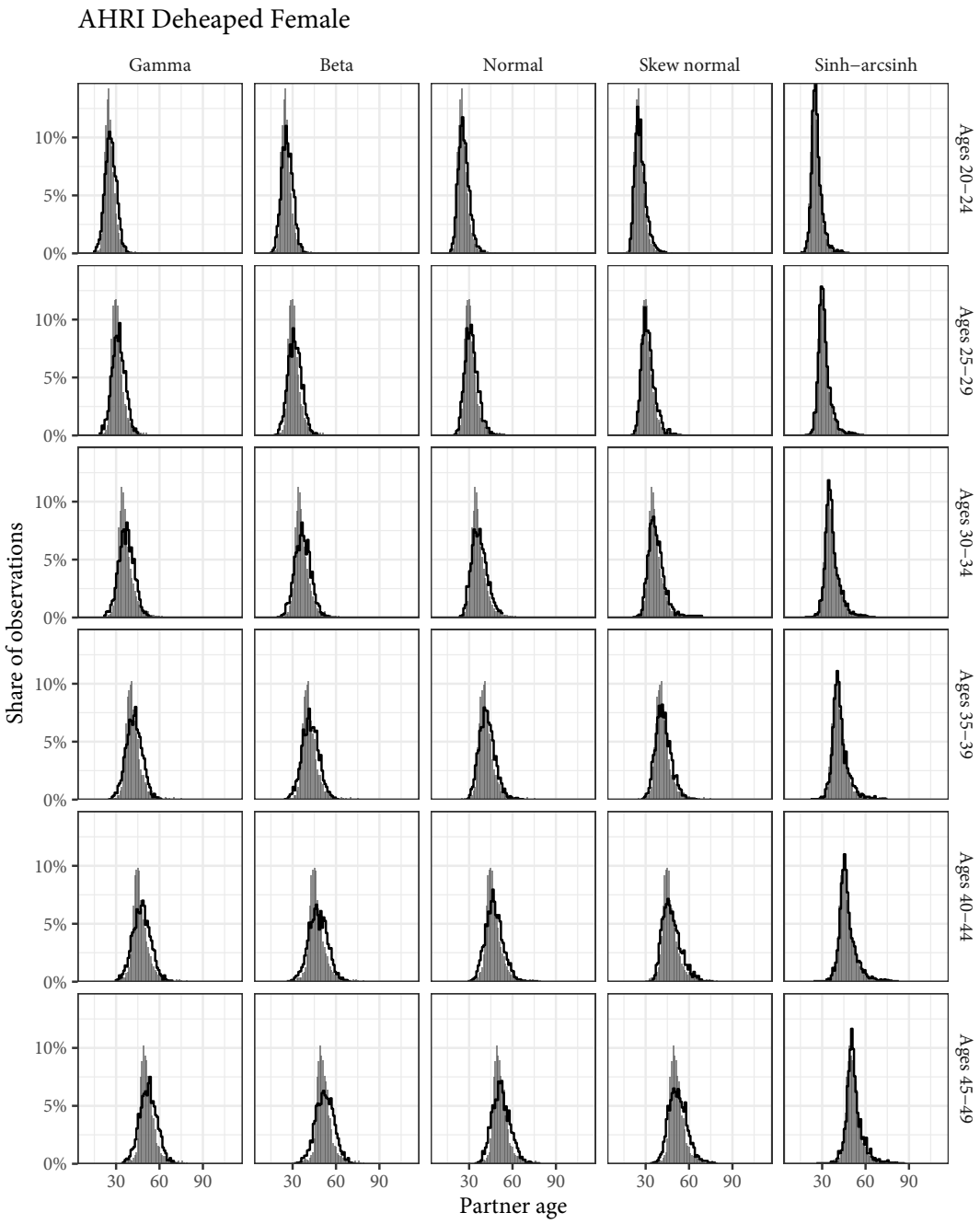

Figure B.5: Observed partner age distributions (grey bars) and posterior predictive partner age distributions (lines) for each probability distribution among women in the AHRI Deheaped data set. Here, we plot the posterior predicitve distribution associated with each distribution's highest-ELPD dependent variable.

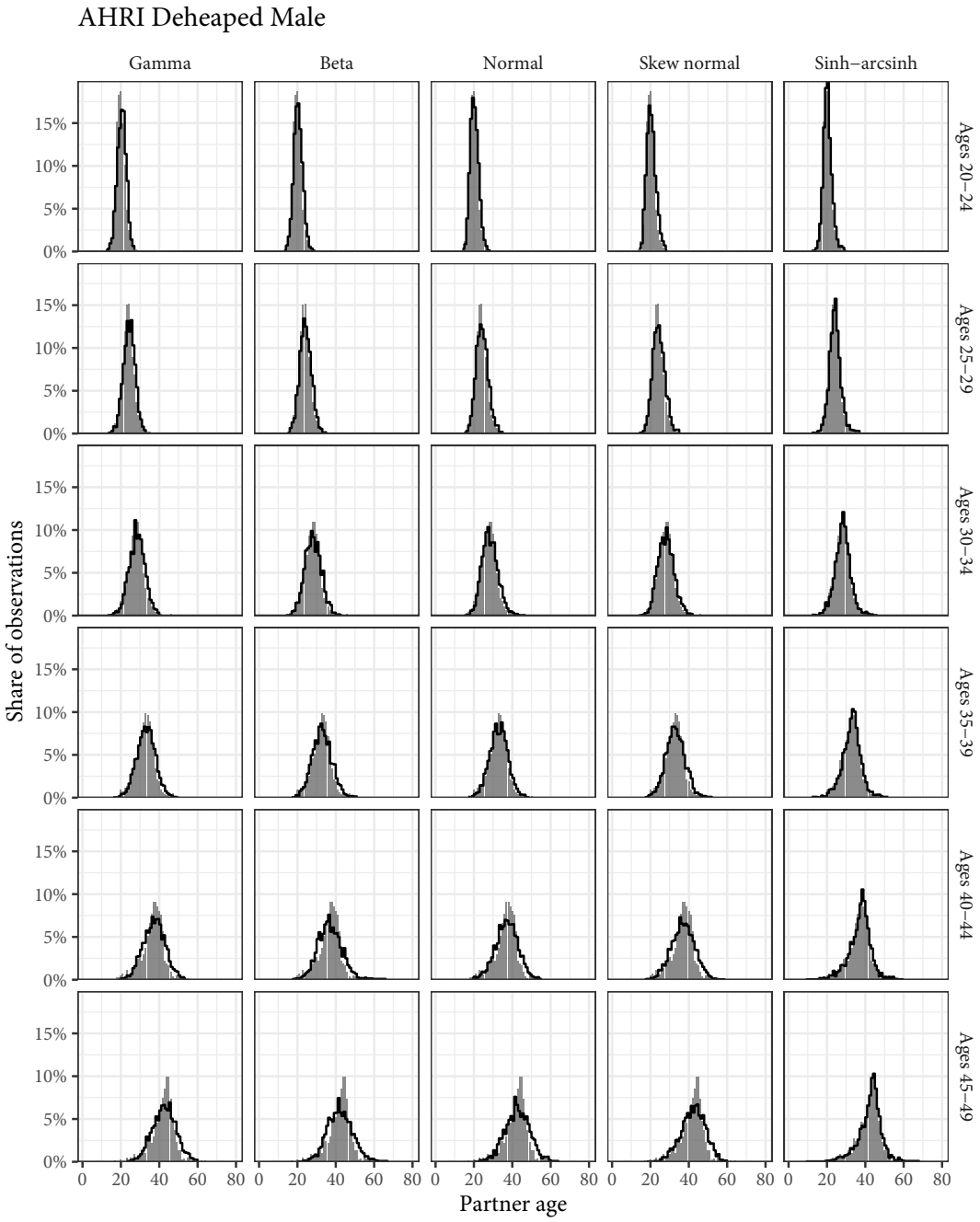

Figure B.6: Observed partner age distributions (grey bars) and posterior predictive partner age distributions (lines) for each probability distribution among men in the AHRI Deheaped data set. Here, we plot the posterior predicitve distribution associated with each distribution's highest-ELPD dependent variable.

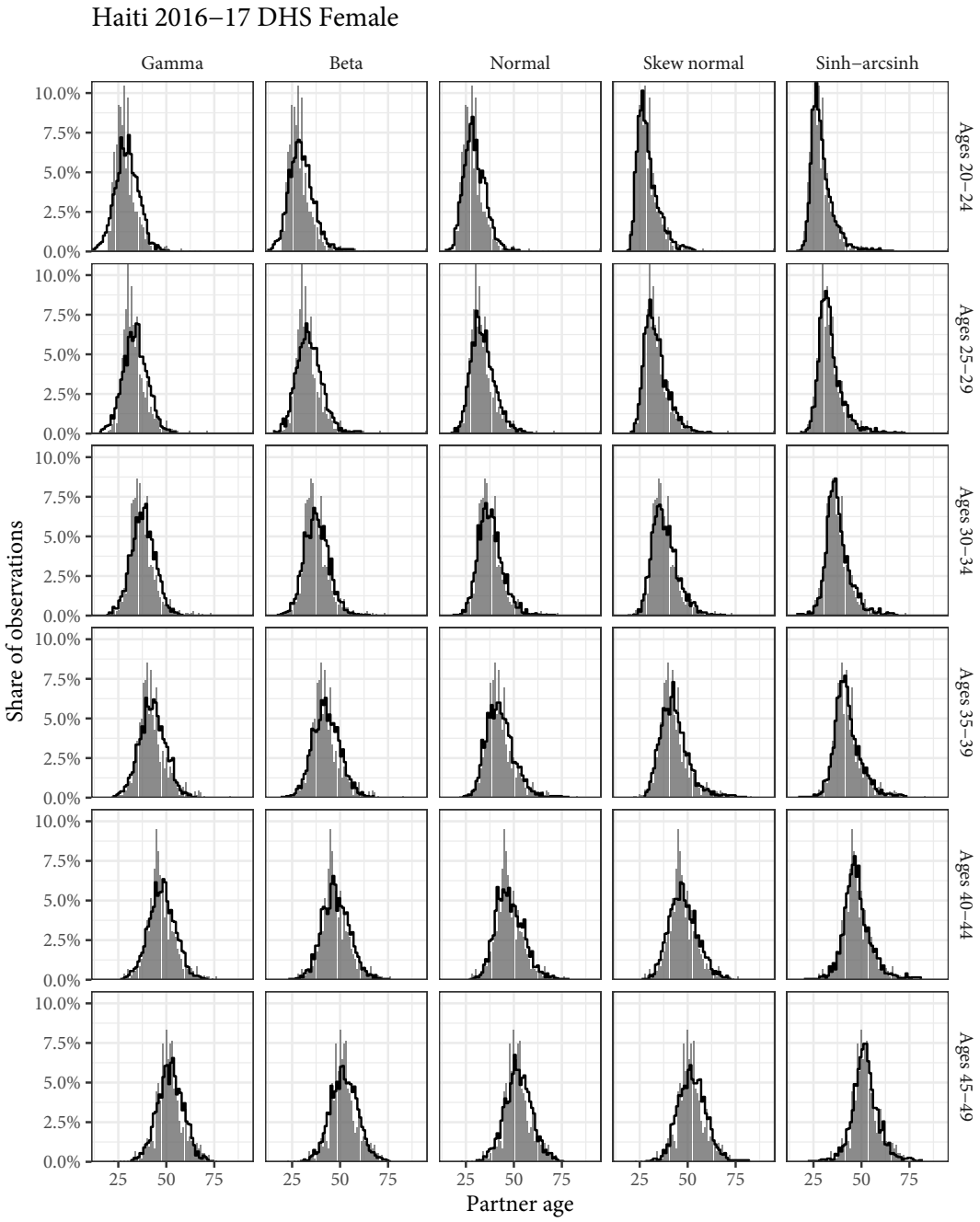

Figure B.7: Observed partner age distributions (grey bars) and posterior predictive partner age distributions (lines) for each probability distribution among women in the Haiti 2016-17 DHS data set. Here, we plot the posterior predicitve distribution associated with each distribution's highest-ELPD dependent variable.

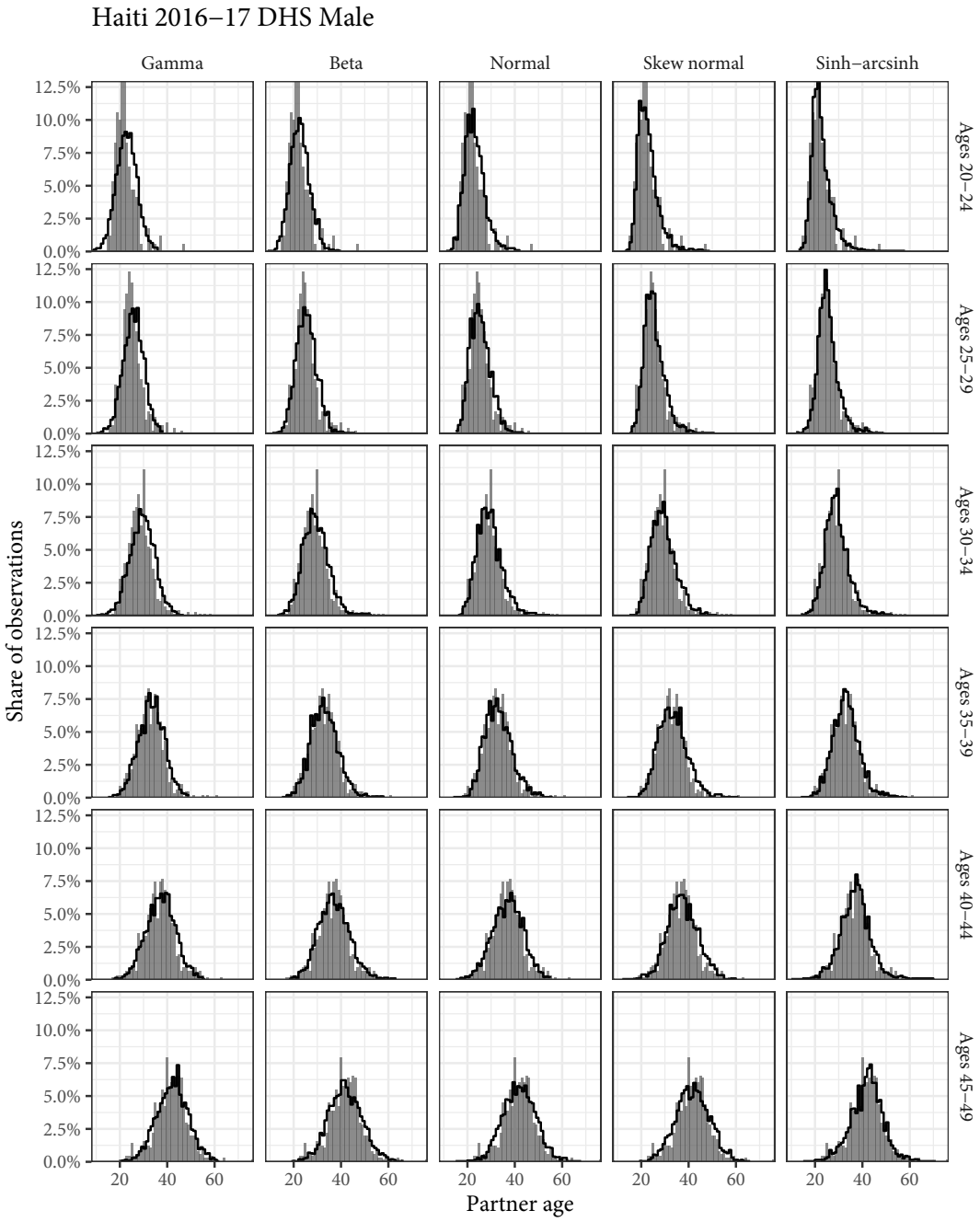

Figure B.8: Observed partner age distributions (grey bars) and posterior predictive partner age distributions (lines) for each probability distribution among men in the Haiti 2016-17 DHS data set. Here, we plot the posterior predicitve distribution associated with each distribution's highest-ELPD dependent variable.

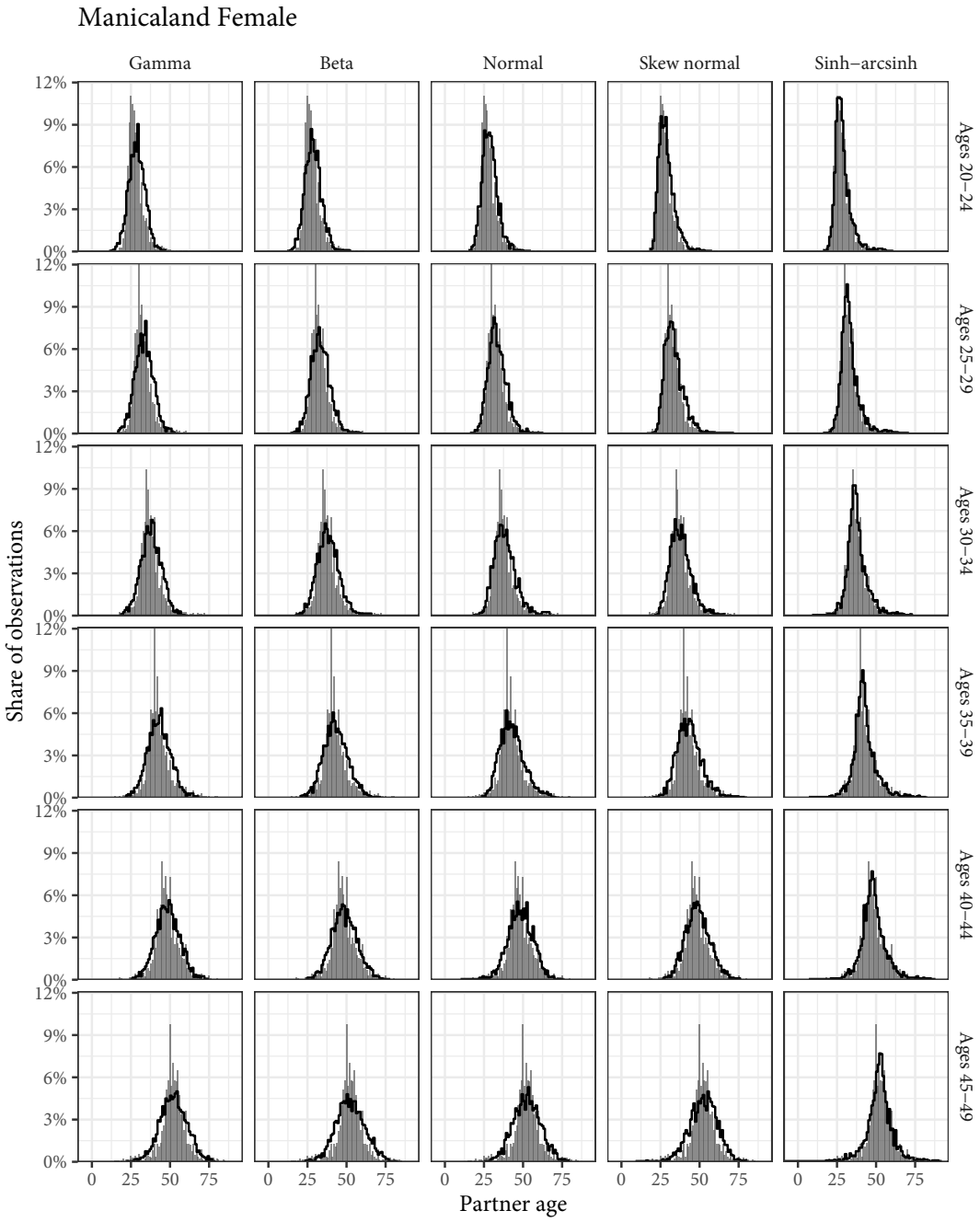

Figure B.9: Observed partner age distributions (grey bars) and posterior predictive partner age distributions (lines) for each probability distribution among women in the Manicaland data set. Here, we plot the posterior predicitve distribution associated with each distribution's highest-ELPD dependent variable.

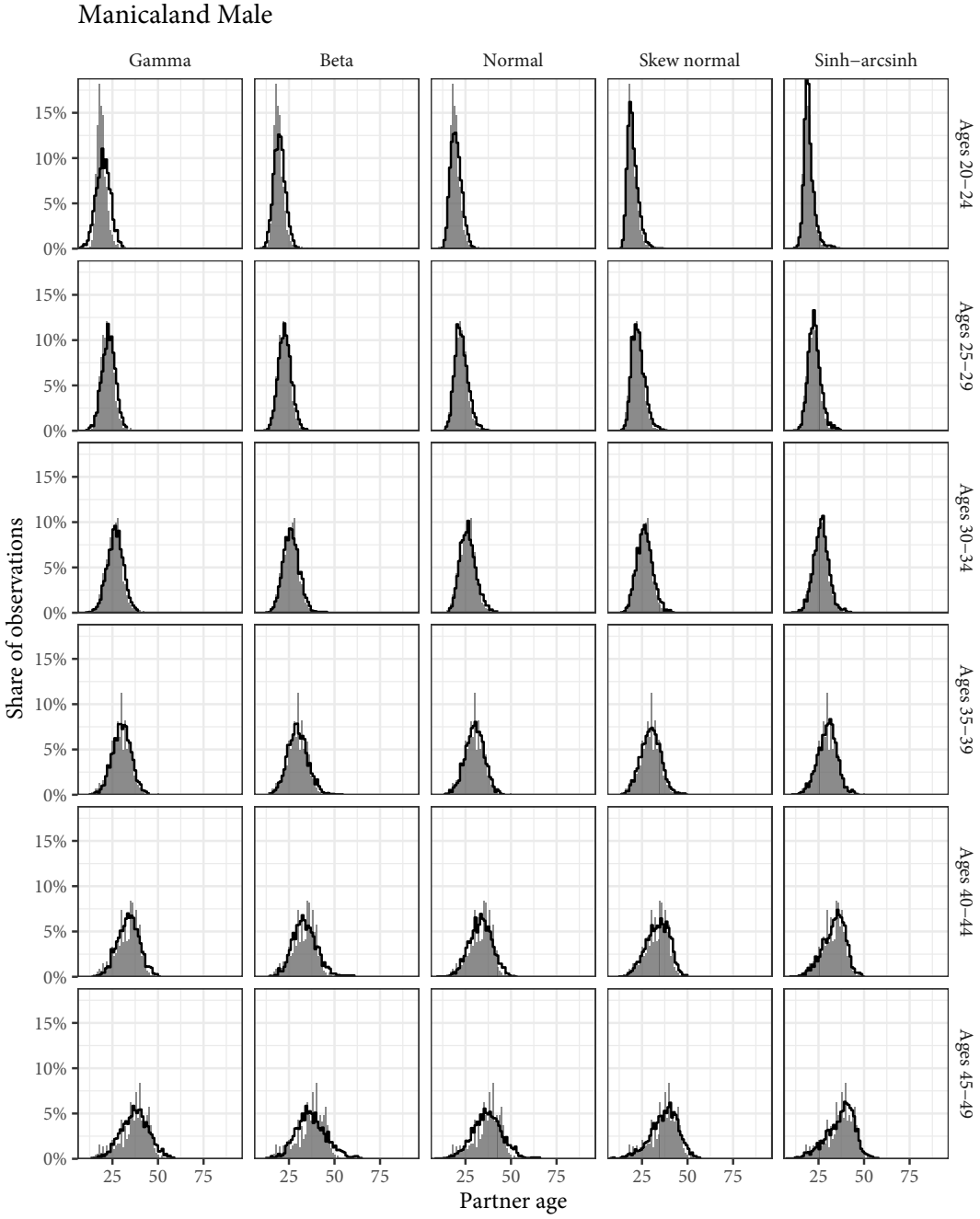

Figure B.10: Observed partner age distributions (grey bars) and posterior predictive partner age distributions (lines) for each probability distribution among men in the Manicaland data set. Here, we plot the posterior predicitve distribution associated with each distribution's highest-ELPD dependent variable.

B.3.2 Supplementary Tables

Rank	Distribution	ELPD	ELPD Diff	SE of Diff	QQ RMSE
AHRI Female 20-24					
1	Sinh-arcsinh	-31750.94	0.00	0.00	0.32
2	Skew normal	-32056.39	-305.46	48.63	0.47
3	Normal	-32414.54	-663.61	60.54	0.62
4	Beta	-32953.92	-1202.98	112.08	0.77
5	Gamma	-33461.85	-1710.92	148.15	0.80
AHRI Female 25-29					
1	Sinh-arcsinh	-24647.65	0.00	0.00	0.28
2	Skew normal	-24906.22	-258.57	43.27	0.52
3	Normal	-25238.71	-591.06	54.82	0.68
4	Beta	-25701.13	-1053.48	114.84	0.89
5	Gamma	-25995.81	-1348.16	132.15	0.90
AHRI Female 30-34					
1	Sinh-arcsinh	-19831.53	0.00	0.00	0.44
2	Skew normal	-20200.44	-368.91	69.40	0.51
3	Normal	-20314.79	-483.26	52.24	0.80
4	Beta	-20575.61	-744.08	67.46	0.93
5	Gamma	-20708.35	-876.82	73.89	0.91
AHRI Female 35-39					
1	Sinh-arcsinh	-15469.18	0.00	0.00	0.31
2	Skew normal	-15749.79	-280.61	53.04	0.77
3	Normal	-15834.32	-365.14	41.23	0.80
4	Beta	-16026.51	-557.33	53.99	1.18
5	Gamma	-16087.40	-618.22	57.06	1.04
AHRI Female 40-44					
1	Sinh-arcsinh	-12556.61	0.00	0.00	0.45
2	Skew normal	-12876.71	-320.10	45.85	1.27
3	Normal	-12935.34	-378.73	52.38	0.92
4	Beta	-13137.69	-581.08	69.18	1.38
5	Gamma	-13150.66	-594.05	62.73	1.19
AHRI Female 45-49					
1	Sinh-arcsinh	-10059.21	0.00	0.00	0.59
2	Skew normal	-10391.95	-332.74	42.75	1.36
3	Normal	-10433.64	-374.43	48.91	1.53
4	Gamma	-10527.00	-467.79	50.72	1.35
5	Beta	-10545.33	-486.12	56.02	1.58

Table B.2: Full ELPD and QQ RMSE table for women in the AHRI dataset. Higher ELPD values and lower QQ RMSE values are better.

Rank	Distribution	ELPD	ELPD Diff	SE of Diff	QQ RMSE
AHRI Male 20-24					
1	Sinh-arcsinh	-20428.11	0.00	0.00	0.23
2	Skew normal	-20499.86	-71.75	17.12	0.25
3	Normal	-20503.89	-75.79	16.85	0.22
4	Beta	-20545.59	-117.49	23.21	0.22
5	Gamma	-20700.24	-272.13	43.53	0.29
AHRI Male 25-29					
1	Sinh-arcsinh	-12664.21	0.00	0.00	0.26
2	Skew normal	-12727.03	-62.82	17.86	0.28
3	Beta	-12739.03	-74.82	18.65	0.31
4	Normal	-12753.25	-89.04	19.35	0.29
5	Gamma	-12788.26	-124.05	35.07	0.38
AHRI Male 30-34					
1	Sinh-arcsinh	-9301.03	0.00	0.00	0.29
2	Skew normal	-9357.18	-56.15	14.08	0.43
3	Beta	-9371.86	-70.83	16.48	0.37
4	Normal	-9385.63	-84.60	14.67	0.46
5	Gamma	-9419.34	-118.31	35.11	0.27
AHRI Male 35-39					
1	Sinh-arcsinh	-6746.89	0.00	0.00	0.30
2	Skew normal	-6812.77	-65.88	17.73	0.64
3	Normal	-6817.86	-70.97	23.24	0.70
4	Beta	-6830.95	-84.06	17.95	0.71
5	Gamma	-6832.47	-85.58	32.03	0.44
AHRI Male 40-44					
1	Sinh-arcsinh	-4610.95	0.00	0.00	0.35
2	Skew normal	-4711.78	-100.84	18.66	0.92
3	Normal	-4713.78	-102.83	18.82	0.78
4	Gamma	-4718.28	-107.33	24.83	0.63
5	Beta	-4742.70	-131.75	17.28	1.07
AHRI Male 45-49					
1	Sinh-arcsinh	-3683.47	0.00	0.00	0.34
2	Skew normal	-3770.59	-87.12	16.56	0.81
3	Gamma	-3776.33	-92.86	15.56	0.87
4	Normal	-3778.84	-95.37	14.50	1.17
5	Beta	-3805.78	-122.31	17.40	1.36

Table B.3: Full ELPD and QQ RMSE table for men in the AHRI dataset. Higher ELPD values and lower QQ RMSE values are better.

Rank	Distribution	ELPD	ELPD Diff	SE of Diff	QQ RMSE
AHRI Deheaped Female 20-24					
1	Sinh-arcsinh	-31411.24	0.00	0.00	0.26
2	Skew normal	-31797.37	-386.13	53.15	0.59
3	Normal	-32179.29	-768.05	65.50	0.56
4	Beta	-32737.57	-1326.32	118.47	0.76
5	Gamma	-33254.17	-1842.92	155.14	0.78
AHRI Deheaped Female 25-29					
1	Sinh-arcsinh	-24439.47	0.00	0.00	0.27
2	Skew normal	-24768.06	-328.59	46.71	0.65
3	Normal	-25104.46	-664.99	58.32	0.82
4	Beta	-25574.33	-1134.86	119.65	1.03
5	Gamma	-25870.30	-1430.83	137.51	1.05
AHRI Deheaped Female 30-34					
1	Sinh-arcsinh	-19680.77	0.00	0.00	0.41
2	Skew normal	-20112.70	-431.94	72.95	0.55
3	Normal	-20228.52	-547.76	56.19	0.81
4	Beta	-20492.23	-811.46	70.53	0.92
5	Gamma	-20624.82	-944.06	76.98	0.80
AHRI Deheaped Female 35-39					
1	Sinh-arcsinh	-15381.68	0.00	0.00	0.26
2	Skew normal	-15703.77	-322.09	55.30	0.68
3	Normal	-15788.73	-407.05	43.67	0.82
4	Beta	-15983.57	-601.90	56.31	1.13
5	Gamma	-16044.22	-662.54	59.17	1.04
AHRI Deheaped Female 40-44					
1	Sinh-arcsinh	-12491.91	0.00	0.00	0.25
2	Skew normal	-12846.63	-354.72	47.38	0.99
3	Normal	-12905.04	-413.12	54.14	0.89
4	Beta	-13109.82	-617.91	70.96	1.31
5	Gamma	-13121.45	-629.53	64.12	1.13
AHRI Deheaped Female 45-49					
1	Sinh-arcsinh	-9981.83	0.00	0.00	0.53
2	Skew normal	-10357.85	-376.01	45.08	1.43
3	Normal	-10401.64	-419.80	51.57	1.46
4	Gamma	-10493.73	-511.90	52.90	1.37
5	Beta	-10513.46	-531.63	58.21	1.61

Table B.4: Full ELPD and QQ RMSE table for women in the AHRI Deheaped dataset. Higher ELPD values and lower QQ RMSE values are better.

Rank	Distribution	ELPD	ELPD Diff	SE of Diff	QQ RMSE
AHRI Deheaped Male 20-24					
1	Sinh-arcsinh	-20310.35	0.00	0.00	0.27
2	Skew normal	-20429.90	-119.55	27.09	0.22
3	Normal	-20459.73	-149.38	35.78	0.29
4	Beta	-20574.15	-263.80	75.13	0.22
5	Gamma	-20899.52	-589.17	175.99	0.27
AHRI Deheaped Male 25-29					
1	Sinh-arcsinh	-12585.54	0.00	0.00	0.28
2	Skew normal	-12680.59	-95.05	21.53	0.44
3	Beta	-12697.00	-111.46	23.31	0.37
4	Normal	-12701.76	-116.23	22.96	0.41
5	Gamma	-12763.81	-178.27	41.24	0.39
AHRI Deheaped Male 30-34					
1	Sinh-arcsinh	-9227.26	0.00	0.00	0.37
2	Skew normal	-9302.42	-75.16	16.15	0.41
3	Beta	-9318.24	-90.97	19.07	0.39
4	Normal	-9327.58	-100.31	16.18	0.41
5	Gamma	-9372.32	-145.06	38.27	0.27
AHRI Deheaped Male 35-39					
1	Sinh-arcsinh	-6694.86	0.00	0.00	0.30
2	Skew normal	-6774.11	-79.26	19.32	0.61
3	Normal	-6780.69	-85.84	25.42	0.44
4	Beta	-6791.95	-97.10	19.81	0.69
5	Gamma	-6796.41	-101.55	34.45	0.40
AHRI Deheaped Male 40-44					
1	Sinh-arcsinh	-4591.04	0.00	0.00	0.49
2	Skew normal	-4700.54	-109.51	19.38	1.16
3	Normal	-4703.52	-112.49	19.93	1.00
4	Gamma	-4708.43	-117.40	25.94	0.89
5	Beta	-4731.41	-140.37	17.84	1.30
AHRI Deheaped Male 45-49					
1	Sinh-arcsinh	-3680.18	0.00	0.00	0.30
2	Normal	-3796.06	-115.88	19.24	1.15
3	Skew normal	-3797.14	-116.95	23.48	1.02
4	Gamma	-3801.02	-120.83	24.51	0.98
5	Beta	-3817.97	-137.79	19.37	1.39

Table B.5: Full ELPD and QQ RMSE table for men in the AHRI Deheaped dataset. Higher ELPD values and lower QQ RMSE values are better.

Rank	Distribution	ELPD	ELPD Diff	SE of Diff	QQ RMSE
Haiti 2016-17 DHS Female 20-24					
1	Sinh-arcsinh	-3259.31	0.00	0.00	0.49
2	Skew normal	-3263.46	-4.15	4.95	0.53
3	Normal	-3338.23	-78.92	19.54	0.91
4	Beta	-3441.91	-182.60	45.77	1.24
5	Gamma	-3504.85	-245.54	53.90	1.29
Haiti 2016-17 DHS Female 25-29					
1	Sinh-arcsinh	-4447.43	0.00	0.00	0.26
2	Skew normal	-4471.22	-23.78	8.41	0.57
3	Normal	-4527.25	-79.82	18.72	0.86
4	Beta	-4625.97	-178.54	40.88	1.23
5	Gamma	-4678.20	-230.77	45.81	1.22
Haiti 2016-17 DHS Female 30-34					
1	Sinh-arcsinh	-4720.12	0.00	0.00	0.44
2	Skew normal	-4749.57	-29.45	9.06	0.68
3	Normal	-4763.78	-43.66	10.51	0.62
4	Beta	-4809.11	-88.99	17.19	0.85
5	Gamma	-4836.82	-116.70	20.32	0.83
Haiti 2016-17 DHS Female 35-39					
1	Sinh-arcsinh	-4490.82	0.00	0.00	0.33
2	Skew normal	-4518.58	-27.75	8.14	0.57
3	Normal	-4526.55	-35.73	8.59	0.73
4	Beta	-4561.27	-70.45	13.60	0.94
5	Gamma	-4577.84	-87.01	15.27	0.86
Haiti 2016-17 DHS Female 40-44					
1	Sinh-arcsinh	-3601.02	0.00	0.00	0.35
2	Skew normal	-3629.45	-28.43	7.51	0.83
3	Normal	-3633.14	-32.11	7.96	0.71
4	Beta	-3641.61	-40.59	9.76	0.73
5	Gamma	-3644.89	-43.86	10.47	0.64
Haiti 2016-17 DHS Female 45-49					
1	Sinh-arcsinh	-3106.27	0.00	0.00	0.39
2	Skew normal	-3133.10	-26.82	7.68	0.88
3	Gamma	-3133.61	-27.33	7.50	0.68
4	Normal	-3134.62	-28.35	7.46	0.81
5	Beta	-3136.89	-30.62	8.61	0.88

Table B.6: Full ELPD and QQ RMSE table for women in the Haiti 2016-17 DHS dataset. Higher ELPD values and lower QQ RMSE values are better.

Rank	Distribution	ELPD	ELPD Diff	SE of Diff	QQ RMSE
Haiti 2016-17 DHS Male 20-24					
1	Skew normal	-468.98	0.00	0.00	0.43
2	Sinh-arcsinh	-469.60	-0.62	1.12	0.41
3	Normal	-475.31	-6.33	4.28	0.67
4	Beta	-483.53	-14.55	7.33	0.65
5	Gamma	-500.53	-31.55	13.35	0.94
Haiti 2016-17 DHS Male 25-29					
1	Sinh-arcsinh	-1386.13	0.00	0.00	0.38
2	Skew normal	-1390.54	-4.41	3.19	0.49
3	Normal	-1395.47	-9.34	4.79	0.60
4	Beta	-1407.46	-21.32	7.31	0.62
5	Gamma	-1434.18	-48.04	11.75	0.79
Haiti 2016-17 DHS Male 30-34					
1	Sinh-arcsinh	-2217.20	0.00	0.00	0.44
2	Skew normal	-2222.10	-4.89	3.42	0.69
3	Normal	-2223.97	-6.76	4.48	0.45
4	Beta	-2240.58	-23.37	9.32	0.52
5	Gamma	-2281.18	-63.98	17.91	0.73
Haiti 2016-17 DHS Male 35-39					
1	Sinh-arcsinh	-2185.96	0.00	0.00	0.28
2	Skew normal	-2189.87	-3.91	2.67	0.69
3	Beta	-2191.05	-5.10	3.68	0.48
4	Normal	-2191.11	-5.16	3.55	0.49
5	Gamma	-2205.69	-19.73	9.57	0.52
Haiti 2016-17 DHS Male 40-44					
1	Sinh-arcsinh	-2051.62	0.00	0.00	0.39
2	Skew normal	-2060.16	-8.54	4.21	0.72
3	Normal	-2060.38	-8.75	4.57	0.69
4	Beta	-2062.00	-10.37	4.87	0.70
5	Gamma	-2063.79	-12.17	5.73	0.47
Haiti 2016-17 DHS Male 45-49					
1	Sinh-arcsinh	-2138.34	0.00	0.00	0.23
2	Normal	-2150.53	-12.19	6.38	0.35
3	Skew normal	-2151.51	-13.17	5.97	0.56
4	Gamma	-2152.88	-14.54	8.93	0.25
5	Beta	-2156.13	-17.79	6.14	0.48

Table B.7: Full ELPD and QQ RMSE table for men in the Haiti 2016-17 DHS dataset. Higher ELPD values and lower QQ RMSE values are better.

Rank	Distribution	ELPD	ELPD Diff	SE of Diff	QQ RMSE
Manicaland Female 20-24					
1	Sinh-arcsinh	-16390.77	0.00	0.00	0.31
2	Skew normal	-16502.01	-111.25	21.22	0.44
3	Normal	-16779.93	-389.16	37.05	0.67
4	Beta	-17111.57	-720.80	62.02	0.86
5	Gamma	-17387.38	-996.61	76.80	1.02
Manicaland Female 25-29					
1	Sinh-arcsinh	-18702.50	0.00	0.00	0.53
2	Skew normal	-18923.04	-220.53	25.27	0.94
3	Normal	-19080.66	-378.16	36.05	0.83
4	Beta	-19405.80	-703.30	64.97	1.05
5	Gamma	-19615.53	-913.03	76.38	1.09
Manicaland Female 30-34					
1	Sinh-arcsinh	-16523.81	0.00	0.00	0.48
2	Skew normal	-16877.96	-354.15	40.36	0.87
3	Normal	-16886.62	-362.80	36.41	0.99
4	Beta	-17021.26	-497.44	43.60	1.12
5	Gamma	-17094.58	-570.76	49.53	0.93
Manicaland Female 35-39					
1	Sinh-arcsinh	-14397.76	0.00	0.00	0.48
2	Skew normal	-14736.64	-338.88	28.35	1.25
3	Normal	-14798.55	-400.79	36.87	1.39
4	Beta	-14824.80	-427.04	33.02	1.47
5	Gamma	-14835.11	-437.35	34.49	1.14
Manicaland Female 40-44					
1	Sinh-arcsinh	-12293.13	0.00	0.00	0.68
2	Skew normal	-12488.28	-195.15	21.36	1.03
3	Gamma	-12500.93	-207.80	22.18	1.03
4	Normal	-12508.91	-215.78	23.29	1.28
5	Beta	-12537.14	-244.01	25.41	1.22
Manicaland Female 45-49					
1	Sinh-arcsinh	-9183.03	0.00	0.00	0.56
2	Skew normal	-9455.87	-272.83	23.57	1.68
3	Normal	-9477.33	-294.30	23.55	1.62
4	Gamma	-9497.31	-314.27	25.08	1.44
5	Beta	-9576.44	-393.40	32.07	1.94

Table B.8: Full ELPD and QQ RMSE table for women in the Manicaland dataset. Higher ELPD values and lower QQ RMSE values are better.

Rank	Distribution	ELPD	ELPD Diff	SE of Diff	QQ RMSE
Manicaland Male 20-24					
1	Sinh-arcsinh	-9770.00	0.00	0.00	0.30
2	Skew normal	-9895.82	-125.83	33.35	0.40
3	Normal	-10139.11	-369.11	79.13	0.49
4	Beta	-10587.64	-817.64	181.23	0.56
5	Gamma	-11594.58	-1824.59	388.26	1.15
Manicaland Male 25-29					
1	Sinh-arcsinh	-13978.59	0.00	0.00	0.40
2	Skew normal	-13990.39	-11.80	8.51	0.48
3	Normal	-14018.60	-40.00	17.48	0.45
4	Beta	-14152.35	-173.76	48.77	0.40
5	Gamma	-14500.47	-521.87	117.58	0.55
Manicaland Male 30-34					
1	Sinh-arcsinh	-12949.24	0.00	0.00	0.31
2	Skew normal	-13016.44	-67.21	25.01	0.37
3	Normal	-13037.46	-88.22	16.57	0.49
4	Beta	-13070.31	-121.07	54.74	0.41
5	Gamma	-13285.47	-336.23	171.92	0.42
Manicaland Male 35-39					
1	Sinh-arcsinh	-11496.14	0.00	0.00	0.27
2	Skew normal	-11528.36	-32.22	9.83	0.39
3	Normal	-11530.43	-34.29	9.72	0.26
4	Gamma	-11531.75	-35.61	12.47	0.24
5	Beta	-11582.63	-86.49	12.97	0.48
Manicaland Male 40-44					
1	Sinh-arcsinh	-8714.06	0.00	0.00	0.35
2	Skew normal	-8749.78	-35.72	10.11	0.51
3	Gamma	-8777.08	-63.02	10.38	0.55
4	Normal	-8791.45	-77.38	12.23	0.76
5	Beta	-8860.22	-146.16	18.15	0.93
Manicaland Male 45-49					
1	Sinh-arcsinh	-7110.27	0.00	0.00	0.42
2	Skew normal	-7177.03	-66.75	25.08	0.76
3	Gamma	-7213.99	-103.72	13.02	1.07
4	Normal	-7232.04	-121.77	13.09	1.28
5	Beta	-7312.35	-202.08	18.61	1.61

Table B.9: Full ELPD and QQ RMSE table for men in the Manicaland dataset. Higher ELPD values and lower QQ RMSE values are better.

Rank	Model	ELPD	ELPD Diff	SE of Diff	QQ RMSE
AHRI					
1	Distributional 4	55841.91	0.00	0.00	0.66
2	Distributional 3	55534.16	-307.75	32.36	0.93
3	Distributional 2	54794.79	-1047.12	51.69	1.21
4	Distributional 1	54335.19	-1506.72	72.32	1.15
5	Conventional	52689.21	-3152.70	100.59	1.30
AHRI Deaheaped					
1	Distributional 4	58503.98	0.00	0.00	0.62
2	Distributional 3	58219.23	-284.75	28.64	0.92
3	Distributional 2	57503.68	-1000.30	47.14	1.14
4	Distributional 1	57097.39	-1406.59	64.48	1.06
5	Conventional	55296.25	-3207.73	99.42	1.26
Haiti 2016-17 DHS					
1	Distributional 4	5207.57	0.00	0.00	0.84
2	Distributional 3	5196.69	-10.89	6.54	0.91
3	Distributional 1	5140.77	-66.80	12.27	0.98
4	Distributional 2	5138.75	-68.83	12.24	0.99
5	Conventional	4777.78	-429.80	30.54	1.33
Manicaland					
1	Distributional 4	24516.15	0.00	0.00	1.04
2	Distributional 3	24313.74	-202.40	20.52	1.34
3	Distributional 2	23472.07	-1044.08	47.77	1.80
4	Distributional 1	23192.49	-1323.66	54.97	1.89
5	Conventional	21011.29	-3504.86	89.01	2.05

Table B.10: LOO-CV estimated ELPD values, differences, and standard errors of differences, as well as QQ RMSE values, for all five regression models fit to all four datasets. The "difference" value of a row is the difference between that row's ELPD value and dataset-specific best ELPD value. Higher ELPD values and lower QQ RMSE values are better.

References

Ades, A. E., & Medley, G. F. (1994). Estimates of disease incidence in women based on antenatal or neonatal seroprevalence data: HIV in new york city. *Statistics in Medicine*, *13*(18), 1881–1894. http://doi.org/10.1002/sim.4780131809

Akullian, A., Bershteyn, A., Klein, D., Vandormael, A., Bärnighausen, T., & Tanser, F. (2017). Sexual partnership age pairings and risk of HIV acquisition in rural south africa. *AIDS (London, England)*, *31*(12), 1755–1764. http://doi.org/10.1097/QAD.0000000000001553

Anderson, R. M., May, R. M., Ng, T. W., & Rowley, J. T. (1992). Age-dependent choice of sexual partners and the transmission dynamics of HIV in sub-saharan africa. *Philosophical Transactions of the Royal Society of London. Series B: Biological Sciences*, *336*(1277), 135–155. http://doi.org/10.1098/rstb.1992.0052

Anderson, S.-J., Cherutich, P., Kilonzo, N., Cremin, I., Fecht, D., Kimanga, D., … Hallett, T. B. (2014). Maximising the effect of combination HIV prevention through prioritisation of the people and places in greatest need: A modelling study. *The Lancet*, *384*(9939), 249–256. http://doi.org/10.1016/S0140-6736(14)61053-9

Arias Garcia, S., Chen, J., Calleja, J. G., Sabin, K., Ogbuanu, C., Lowrance, D., & Zhao, J. (2020). Availability and quality of surveillance and survey data on HIV prevalence among sex workers, men who have sex with men, people who inject drugs, and transgender women in low- and middle-income countries: Review of available data (2001-2017). *JMIR Public Health and Surveillance*, *6*(4), e21688. http://doi.org/10.2196/21688

Ataguba, J. E.-O. (2018). A reassessment of global antenatal care coverage for improving maternal health using sub-saharan africa as a case study. *PLOS ONE*, *13*(10), e0204822. http://doi.org/10.1371/journal.pone.0204822

Bacchetti, P., Segal, M. R., & Jewell, N. P. (1993). Backcalculation of HIV infection rates. *Statistical Science*, *8*(2), 82–101. Retrieved from `https://www.jstor.org/stable/2246140`

Bailey, J. (2018). *Geogrid: Turn geospatial polygons into regular or hexagonal grids*. Retrieved from `https://CRAN.R-project.org/package=geogrid`

Bao, L. (2012). A new infectious disease model for estimating and projecting HIV/AIDS epidemics. *Sexually Transmitted Infections*, *88*, i58–i64.

http://doi.org/10.1136/sextrans-2012-050689

Beauclair, R., Hens, N., & Delva, W. (2018). The role of age-mixing patterns in HIV transmission dynamics: Novel hypotheses from a field study in cape town, south africa. *Epidemics*, *25*, 61–71. http://doi.org/10.1016/j.epidem.2018.05.006

Bhatt, S., Eaton, J. W., Tait, A., Mahy, M., Wanyeki, I., Ghys, P. D., … Gething, P. W. (2019). *Subnational mapping of HIV prevalence, ART coverage and relative incidence risk using the HIVE map model*.

Boerma, J. T., & Weir, S. S. (2005). Integrating demographic and epidemiological approaches to research on HIV/AIDS: The proximate-determinants framework. *The Journal of Infectious Diseases*, *191*, S61–S67. http://doi.org/10.1086/425282

Brookmeyer, R. (2010). Measuring the HIV/AIDS epidemic: Approaches and challenges. *Epidemiologic Reviews*, *32*(1), 26–37. http://doi.org/10.1093/epirev/mxq002

Brookmeyer, R., & Gail, M. H. (1988). A method for obtaining short-term projections and lower bounds on the size of the aids epidemic. *Journal of the American Statistical Association*, *83*(402), 301–308. http://doi.org/10.1080/01621459.1988.10478599

Brown, T., Bao, L., Eaton, J. W., Hogan, D. R., Mahy, M., Marsh, K., … Puckett, R. (2014). Improvements in prevalence trend fitting and incidence estimation in EPP 2013. *AIDS*, *28*, S415. http://doi.org/10.1097/QAD.0000000000000454

Bürkner, P.-C. (2018). Advanced bayesian multilevel modeling with the r package brms. *The R Journal*, *10*(1), 395–411. http://doi.org/10.32614/RJ-2018-017

Centers for Disease Control and Prevention. (2021). *Estimated HIV incidence and prevalence in the united states, 2015–2019* (No. 26(No 1)). Retrieved from `https://www.cdc.gov/hiv/library/reports/hiv-surveillance.html`

Chatzilena, A., Leeuwen, E. van, Ratmann, O., Baguelin, M., & Demiris, N. (2019). Contemporary statistical inference for infectious disease models using stan. *arXiv:1903.00423 [q-Bio, Stat]*. Retrieved from `http://arxiv.org/abs/1903.00423`

Chen, C., Wakefield, J., & Lumley, T. (2014). The use of sampling weights in bayesian hierarchical models for small area estimation. *Spatial and Spatio-Temporal Epidemiology*, *11*, 33–43. http://doi.org/10.1016/j.sste.2014.07.002

Collaborative Group on AIDS Incubation and HIV Survival. (2000). Time from HIV-1 seroconversion to AIDS and death before widespread use of highly-active antiretroviral therapy: A collaborative re-analysis. *The Lancet*, *355*(9210), 1131–1137. http://doi.org/10.1016/S0140-6736(00)02061-4

Conrad, C., Bradley, H. M., Broz, D., Buddha, S., Chapman, E. L., Galang, R. R., … Duwve, J. M. (2015). Community outbreak of HIV infection linked to injection drug use of oxymorphone — indiana, 2015. *Morbidity and Mortality Weekly Report*, *64*(16), 443–444. Retrieved from `https://www.ncbi.nlm.nih.gov/pmc/articles/PMC4584812/`

Cuadros, D. F., Li, J., Branscum, A. J., Akullian, A., Jia, P., Mziray, E. N., & Tanser, F. (2017). Mapping the spatial variability of HIV infection in sub-saharan africa: Effective information for localized HIV prevention and control. *Scientific Reports*, *7*(1), 9093. http://doi.org/10.1038/s41598-017-09464-y

Cuadros, D. F., Oliveira, T. de, Gräf, T., Junqueira, D. M., Wilkinson, E., Lemey, P., … Tanser, F. (2022). The role of high-risk geographies in the perpetuation of the HIV epidemic in rural south africa: A spatial molecular epidemiology study. *PLOS Global Public Health*, *2*(2), e0000105. http://doi.org/10.1371/journal.pgph.0000105

Cuadros, D. F., Sartorius, B., Hall, C., Akullian, A., Bärnighausen, T., & Tanser, F. (2018). Capturing the spatial variability of HIV epidemics in south africa and tanzania using routine healthcare facility data. *International Journal of Health Geographics*, *17*(1), 27. http://doi.org/10.1186/s12942-018-0146-8

Dwyer-Lindgren, L., Cork, M. A., Sligar, A., Steuben, K. M., Wilson, K. F., Provost, N. R., … Hay, S. I. (2019). Mapping HIV prevalence in sub-saharan africa between 2000 and 2017. *Nature*, 1. http://doi.org/10.1038/s41586-019-1200-9

Eaton, J. W., & Bao, L. (2017). Accounting for nonsampling error in estimates of HIV epidemic trends from antenatal clinic sentinel surveillance. *AIDS (London, England)*, *31*, S61–S68. http://doi.org/10.1097/QAD.0000000000001419

Eaton, J. W., Brown, T., Puckett, R., Glaubius, R., Mutai, K., Bao, L., … Hallett, T. B. (2019). The estimation and projection package age-sex model and the r-hybrid model: New tools for estimating HIV incidence trends in sub-saharan africa. *AIDS*, *33*, S235. http://doi.org/10.1097/QAD.0000000000002437

Eaton, J. W., Dwyer-Lindgren, L., Gutreuter, S., O'Driscoll, M., Stevens, O., Bajaj, S., … Shiraishi, R. W. (2021). Naomi: A new modelling tool for estimating HIV epidemic indicators at the district level in sub-saharan africa. *Journal of the International AIDS Society*, *24*, e25788. http://doi.org/10.1002/jia2.25788

Eaton, J. W., Rehle, T. M., Jooste, S., Nkambule, R., Kim, A. A., Mahy, M., & Hallett, T. B. (2014). Recent HIV prevalence trends among pregnant women and all women in sub-saharan africa: Implications for HIV estimates. *AIDS (London, England)*, *28*(4), S507. http://doi.org/10.1097/QAD.0000000000000412

Eisinger, R. W., Dieffenbach, C. W., & Fauci, A. S. (2019). HIV viral load and transmissibility of HIV infection: Undetectable equals untransmittable. *JAMA*, *321*(5), 451–452. http://doi.org/10.1001/jama.2018.21167

Faria, N. R., Rambaut, A., Suchard, M. A., Baele, G., Bedford, T., Ward, M. J., … Lemey, P. (2014). The early spread and epidemic ignition of HIV-1 in human populations. *Science (New York, N.Y.)*, *346*(6205), 56–61. http://doi.org/10.1126/science.1256739

Flaxman, S., Mishra, S., Gandy, A., Unwin, H. J. T., Mellan, T. A., Coupland, H., … Imperial College COVID-19 Response Team. (2020). Estimating the effects of non-pharmaceutical interventions on COVID-19 in europe nature. *Nature*. http://doi.org/10.1038/s41586-020-2405-7

Flaxman, S., Wilson, A. G., Neill, D. B., Nickisch, H., & Smola, A. J. (2015). Fast kronecker inference in gaussian processes with non-gaussian likelihoods. In *Proceedings of the 32Nd international conference on international conference on machine learning - volume 37* (pp. 607–616). JMLR.org. Retrieved from `http://dl.acm.org/citation.cfm?id=3045118.3045184`

Fuglstad, G.-A., Li, Z. R., & Wakefield, J. (2021). The two cultures for prevalence mapping: Small area estimation and spatial statistics. *arXiv:2110.09576 [Stat]*. Retrieved from

http://arxiv.org/abs/2110.09576

Gareta, D., Baisley, K., Mngomezulu, T., Smit, T., Khoza, T., Nxumalo, S., … Herbst, K. (2021). Cohort profile update: Africa centre demographic information system (ACDIS) and population-based HIV survey. *International Journal of Epidemiology*, *50*(1), 33–34. http://doi.org/10.1093/ije/dyaa264

Gareta, D., Dube, S., & Herbst, K. (2020a). AHRI.PIP.men's general health.all.release 2020-07. Africa Health Research Institute (AHRI). http://doi.org/10.23664/AHRI.PIP.RD04-99.MGH.ALL.202007

Gareta, D., Dube, S., & Herbst, K. (2020b). AHRI.PIP.women's general health.all.release 2020-07. Africa Health Research Institute (AHRI). http://doi.org/10.23664/AHRI.PIP.RD03-99.WGH.ALL.202007

Garnett, G. P., & Anderson, R. M. (1994). Balancing sexual partnerships in an age and activity stratified model of HIV transmission in heterosexual populations. *IMA Journal of Mathematics Applied in Medicine and Biology*, *11*(3), 161–192. http://doi.org/10.1093/imammb/11.3.161

Gething, P. W., Patil, A. P., Smith, D. L., Guerra, C. A., Elyazar, I. R., Johnston, G. L., … Hay, S. I. (2011). A new world malaria map: Plasmodium falciparum endemicity in 2010. *Malaria Journal*, *10*, 378. http://doi.org/10.1186/1475-2875-10-378

Ghys, P. D., Williams, B. G., Over, M., Hallett, T. B., & Godfrey-Faussett, P. (2018). Epidemiological metrics and benchmarks for a transition in the HIV epidemic. *PLoS Medicine*, *15*(10), e1002678. http://doi.org/10.1371/journal.pmed.1002678

Gregson, S., Mugurungi, O., Eaton, J., Takaruza, A., Rhead, R., Maswera, R., … Nyamukapa, C. A. (2017). Documenting and explaining the HIV decline in east zimbabwe: The manicaland general population cohort. *BMJ Open*, *7*(10), e015898. http://doi.org/10.1136/bmjopen-2017-015898

Gregson, S., Nyamukapa, C. A., Garnett, G. P., Mason, P. R., Zhuwau, T., Caraël, M., … Anderson, R. M. (2002). Sexual mixing patterns and sex-differentials in teenage exposure to HIV infection in rural zimbabwe. *Lancet (London, England)*, *359*(9321), 1896–1903. http://doi.org/10.1016/S0140-6736(02)08780-9

Grinsztajn, L., Semenova, E., Margossian, C. C., & Riou, J. (2021). Bayesian workflow for disease transmission modeling in stan. *Statistics in Medicine*, *40*(27), 6209–6234. http://doi.org/10.1002/sim.9164

Gutreuter, S., Igumbor, E., Wabiri, N., Desai, M., & Durand, L. (2019). Improving estimates of district HIV prevalence and burden in south africa using small area estimation techniques. *PLOS ONE*, *14*(2), e0212445. http://doi.org/10.1371/journal.pone.0212445

Hall, M., Golubchik, T., Bonsall, D., Abeler-Dörner, L., Limbada, M., Kosloff, B., … Team, on behalf of the H. 071. (PopART). P. protocol team {and}. the P. protocol. (2021). *Demographic characteristics of sources of HIV-1 transmission in zambia* (p. 2021.10.04.21263560). Retrieved from https://www.medrxiv.org/content/10.1101/2021.10.04.21263560v1

Hallett, T. B. (2011). Estimating the HIV incidence rate: Recent and future developments. *Current Opinion in HIV and AIDS*, *6*(2), 102–107.

http://doi.org/10.1097/COH.0b013e328343bfdb

Hallett, T. B., Gregson, S., Lewis, J. J. C., Lopman, B. A., & Garnett, G. P. (2007). Behaviour change in generalised HIV epidemics: Impact of reducing cross-generational sex and delaying age at sexual debut. *Sexually Transmitted Infections*, *83 Suppl 1*, i50–54. http://doi.org/10.1136/sti.2006.023606

Harling, G., Newell, M.-L., Tanser, F., Kawachi, I., Subramanian, S., & Bärnighausen, T. (2014). Do age-disparate relationships drive HIV incidence in young women? Evidence from a population cohort in rural KwaZulu-natal, south africa. *Journal of Acquired Immune Deficiency Syndromes (1999)*, *66*(4), 443–451. http://doi.org/10.1097/QAI.0000000000000198

Hastie, T., & Tibshirani, R. (1986). Generalized additive models. *Statistical Science*, *1*(3). http://doi.org/10.1214/ss/1177013604

Held, L., Höhle, M., & Hofmann, M. (2005). A statistical framework for the analysis of multivariate infectious disease surveillance counts. *Statistical Modelling*, *5*(3), 187–199. http://doi.org/10.1191/1471082X05st098oa

Hoffman, M. D., & Gelman, A. (2011). The no-u-turn sampler: Adaptively setting path lengths in hamiltonian monte carlo. *arXiv:1111.4246 [Cs, Stat]*. Retrieved from `http://arxiv.org/abs/1111.4246`

Hogan, D. R., & Salomon, J. A. (2012). Spline-based modelling of trends in the force of HIV infection, with application to the UNAIDS estimation and projection package. *Sexually Transmitted Infections*, *88*, i52–i57. http://doi.org/10.1136/sextrans-2012-050652

Hunter, D. R., Goodreau, S. M., & Handcock, M. S. (2008). Goodness of fit of social network models. *Journal of the American Statistical Association*, *103*(481), 248–258. http://doi.org/10.1198/016214507000000446

Hunter, D. R., Handcock, M. S., Butts, C. T., Goodreau, S. M., & Morris, M. (2008). Ergm: A package to fit, simulate and diagnose exponential-family models for networks. *Journal of Statistical Software*, *24*(3), nihpa54860.

Hyndman, R. J., & Athanasopoulos, G. (2018). *Forecasting: Principles and practice*. OTexts.

ICAP at Columbia University, & PEPFAR. (2019). Methodology. PHIA. Retrieved June 13, 2019, from `https://phia.icap.columbia.edu/methodology/`

Institut Haïtien de l'Enfance - IHE/Haiti, & ICF. (2018). *Haiti enquête mortalité, morbidité et utilisation des services 2016-2017 - EMMUS-VI [dataset]*. Rockville, Maryland: IHE/Haiti, ICF [Producers]. Retrieved from `http://dhsprogram.com/pubs/pdf/FR326/FR326.pdf`

Ismay, C., & Solomon, N. (2021). *Thesisdown: An updated r markdown thesis template using the bookdown package*.

Jenness, S. M., Goodreau, S. M., & Morris, M. (2018). EpiModel: An r package for mathematical modeling of infectious disease over networks. *Journal of Statistical Software*, *84*. http://doi.org/10.18637/jss.v084.i08

Johnson, L., & Dorrington, R. (2019). *Thembisa version 4.2: A model for evaluating the impact of HIV/AIDS in south africa*. Retrieved from

https://www.thembisa.org/content/downloadPage/Thembisa4_2report

Johnson, N. L. (1949). Systems of frequency curves generated by methods of translation. *Biometrika*, *36*(1), 149. http://doi.org/10.2307/2332539

Joint United Nations Programme on HIV/AIDS, UNAIDS/WHO Working Group on Global HIV/AIDS and STI Surveillance, World Health Organization, & Centers for Disease Control and Prevention (U.S.). (2004). *Guidelines for conducting HIV sentinel serosurveys among pregnant women and other groups*. Geneva: Joint United Nations Program on HIV/AIDS : World Health Organization.

Jones, M. C., & Pewsey, A. (2009). Sinh-arcsinh distributions. *Biometrika*, *96*(4), 761–780. http://doi.org/10.1093/biomet/asp053

Kassanjee, R., McWalter, T. A., & Welte, A. (2014). Short communication: Defining optimality of a test for recent infection for HIV incidence surveillance. *AIDS Research and Human Retroviruses*, *30*(1), 45–49. http://doi.org/10.1089/aid.2013.0113

Kate Grabowski, M., Lessler, J., Bazaale, J., Nabukalu, D., Nankinga, J., Nantume, B., … Gray, R. H. (2020). Migration, hotspots, and dispersal of HIV infection in rakai, uganda. *Nature Communications*, *11*(1), 976. http://doi.org/10.1038/s41467-020-14636-y

Kermack, W. O., & McKendrick, A. G. (1927). A contribution to the mathematical theory of epidemics. *Proceedings of the Royal Society of London. Series A, Containing Papers of a Mathematical and Physical Character*, *115*(772), 700–721. Retrieved from https://www.jstor.org/stable/94815

Kerr, C. C., Stuart, R. M., Gray, R. T., Shattock, A. J., Fraser-Hurt, N., Benedikt, C., … Wilson, D. P. (2015). Optima: A model for HIV epidemic analysis, program prioritization, and resource optimization. *Journal of Acquired Immune Deficiency Syndromes (1999)*, *69*(3), 365–376. http://doi.org/10.1097/QAI.0000000000000605

Kharsany, A. B. M., & Karim, Q. A. (2016). HIV infection and AIDS in sub-saharan africa: Current status, challenges and opportunities. *The Open AIDS Journal*, *10*, 34. http://doi.org/10.2174/1874613601610010034

Kneib, T., & Umlauf, N. (2017). *A primer on bayesian distributional regression* (Working Paper No. 2017-13). Working Papers in Economics; Statistics. Retrieved from https://www.econstor.eu/handle/10419/180164

Kristensen, K., Nielsen, A., Berg, C. W., Skaug, H., & Bell, B. (2016). TMB: Automatic differentiation and laplace approximation. *Journal of Statistical Software*, *70*(5). http://doi.org/10.18637/jss.v070.i05

Krivitsky, P. N., & Handcock, M. S. (2014). A separable model for dynamic networks. *Journal of the Royal Statistical Society. Series B, Statistical Methodology*, *76*(1), 29–46. http://doi.org/10.1111/rssb.12014

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

Krivitsky, P. N., & Morris, M. (2017). INFERENCE FOR SOCIAL NETWORK MODELS FROM EGOCENTRICALLY SAMPLED DATA, WITH APPLICATION TO UNDERSTANDING PERSISTENT RACIAL DISPARITIES IN HIV PREVALENCE

IN THE US. *The Annals of Applied Statistics*, *11*(1), 427–455. http://doi.org/10.1214/16-AOAS1010

Larmarange, J., Vallo, R., Yaro, S., Msellati, P., & Meda, N. (2011). Methods for mapping regional trends of HIV prevalence from demographic and health surveys (DHS). *Cybergeo: European Journal of Geography*, (558). http://doi.org/10.4000/cybergeo.24606

Lewis, J. J. C., Donnelly, C. A., Mare, P., Mupambireyi, Z., Garnett, G. P., & Gregson, S. (2007). Evaluating the proximate determinants framework for HIV infection in rural zimbabwe. *Sexually Transmitted Infections*, *83 Suppl 1*, i61–69. http://doi.org/10.1136/sti.2006.023671

Lindén, A., & Mäntyniemi, S. (2011). Using the negative binomial distribution to model overdispersion in ecological count data. *Ecology*, *92*(7), 1414–1421. http://doi.org/10.1890/10-1831.1

Lindgren, F., Rue, H., & Lindström, 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: Series B (Statistical Methodology)*, *73*(4), 423–498. http://doi.org/10.1111/j.1467-9868.2011.00777.x

Maheu-Giroux, M., Marsh, K., Doyle, C. M., Godin, A., Lanièce Delaunay, C., Johnson, L. F., … Eaton, J. W. (2019). National HIV testing and diagnosis coverage in sub-saharan africa: A new modeling tool for estimating the 'first 90' from program and survey data. *AIDS (London, England)*, *33 Suppl 3*, S255–S269. http://doi.org/10.1097/QAD.0000000000002386

Mahiane, S. G., Marsh, K., Glaubius, R., & Eaton, J. W. (2019). Estimating and projecting the number of new HIV diagnoses and incidence in spectrum's case surveillance and vital registration tool. *AIDS*, *33*, S245. http://doi.org/10.1097/QAD.0000000000002324

Martins, T. G., Simpson, D., Lindgren, F., & Rue, H. (2013). Bayesian computing with INLA: New features. *Computational Statistics & Data Analysis*, *67*, 68–83. http://doi.org/10.1016/j.csda.2013.04.014

Maughan-Brown, B., Evans, M., & George, G. (2016). Sexual behaviour of men and women within age-disparate partnerships in south africa: Implications for young women's HIV risk. *PLOS ONE*, *11*(8), e0159162. http://doi.org/10.1371/journal.pone.0159162

McGillen, J. B., Anderson, S.-J., Dybul, M. R., & Hallett, T. B. (2016). Optimum resource allocation to reduce HIV incidence across sub-saharan africa: A mathematical modelling study. *The Lancet HIV*, *3*(9), e441–e448. http://doi.org/10.1016/S2352-3018(16)30051-0

Meredith, H. R., Giles, J. R., Perez-Saez, J., Mande, T., Rinaldo, A., Mutembo, S., … Wesolowski, A. (2021). Characterizing human mobility patterns in rural settings of sub-saharan africa. *eLife*, *10*, e68441. http://doi.org/10.7554/eLife.68441

Meyer-Rath, G., McGillen, J. B., Cuadros, D. F., Hallett, T. B., Bhatt, S., Wabiri, N., … Rehle, T. (2018). Targeting the right interventions to the right people and places: The role of geospatial analysis in HIV program planning. *AIDS (London, England)*, *32*(8), 957–963. http://doi.org/10.1097/QAD.0000000000001792

Médecins Sans Frontières, Malawi, & Ministry of Health and Population, Chiradzulu District, Malawi. (2004). *Antiretroviral therapy in primary health care: Experience of the chiradzulu programme in malawi* (Case Study). Geneva: World Health Organization. Retrieved from https://www.who.int/hiv/pub/prev_care/en/chiradzulu.pdf

Ministry of Health, Malawi. (2018). *Malawi population-based HIV impact assessment (MPHIA) 2015-2016: Final report*. Lilongwe: Ministry of Health.

Monnahan, C. C., & Kristensen, K. (2018). No-u-turn sampling for fast bayesian inference in ADMB and TMB: Introducing the adnuts and tmbstan r packages. *PLoS ONE*, *13*(5). http://doi.org/10.1371/journal.pone.0197954

Morris, M. (1993). Epidemiology and social networks:: Modeling structured diffusion. *Sociological Methods & Research*, *22*(1), 99–126. http://doi.org/10.1177/0049124193022001005

National Statistical Office/Malawi, & ICF. (2017). *Malawi demographic and health survey 2015-16*. Zomba, Malawi: National Statistical Office; ICF. Retrieved from http://dhsprogram.com/pubs/pdf/FR319/FR319.pdf

National Statistical Office - NSO/Malawi, & ICF Macro. (2011). *Malawi demographic and health survey 2010*. Zomba, Malawi: NSO/Malawi; ICF Macro. Retrieved from http://dhsprogram.com/pubs/pdf/FR247/FR247.pdf

National Statistical Office - NSO/Malawi, & ORC Macro. (2005). *Malawi demographic and health survey 2004*. Calverton, Maryland, USA: NSO/Malawi; ORC Macro. Retrieved from http://dhsprogram.com/pubs/pdf/FR175/FR175.pdf

Novitsky, V., Zahralban-Steele, M., Moyo, S., Nkhisang, T., Maruapula, D., McLane, M. F., … Essex, M. (2020). Mapping of HIV-1C transmission networks reveals extensive spread of viral lineages across villages in botswana treatment-as-prevention trial. *The Journal of Infectious Diseases*, *222*(10), 1670–1680. http://doi.org/10.1093/infdis/jiaa276

Okano, J. T., Sharp, K., Valdano, E., Palk, L., & Blower, S. (2020). HIV transmission and source–sink dynamics in sub-saharan africa. *The Lancet HIV*, *7*(3), e209–e214. http://doi.org/10.1016/S2352-3018(19)30407-2

Over, M. (2012). *Achieving an AIDS transition: Preventing infections to sustain treatment*. Brookings Institution Press.

Pfaff, T., Fortunato, M., Sanchez-Gonzalez, A., & Battaglia, P. W. (2021). Learning mesh-based simulation with graph networks. *arXiv:2010.03409 [Cs]*. Retrieved from http://arxiv.org/abs/2010.03409

Political declaration on HIV and AIDS: On the fast track to accelerating the fight against HIV and to ending the AIDS epidemic by 2030 (2016). Retrieved from https://documents-dds-ny.un.org/doc/UNDOC/GEN/N16/164/34/PDF/N1616434.pdf

R Core Team. (2013). *R: A language and environment for statistical computing*. Vienna, Austria: R Foundation for Statistical Computing. Retrieved from http://www.R-project.org/

R Core Team. (2020). *R: A language and environment for statistical computing*. Vienna, Austria: R Foundation for Statistical Computing. Retrieved from https://www.R-project.org

Ratmann, O., Kagaayi, J., Hall, M., Golubchick, T., Kigozi, G., Xi, X., … Tobian, A. (2020). Quantifying HIV transmission flow between high-prevalence hotspots and surrounding communities: A population-based study in rakai, uganda. *The Lancet HIV*, *7*(3), e173–e183. http://doi.org/10.1016/S2352-3018(19)30378-9

Reniers, G., Slaymaker, E., Nakiyingi-Miiro, J., Nyamukapa, C., Crampin, A. C., Herbst, K., … Network, on behalf of the A. (2014). Mortality trends in the era of antiretroviral therapy: Evidence from the network for analysing longitudinal population based HIV/AIDS data on africa (ALPHA). *AIDS*, *28*, S533. http://doi.org/10.1097/QAD.0000000000000496

Reniers, G., Wamukoya, M., Urassa, M., Nyaguara, A., Nakiyingi-Miiro, J., Lutalo, T., … Zaba, B. (2016). Data resource profile: Network for analysing longitudinal population-based HIV/AIDS data on africa (ALPHA network). *International Journal of Epidemiology*, *45*(1), 83–93. http://doi.org/10.1093/ije/dyv343

Reniers, G., Wamukoya, M., Urassa, M., Nyaguara, A., Nakiyingi-Miiro, J., Lutalo, T., … Zaba, B. (2016). Data resource profile: Network for analysing longitudinal population-based HIV/AIDS data on africa (ALPHA network). *International Journal of Epidemiology*, *45*(1), 83–93. http://doi.org/10.1093/ije/dyv343

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. http://doi.org/10.1177/0962280216660421

Risher, K. A., Cori, A., Reniers, G., Marston, M., Calvert, C., Crampin, A., … Eaton, J. W. (2021). Age patterns of HIV incidence in eastern and southern africa: A modelling analysis of observational population-based cohort studies. *The Lancet HIV*, *8*(7), e429–e439. http://doi.org/10.1016/S2352-3018(21)00069-2

Ritchwood, T. D., Hughes, J. P., Jennings, L., MacPhail, C., Williamson, B., Selin, A., … Pettifor, A. (2016). Characteristics of age-discordant partnerships associated with HIV risk among young south african women (HPTN 068). *Journal of Acquired Immune Deficiency Syndromes (1999)*, *72*(4), 423–429. http://doi.org/10.1097/QAI.0000000000000988

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: Series B (Statistical Methodology)*, *71*(2), 319–392. http://doi.org/10.1111/j.1467-9868.2008.00700.x

Sakarovitch, C., Alioum, A., Ekouevi, D. K., Msellati, P., Leroy, V., & Dabis, F. (2007). Estimating incidence of HIV infection in childbearing age african women using serial prevalence data from antenatal clinics. *Statistics in Medicine*, *26*(2), 320–335. http://doi.org/10.1002/sim.2540

Sanchez-Gonzalez, A., Godwin, J., Pfaff, T., Ying, R., Leskovec, J., & Battaglia, P. W. (2020). Learning to simulate complex physics with graph networks. *arXiv:2002.09405 [Physics, Stat]*. Retrieved from `http://arxiv.org/abs/2002.09405`

Sartorius, B., VanderHeide, J. D., Yang, M., Goosmann, E. A., Hon, J., Haeuser, E., … Dwyer-Lindgren, L. (2021). Subnational mapping of HIV incidence and mortality among individuals aged 15–49 years in sub-saharan africa, 2000–18: A modelling study.

The Lancet HIV, *8*(6), e363–e375. http://doi.org/10.1016/S2352-3018(21)00051-5

Schaefer, R., Gregson, S., Eaton, J. W., Mugurungi, O., Rhead, R., Takaruza, A., … Nyamukapa, C. (2017). Age-disparate relationships and HIV incidence in adolescent girls and young women: Evidence from zimbabwe. *AIDS (London, England)*, *31*(10), 1461–1470. http://doi.org/10.1097/QAD.0000000000001506

Shryock, H. S., Siegel, J. S., & Stockwell, E. G. (1976). *The methods and materials of demography*. Academic Press.

Skaug, H. J., & Fournier, D. A. (2006). Automatic approximation of the marginal likelihood in non-gaussian hierarchical models. *Computational Statistics & Data Analysis*, *51*(2), 699–709. http://doi.org/10.1016/j.csda.2006.03.005

Smid, J. H., Garcia, V., Low, N., Mercer, C. H., & Althaus, C. L. (2018). Age difference between heterosexual partners in britain: Implications for the spread of chlamydia trachomatis. *Epidemics*, *24*, 60–66. http://doi.org/10.1016/j.epidem.2018.03.004

Stover, J., Brown, T., Puckett, R., & Peerapatanapokin, W. (2017). Updates to the spectrum/estimations and projections package model for estimating trends and current values for key HIV indicators. *AIDS*, *31*, S5. http://doi.org/10.1097/QAD.0000000000001322

Subnational Estimates Working Group of the HIV Modelling Consortium. (2016). Evaluation of geospatial methods to generate subnational HIV prevalence estimates for local level planning. *AIDS (London, England)*, *30*(9), 1467–1474. http://doi.org/10.1097/QAD.0000000000001075

The DHS program. (2021). USAID. Retrieved from `https://dhsprogram.com/data/`

The DHS program. (2021). USAID. Retrieved from `https://dhsprogram.com/data/`

Todd, J., Glynn, J. R., Marston, M., Lutalo, T., Biraro, S., Mwita, W., … Żaba, B. (2007). Time from HIV seroconversion to death: A collaborative analysis of eight studies in six low and middle-income countries before highly active antiretroviral therapy. *AIDS (London, England)*, *21*, S55–S63. http://doi.org/10.1097/01.aids.0000299411.75269.e8

Trickey, A., May, M. T., Vehreschild, J.-J., Obel, N., Gill, M. J., Crane, H. M., … Sterne, J. A. C. (2017). Survival of HIV-positive patients starting antiretroviral therapy between 1996 and 2013: A collaborative analysis of cohort studies. *The Lancet HIV*, *4*(8), e349–e356. http://doi.org/10.1016/S2352-3018(17)30066-8

UNAIDS. (2020). AIDSinfo UNAIDS. AIDSinfo. Retrieved December 10, 2021, from `https://aidsinfo.unaids.org/`

UNAIDS. (2021). Global HIV & AIDS statistics — fact sheet. Retrieved January 17, 2022, from `https://www.unaids.org/en/resources/fact-sheet`

United Nations, Department of Economic and Social Affairs, Population Division. (2019). *World population prospects 2019: Methodology of the united nations population estimates and projections* (No. ST/ESA/SER.A/425). New York: United Nations. Retrieved from `https://population.un.org/wpp/Publications/Files/WPP2019_Methodology.pdf`

Valdano, E., Okano, J. T., Colizza, V., Mitonga, H. K., & Blower, S. (2021). Using mobile phone data to reveal risk flow networks underlying the HIV epidemic in namibia. *Nature Communications*, *12*(1), 2837. http://doi.org/10.1038/s41467-021-23051-w

Vehtari, A., Gabry, J., Magnusson, M., Yao, Y., Bürkner, P.-C., Paananen, T., & Gelman, A. (2020). *Loo: Efficient leave-one-out cross-validation and WAIC for bayesian models*. Retrieved from `https://mc-stan.org/loo`

Vehtari, A., Gelman, A., & Gabry, J. (2017). Practical bayesian model evaluation using leave-one-out cross-validation and WAIC. *Statistics and Computing*, *27*(5), 1413–1432. http://doi.org/10.1007/s11222-016-9696-4

Vollmer, M. A. C., Glampson, B., Mellan, T., Mishra, S., Mercuri, L., Costello, C., … Bhatt, S. (2021). A unified machine learning approach to time series forecasting applied to demand at emergency departments. *BMC Emergency Medicine*, *21*(1), 9. http://doi.org/10.1186/s12873-020-00395-y

Wakefield, J., Dong, T. Q., & Minin, V. N. (2017). Spatio-temporal analysis of surveillance data. *arXiv:1711.00555 [Stat]*. Retrieved from `http://arxiv.org/abs/1711.00555`

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. http://doi.org/10.1080/01621459.2012.737729

Wickham, H. (2016). *ggplot2: Elegant graphics for data analysis*. Springer-Verlag New York. Retrieved from `https://ggplot2.tidyverse.org`

Wickham, H. (2016). *ggplot2: Elegant graphics for data analysis*. Springer-Verlag New York. Retrieved from `https://ggplot2.tidyverse.org`

Wolock, T. M., Flaxman, S., Risher, K. A., Dadirai, T., Gregson, S., & Eaton, J. W. (2021). Evaluating distributional regression strategies for modelling self-reported sexual age-mixing. *eLife*, *10*, e68318. http://doi.org/10.7554/eLife.68318

World Health Organization, & UNAIDS. (2015). *Guidelines for conducting HIV surveillance among pregnant women attending antenatal clinics based on routine programme data*. Geneva: World Health Organization. Retrieved from `https://apps.who.int/iris/handle/10665/197864`

World Health Organization, & UNAIDS. (2020, November). Recommended population size estimates of men who have sex with men. World Health Organization. Retrieved from `https://www.unaids.org/en/resources/documents/2020/recommended-population-size-estimates-of-men-who-have-sex-with-men`

Yiannoutsos, C. T., Johnson, L. F., Boulle, A., Musick, B. S., Gsponer, T., Balestre, E., … Egger, M. (2012). Estimated mortality of adult HIV-infected patients starting treatment with combination antiretroviral therapy. *Sexually Transmitted Infections*, *88*, i33–i43. http://doi.org/10.1136/sextrans-2012-050658

Yu, J. K.-L., Chen, S. C.-C., Wang, K.-Y., Chang, C.-S., Makombe, S. D., Schouten, E. J., & Harries, A. D. (2007). True outcomes for patients on antiretroviral therapy who are "lost to follow-up" in malawi. *Bulletin of the World Health Organization*, *85*(7), 550. http://doi.org/10.2471/BLT.06.037739