Bayesian spatio-temporal methods for small-area estimation of HIV indicators

Imperial College London

Adam Howes

Department of Mathematics

Imperial College London

In partial fulfillment of the requirements for the degree of

Doctor of Philosophy

October 2023

Copyright

Statement of Originality

This thesis, and the work presented in it, is work that I conducted myself. In all cases where I describe others' work, I provide appropriate references.

For someone, or something.

Acknowledgements

I would like to start by thanking my supervisors Seth Flaxman and Jeff Eaton for their guidance and mentorship throughout the duration of this thesis. I am grateful for the research environment offered by the Modern Statistics and Statistical Machine Learning Centre for Doctoral Training at Imperial and Oxford, the HIV Inference Group at Imperial, and the Machine Learning and Global Health Network. Thanks to Mike McLaren, Kevin Esvelt, the Nucleic Acid Observatory team, and the Sculpting Evolution lab for hosting my visit to the MIT Media Lab. Thanks to Alex Stringer, and the Department of Statistics and Actuarial Science for hosting my visit to the University of Waterloo. My sense for what matters has been shaped (arguably improved) by the Effective Altruism community. This research was made possible by funding provided by the Bill & Melinda Gates Foundation and EPSRC.

Adam Howes
Imperial College London
October 2023

Abstract

Progress towards ending AIDS as a public health threat by 2030 is faltering. Effective public health response requires accurate, timely, high-resolution estimates of epidemic and demographic indicators. Limitations of available data make obtaining these estimates difficult. I develop and apply Bayesian spatio-temporal methods to meet this challenge. First, I examine models for area-level spatial structure. Second, I estimate district-level HIV risk group proportions, enabling behavioural prioritisation of prevention services, as put forward in the Global AIDS Strategy. Finally, I develop a novel deterministic Bayesian inference method, combining adaptive Gauss-Hermite quadrature with principal component analysis, motivated by the Naomi district-level model of HIV indicators. Together, the contributions in this thesis help to guide precision HIV policy in sub-Saharan Africa, as well as advancing Bayesian methods for spatio-temporal data.

Contents

List of Figures

List of Tables

List of Abbreviations

HIV . . . . . Human Immunodeficiency Virus.

AIDS . . . . . Acquired ImmunoDeficiency Syndrome.

PEPFAR . . . President's Emergency Plan for AIDS Relief.

HIV . . . . . Demographic and Health Surveys.

AIS . . . . . AIDS Indicator Survey.

PrEP . . . . . Pre-Exposure Prophylaxis.

PEP . . . . . Post-Exposure Prophylaxis.

FSW . . . . . Female Sex Worker(s).

MSM . . . . . Men who have Sex with Men.

PWID . . . . . People Who Inject Drugs.

ANC . . . . . Antenatal Clinic.

UNAIDS . . . United Nations Joint Programme on HIV/AIDS.

CDC . . . . . Centers for Disease Control and Prevention.

UAT . . . . . Unlinked Anonymous Testing.

PMTCT . . . Prevention of Mother-to-Child Transmission.

PLHIV . . . . People Living with HIV.

MCMC . . . . Markov Chain Monte Carlo.

VI . . . . . Variational Inference.

INLA . . . . . Integrated Nested Laplace Approximation.

GP . . . . . Gaussian Process.

CAR . . . . . Conditionally Auto-regressive.

ART . . . . . Antiretroviral Therapy.

SAE . . . . . Small Area Estimation.

GMRF . . . . Gaussian Markov Random Field.

HMC . . . . . Hamiltonian Monte Carlo.

GMRF . . . . Gaussian Markov Random Field.

HMC . . . . . Hamiltonian Monte Carlo.

LGM . . . . . Latent Gaussian Model.

ELGM . . . . . Extended Latent Gaussian Model.

List of Notations

ρ	HIV prevalence.
λ	HIV incidence.
α	ART coverage.
$\mathcal{S}$	Spatial study region $\mathcal{S} \subseteq \mathbb{R}^2$.
$s \in \mathcal{S}$	Point location.
$\mathcal{T}$	Temporal study period $\mathcal{T} \subseteq \mathbb{R}$.
$t \in \mathcal{T}$	Time.
$\mathbf{y}$	Data, a n-vector.
$\boldsymbol{\phi}$	Parameters, a d-vector $(\phi_1, \ldots, \phi_d)$.
$\mathbf{x}$	Latent field, a N-vector $(x_1, \ldots, x_N)$.
$\boldsymbol{\theta}$	Hyperparameters, a m-vector $(\theta_1, \ldots, \theta_m)$.
$x \sim p(x)$	x is distributed according to $p(x)$.
A_i	Areal unit.
$A_i \sim A_j$	Adjacency between areal units.
$\mathbf{H}$	Hessian matrix.
$\mathbf{R}$	Structure matrix.
$\mathbf{Q}$	Precision matrix.
$\mathbf{\Sigma}$	Covariance matrix.
$\mathcal{N}$	Gaussian distribution.
$k : \mathcal{X} \times \mathcal{X} \to \mathbb{R}$	Kernel function on the space $\mathcal{X}$.
$A_i \sim A_j$	Adjacency between areal units.
$\mathcal{Q}$	A set of quadrature nodes.
$\omega : \mathcal{Q} \to \mathbb{R}$ . . .	A quadrature weighting function.
$\mathcal{Q}(m, k)$	Gauss-Hermite quadrature points in m dimensions with k nodes per dimension, constructed according to a product rule.

1 Introduction

This thesis is about applied and methodological Bayesian statistics. It is applied and methodological in that I am concerned with real world questions and the means to answer them. It is Bayesian in that I arrive at conclusions based on data using probability models.

The real world questions relate to surveillance of the human immunodeficiency virus (HIV) epidemic in sub-Saharan Africa (SSA). HIV is the largest cause of disability adjusted life years (DALYs) in SSA among those one year and older [Global Burden of Disease Collaborative Network (2019); Figure 1.1]. Using statistics to quantify the epidemic is an important part of the public health response, and the path towards disease control and elimination. However, there are substantial challenges involved in obtaining suitable estimates of relevant indicators.

The estimates in this thesis are based on data recorded from national household surveys or routinely collected from healthcare facilities. An important feature of this data is its location and the time at which it was collected. While diverse, spatio-temporal data have distinctive commonalities which reoccur across settings. As such, I draw on modelling techniques from spatio-temporal statistics.

Computation is an essential part of modern statistical practice. Each project in this thesis, as well as the thesis itself, is accompanied by R code, hosted on GitHub.

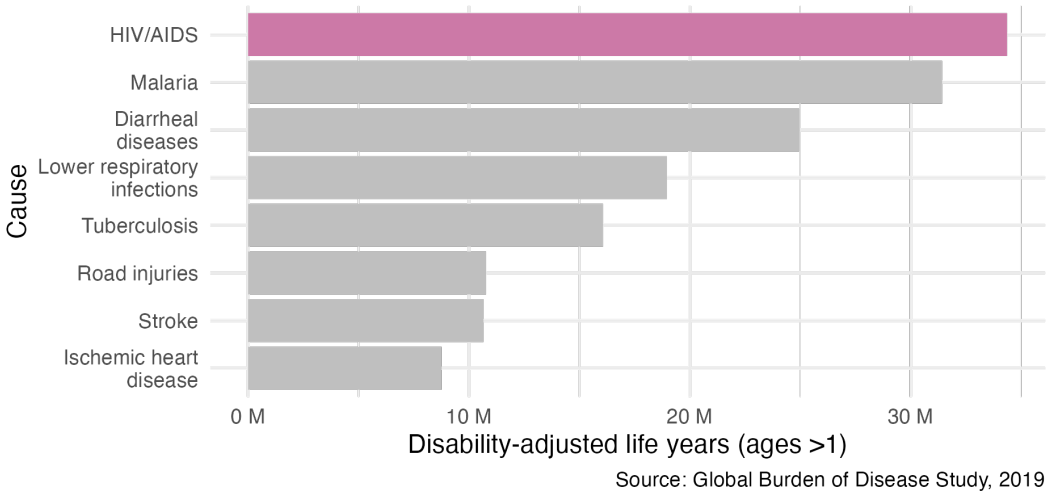

Figure 1.1: HIV/AIDS is the largest cause of DALYs for non-infacts (>1 years) in SSA. One DALY represents the loss of the equivalent of one year of full health, and is calculated by the sum of years of life lost and years lost due to disability. The disability weights used by the Global Burden of Disease Collaborative Network (2019) vary depending on severity of the condition.

1.1 Chapter overview

The structure of this thesis (Figure 1.2) is as follows:

- Chapter 2: I begin by providing background on the HIV/AIDS epidemic, as well as describing the challenges faced by surveillance efforts.
- Chapter 3: I then introduce the statistical concepts and notation used throughout the thesis, focusing on Bayesian modelling and computation, spatio-temporal statistics, and survey methods.
- Chapter 4: The prevailing model for spatial structure used in small-area estimation (Besag et al. 1991) was designed with analysis of a grid of pixels in mind. In disease mapping, we work using the districts of a country, which are not a grid. I evaluate the practical consequences of this this concern (Howes, Eaton, et al. 2023+).
- Chapter 5: Adolescent girls and young women are a demographic group at disproportionate risk of HIV infection. The Global AIDS Strategy suggests prioritising interventions on the basis of behaviour to prevent the most new infections using available resources. I estimate the size of behavioural risk

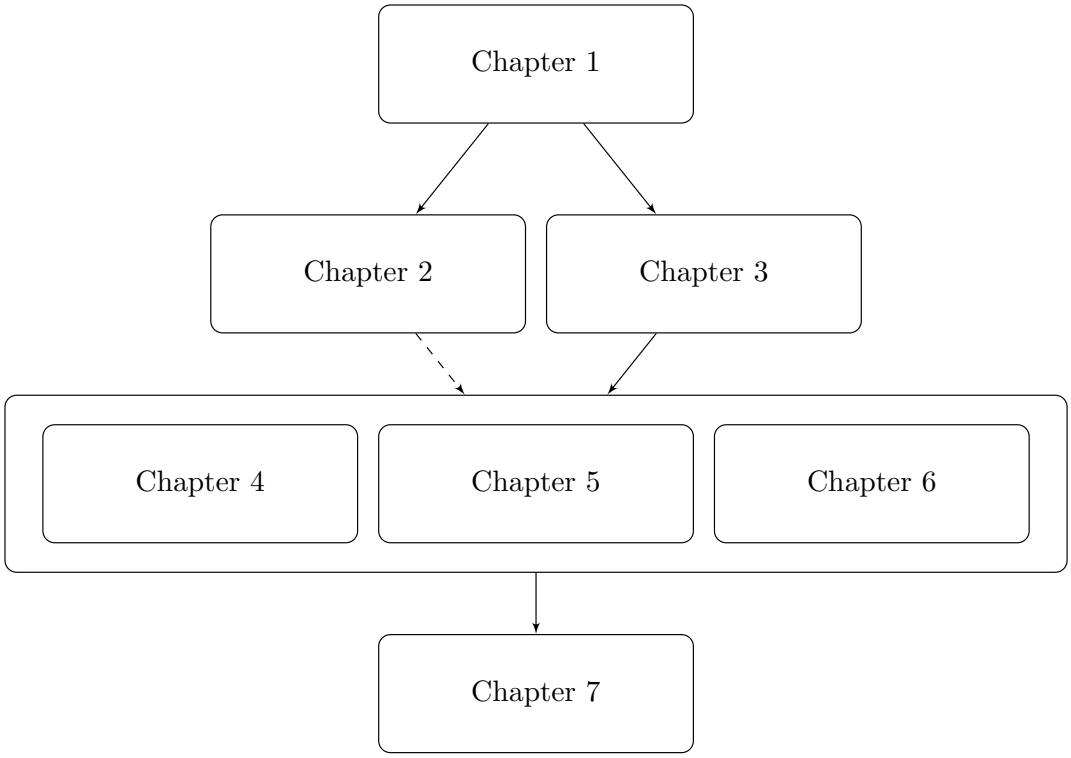

Figure 1.2: The chapters in this thesis are structured to depend on each other. Dashed lines represents a recommended, but not required dependency. Though chronological order is recommended, Chapter 4, Chapter 5 and Chapter 6 may be read in any order as they correspond to separable research projects.

groups across priority countries to enable implementation of this strategy, and assess the potential benefits in terms of numbers of new infections prevented (Howes, Risher, et al. 2023).

- Chapter 6: The Naomi small-area estimation model (Eaton et al. 2021) is used by countries to estimate district-level HIV indicators. With this motivation, I develop an approximate Bayesian inference method combining adaptive Gauss-Hermite quadrature with principal components analysis (Howes, Stringer, et al. 2023+). I apply the method to data from Malawi, and analyse the consequences of inference method choice for policy relevant outcomes. Further, I open the door to a new class of fast, flexible, and accurate Bayesian inference algorithms.
- Chapter 7: Finally, I discuss avenues for future work, and my conclusions regarding the research, as well as its strengths and weaknesses.

2

The HIV/AIDS epidemic

2.1 Background

HIV is a retrovirus which infects humans. If untreated, HIV develop into a more advanced stage known as acquired immunodeficiency syndrome (AIDS). HIV primarily attacks a type of white blood cell vital for the function of the immune system. As a result, AIDS is characterised by increased risk of developing opportunistic infections such as tuberculosis or *Pneumocystis* pneumonias.

The first AIDS cases were reported in Los Angeles in the early 1980s (Gottlieb et al. 1981; Barré-Sinoussi et al. 1983). Since then, HIV has spread globally. Transmission occurs by exposure to specific bodily fluids of an infected person. The most common mode of transmission is via unprotected anal or vaginal sex, though transmission can also occur from a mother to her baby, or when drug injection equipment is shared. Approximately 86 million people have become infected with HIV, and of those 40 million have died of AIDS-related causes.

An ongoing global and multifaceted effort has been made to respond to the epidemic. Groups that have played pivotal roles in shaping the response include local communities, civil society organisations, governments, research institutions, pharmaceutical companies, international agencies like the Joint United Nations Programme on HIV/AIDS (UNAIDS), and global health initiatives such like the

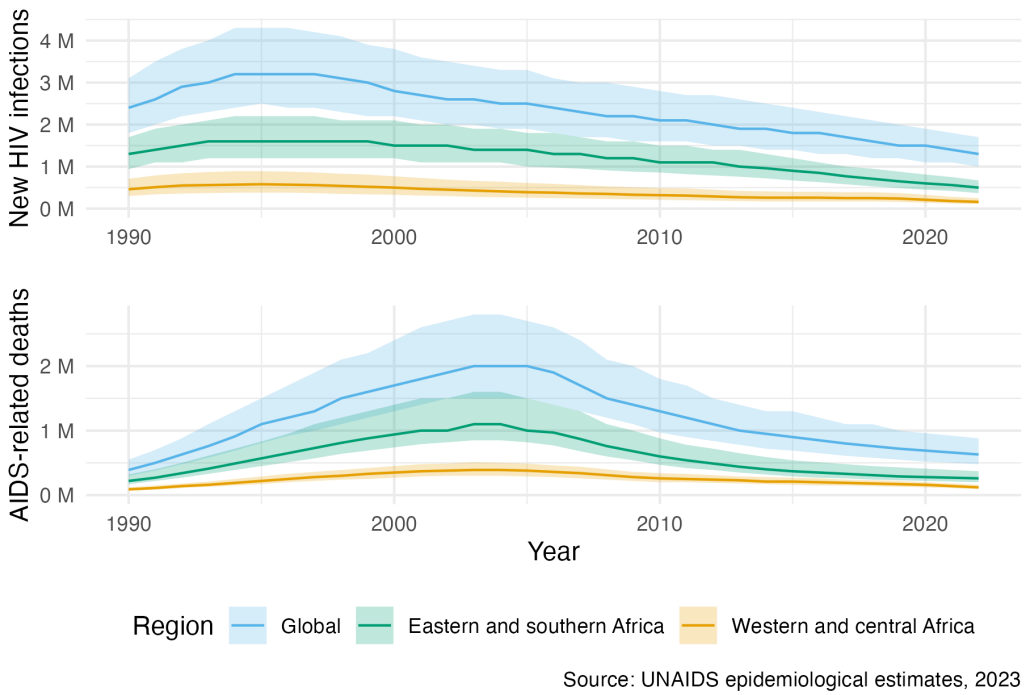

Figure 2.1: Globally, yearly new HIV infections peaked in 1995, and have since decreased by 59% and yearly AIDS-related deaths peaked in 2004, and have since decreased by 68% (UNAIDS 2023a). Much of the disease burden is concentrated in eastern and southern Africa, as well as western and central Africa.

President's Emergency Plan for AIDS Relief (PEPFAR) and the Global Fund to Fight AIDS, Tuberculosis, and Malaria (GFATM). The investment of $100 billion by PEPFAR constituting the "largest commitment by a single nation to address a single disease in history" indicates the scale of the response (U.S. Department of State 2022).

A substantial impact has been made on the course of the epidemic. The number of new HIV infections and AIDS-related deaths per year have both fallen significantly since their peak (Figure 2.1). This progress comes due to implementation HIV prevention and treatment options. I descibe the most significant of these below.

- Antiretroviral therapy (ART) is a drug which stops the virus from replicating in the body. A person living with HIV who takes ART daily can live a full and healthy life. Of the 39 million people living with HIV (PLHIV) in 2022, around 76% were accessing ART. A staggering 21 million AIDS-related

deaths are estimated to have been averted by ART (UNAIDS 2023b). ART reduces the amount of virus in the blood and genital secretions. If the virus is undetectable then there is considerable evidence that it cannot be transmitted sexually (Cohen et al. 2011). For this reason, as well as providing life saving treatment, ART also operates as prevention (TaSP). Particular efforts have been made to provide pregnant women with ART, which is estimated to reduce the chance of mother-to-child transmission (MTCT) by 25-35% in some settings (Siegfried et al. 2011).

- Condoms are an inexpensive and effective method for prevention of HIV and other sexually transmitted infections (STIs) such as *Chlamydia trachomatis*, *Neisseria gonorroeae*, syphilis, and *Trichomonas vaginalis*. There has been a significant scale-up of condom usage since 1990, which is estimated to have averted 117 million new HIV infections (Stover and Teng 2021). That said, there remain significant and difficult to close gaps in the proportion of individuals reporting they used a condom during their last high-risk sexual encounter.
- Voluntary medical male circumcision (VMMC) has been found to provide partial protection against female-to-male HIV transmission. Three landmark randomised control trials (Auvert et al. 2005; Gray et al. 2007; Bailey et al. 2007) found complete surgical removal of the foreskin to result in a 50-60% reduction of HIV incidence in men. Based on this evidence, VMMC has been recommended since 2007 by the World Health Organization (WHO) and UNAIDS as a key HIV intervention in high-prevalence settings. Scale up of VMMC across 15 priority countries between 2008 and 2019 is estimated to have already averted 340 thousand new HIV infections, though the future number of new HIV infections averted is likely to be much higher.
- Pre-exposure prophylaxis (PrEP) and post-exposure prophylaxis (PEP) are drugs which can be taken before and after exposure to prevent transmission. More costly than some other prevention options, they are primarily useful in high risk settings.

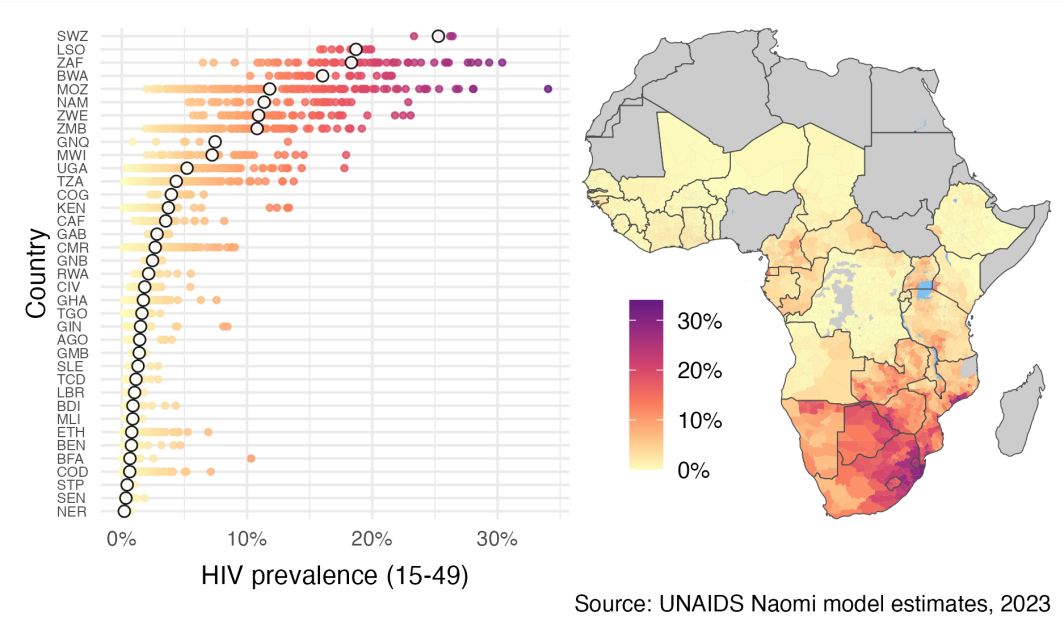

Figure 2.2: Adult (15-49) HIV prevalence varies substantially both within and between countries in SSA. These estimates from 2023 were generated by country teams using the Naomi small-area estimation model in a process supported by UNAIDS and are available from UNAIDS (2023a). White filled points are country-level estimates, and coloured points are district-level estimates. Results from Nigeria are yet to be approved and have been redacted. The estimates process in the Cabo Delgado province of Mozambique was disrupted by conflict. Obtaining results for the Democratic Republic of the Congo required removing some districts from the model. Country names are given by three-leter codes as published by the International Organization for Standardization (ISO).

Though progress had been made, there remains much more to do. In 2022 there were 1.3 million people newly infected with HIV and 630 thousand AIDS-related deaths, more than one every minute (UNAIDS 2022). Bold fast-track targets have been set to accelerate the end of AIDS as global public health threat by 2030. Meeting these targets in the context of disruption to HIV services caused by the COVID-19 pandemic and a shortfall in HIV funding (Economist Impact 2023) requires renewed commitment.

For available resources to have the greatest impact, it is important that HIV interventions are prioritised. Under the precision public health paradigm, the right interventions should be provided to the right populations, in the right place, at the right time (Khoury et al. 2016). Some interventions might be orders of magnitude

more impactful than others (Ord 2013).

Disease burden varies substantially across multiple spatial scales. In some countries, the epidemic is concentrated within small populations, and HIV prevalence is low. In others, transmission is sustained in the general population, and HIV prevalence is higher. Most of the countries severely affected by HIV are in sub-Saharan Africa (SSA). It is estimated that 66% of the 39 million PLHIV worldwide live in SSA. Adult HIV prevalence (ages 15-49) is higher than 10% (Figure 2.2) in some countries in southern Africa. Just as there is variation between countries, there is variation within countries. For example, adult HIV prevalence at the district municipality level in South Africa ranges from 6% in Namakwa to 30% in uMkhanyakude.

In all countries and contexts, some groups of people are at much higher risk than others. Groups of people at increased risk of HIV infection are known as key populations (KPs). Examples include men who have sex with men (MSM), female sex workers (FSW), people who inject drugs (PWID), and transgender people (TGP) (Stevens, Sabin, Garcia, et al. 2022). KPs are often marginalised, and face legal and social issues. In concentrated settings, the majority of new HIV infections occur in KPs and their sexual partners. In generalised settings like SSA, risk is more diffuse across the population. For example, in SSA adolescent girls and young women (AGYW) are a large demographic group at increased risk of HIV infection (Risher et al. 2021; Monod et al. 2023), but typically not considered a KP.

2.2 HIV surveillance

HIV surveillance refers to the collection, analysis, interpretation and dissemination of data relating to HIV/AIDS. Surveillance can used to track epidemic indicators, identify at-risk populations, find drivers of transmission, and evaluate the impact of prevention and treatment programs. Important indicators include:

- **HIV prevalence** is the proportion $\rho \in [0, 1]$ of the population who have HIV, typically written as a percentage. Both new infections and more PLHIV

remaining alive by taking treatment result in an increase in HIV prevalence. As such, HIV prevalence should not be interpreted in isolation, and is primarily used indirectly to calculate other indicators, rather than directly in policy. In some circumstances, when other indicators are difficult to estimate, HIV prevalence can be a useful proxy. The number of PLHIV is given by $N\rho$, where N is the population size.

- **HIV incidence** is the rate $\lambda \in \mathbb{R}$ of new HIV infections, typically written as number of new infections per 1000 person years. HIV incidence can be specified in terms of HIV prevalence by $\lambda = N\Delta\rho$, where $\Delta\rho$ is the change in HIV prevalence over some time period. Planning, delivery, and evaluation of prevention programming relies on estimates of HIV incidence.
- **ART coverage** is the proportion $\alpha \in [0, 1]$ of PLHIV who are on ART, typically written as a percentage. Estimates of ART coverage play a direct role in the provision and target setting for treatment services.

2.2.1 Challenges

Obtaining reliable, timely, estimates of these indicators at an appropriate spatial resolution is challenging. Among the most significant difficulties are:

1. **Data sparsity** Collection of data is costly and time consuming. As a result, limited direct data might be available for the particular time, location, and sub-population of interest. For example, in many countries the last household survey conducted is several years out of date.
2. **Missing data** The sampling frame may not correspond to the target population. For example, many KPs are difficult to reach, and may be ommited from sampling frames. Individuals included on the sampling frame may choose not to respond. All surveys are subject to sampling error, as only a subset of the target population are sampled. Each of these issues can be characterised as being due to missing data. I characterise data sparsity as referring to limited study availability, and missing data as referring to the shortfalls of any given study.

3. **Survey biases** Individuals may be hesitant to disclose their HIV status, or report higher risk behaviours, due to social desirability bias and a fear of discrimination or stigma. When available, biomarker data can be used to overcome under-reporting, but still may be subject to measurement errors.
4. **Denominators and demography** Many indicators are rates or proportions, which rely on estimates of the population at risk in the denominator. Accurately estimating population denominators is itself a challenging task (Tatem 2017). Taking a ratio of uncertain quantities amplifies uncertainty, but is rarely properly accounted for.
5. **Reliance on epidemiological parameters** Indicators rely on estimates of epidemiological parameters obtained from cohort studies. These parameters may not generalise to the setting of interest. Further, they are typically applied coarsely, and without proper accounting for uncertainty.

2.2.2 Statistical approaches

The challenges above make direct interpretation of the data often misleading. Careful statistical modelling is required to overcome these limitations.

1. **Borrowing information** When little direct data is available, data judged to be indirectly related can be used to help improve estimation. For example, if limited data is available for men aged 30-34 is a particular country, it is likely reasonable to make use of the data for men aged 35-39. As well as over age groups, information can be borrowed between and within countries, and across times.
2. **Evidence synthesis** Multiple sources of evidence can be combined to overcome the limitations of any one data source. For example, infrequently run household surveys can be complimented by up to date programmatic data.
3. **Expert guidance** Expert epidemiological, demographic, and local stakeholder guidance can be used to improve estimates. Ensuring the quality of any data used in the estimation process is essential.

2.2.3 Future directions

Aims for HIV response going forward, and the surveillance capabilities needed to meet them.

1. **Greater reliance on routine health system data** It is not recommended to include HIV testing in nationally representative household surveys in low (<2%) HIV prevalence settings (World Health Organization 2005). Patient-level HIV data systems (World Health Organization 2017). Case-based surveillance (CBS).
2. **Integration with other health programs** Strengthen health systems.

3

Bayesian spatio-temporal statistics

3.1 Bayesian statistics

Bayesian statistics is a mathematical paradigm for learning from data. It is well suited for to the challenges posed by Section 2.2 because it allows principled and flexible integration of data and scientific knowledge. In this section I provide brief overview. For a more complete introduction, I recommend McElreath (2020) or Gelman, Carlin, et al. (2013).

3.1.1 Bayesian modelling

At its best, the Bayesian paradigm allows the analyst focus on how best to model the data. This is achieved by the construction of a generative model $p(\mathbf{y}, \boldsymbol{\phi})$ for the observed data $\mathbf{y} = (y_1, \ldots, y_n)$ together with parameters $\boldsymbol{\phi} = (\phi_1, \ldots, \phi_d)$. Here n is the dimension of the data and d is the number of parameters. The model is generative in the sense that one can simulate from it to generate draws

$$(\mathbf{y}, \boldsymbol{\phi}) \sim p(\mathbf{y}, \boldsymbol{\phi}) \tag{3.1}$$

If these draws differ too greatly from what the analyst would expect, then the generative model does not capture their scientific understanding, and can be refined. In this way, models can be built iteratively, with complexity added gradually.

The model is usually constructed from two parts, known respectively as the likelihood $p(\mathbf{y} \mid \boldsymbol{\phi})$ and the prior distribution $p(\boldsymbol{\phi})$. The joint distribution is obtained by the product $p(\mathbf{y}, \boldsymbol{\phi}) = p(\mathbf{y} \mid \boldsymbol{\phi})p(\boldsymbol{\phi})$. The likelihood, as a function of $\boldsymbol{\phi}$ with $\mathbf{y}$ fixed, reflects the probability of observing the data when the true value of the parameters is $\boldsymbol{\phi}$. The prior distribution encapsulates beliefs about the parameters $\boldsymbol{\phi}$ before the data is observed.

Recommendations for specifying the prior distribution vary. A central issue is the extent to which subjective information should be incorporated into the prior distribution, and thereby influence the posterior distribution. Proponents of the objective Bayesian paradigm suggest that the prior distribution should be non-informative, so as not to introduce subjectivity into the analysis. That said, we shall see that the distinction between likelihood and prior distribution can be blurred (Section 3.3). As such, it may be argued that issues of subjectivity are not unique to the prior distribution, and ultimately the challenge of specifying the data generating process is better thought of more holistically (Gelman, Simpson, et al. 2017).

3.1.2 Bayesian computation

The posterior distribution $p(\boldsymbol{\phi} \mid \mathbf{y})$ encapsulates probabilistic beliefs about the parameters given the observed data. Using the eponymous Bayes' theorem, it is given by

$$p(\boldsymbol{\phi} \mid \mathbf{y}) = \frac{p(\mathbf{y} \mid \boldsymbol{\phi})p(\boldsymbol{\phi})}{p(\mathbf{y})}. \tag{3.2}$$

Unfortunately, it is usually intractable to calculate Equation (3.2) directly because of the potentially high-dimensional integral $p(\mathbf{y}) = \int p(\mathbf{y}, \boldsymbol{\phi}) \mathrm{d}\boldsymbol{\phi}$ in the denominator, sometimes known as the marginal likelihood or evidence. As such, although it is easy to evaluate $p(\boldsymbol{\phi} \mid \mathbf{y}) \propto p(\mathbf{y} \mid \boldsymbol{\phi})p(\boldsymbol{\phi})$, it is typically difficult to evaluate the posterior distribution itself.

A variety of computational methods have been developed to tackle this problem. The main categories are:

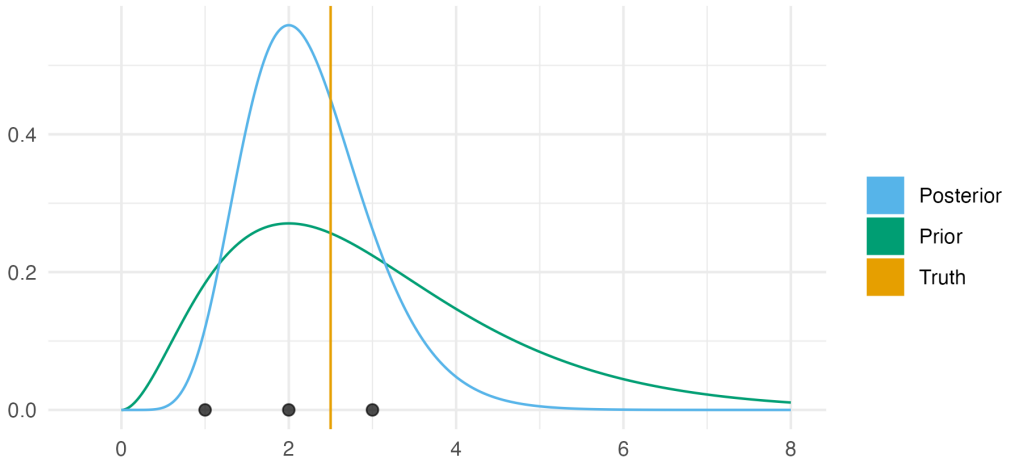

Figure 3.1: An example of Bayesian modelling and computation for a simple one parameter model. The likelihood is $y_i \sim \text{Poisson}(\phi)$ for $i = 1, 2, 3$ and prior distribution is $\phi \sim \text{Gamma}(3, 1)$. Given observed data $\mathbf{y} = (1, 2, 3)$ the posterior distribution is available in closed form as $\text{Gamma}(9, 4)$. This is because the model is conjugate and the posterior distribution is in the same family as the prior distribution. Conjugate models are frequently used because of their convenience, in preference to other perhaps more appropriate models which might be more computationally demanding.

1. **Sampling algorithms** These approaches look to generate samples from the posterior distribution, and are known as Monte Carlo methods, named after the casino (Robert and Casella 2005). The most popular is Markov chain Monte Carlo (MCMC), which proceeds by simulating from a Markov chain with the posterior distribution as its stationary distribution. In this thesis, I make use of the No-U-Turn sampler [NUTS; Hoffman, Gelman, et al. (2014)], a Hamiltonian Monte Carlo [HMC; Duane et al. (1987); Neal et al. (2011)] algorithm, implemented in the Stan (Carpenter et al. 2017) probabilistic programming language (PPL).
2. **Variational inference** In variational inference (VI), the posterior distribution is assumed to belong to a particular class of functions. Optimisation algorithms are then used to choose the best member of that class. Though VI is fast, it may not be accurate (Yao et al. 2018).
3. **Expectation maximisation**
4. **Deterministic approximations** These approximations are the subject of Chapter 6.

3.1.3 Interplay between modelling and computation

Bayesian computation aspires to abstract away calculation of the posterior distribution from the analyst. Modern computational techniques and software have made this aspiration a reality for many models. However, computation of the posterior distribution remains intractable for a majority of models. As such, the analyst need not only to be concerned with choosing a model suitable for the data, but also choosing a model for which the posterior distribution is tractable in reasonable time. As such, there is an important interplay between modelling and computation, wherein models are bound by the limits of computation. As computation improves, the space of models available to the analyst expands.

3.2 Spatio-temporal statistics

In spatio-temporal statistics (Cressie and Wikle 2015), we observe data indexed by spatial or temporal location. In this thesis we assume that the spatial study region $\mathcal{S} \subseteq \mathbb{R}^2$ has two dimensions, corresponding to latitude and longitude. Data may be associated to a point $s \in \mathcal{S}$ or area $A \subseteq \mathcal{S}$ in the study region. The temporal study period $\mathcal{T} \subseteq \mathbb{R}$ can more generally be assumed to be one dimensional. Similarly, data may be associated to a point $t \in \mathcal{T}$ or period of time $T \in \mathcal{T}$.

Spatio-temporal data has some important properties. Foremost among them is correlation structure. Tobler's first law of geography, also expressed by Fisher (1936), is that "everything is related to everything else, but near things are more related than distant things" (Tobler 1970). This law extends not only to space but also time.

3.3 Model classes

3.3.1 Hierarchical models

Often it is natural to group some observations together. Bayesian hierarchical or multilevel models, comprised of multiple stages, allow for natural handling data of this sort, even with complex nested or crossed grouping structures. In a three-stage

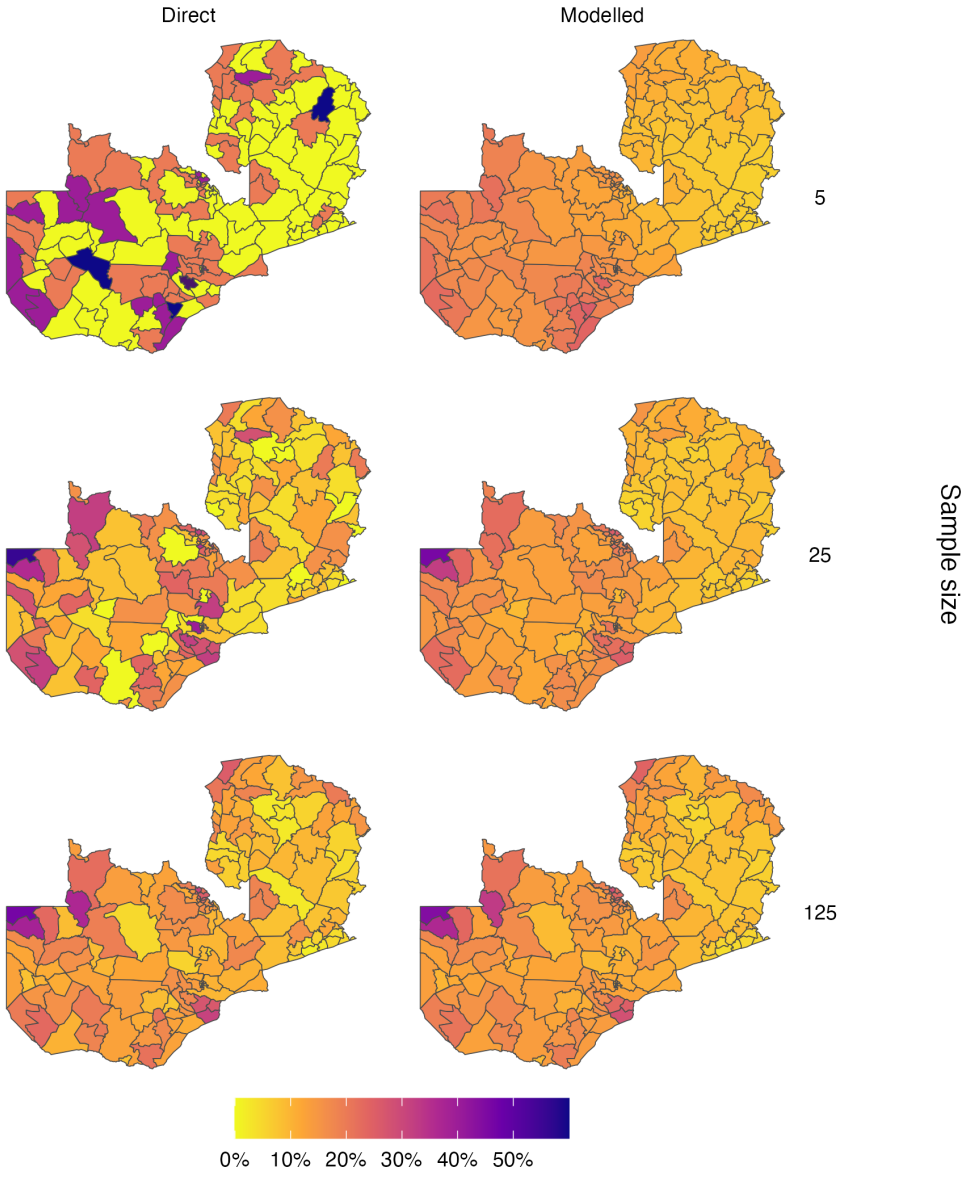

Figure 3.2: To demonstrate the benefits of spatial modelling, I simulated simple random samples with varying sample size in each of the 156 constituencies of Zambia. I then calculated direct and modelled estimates for each survey. The model was a logistic regression with linear predictor given by an intercept and a Besag spatial random effect. HIV estimates for Zambia have previously been generated at the district-level comprising 116 spatial units. Moving forward, there is interest in generating estimates at the constituency level, as program planning is more locally devolved. This figure is adapted from a presentation I gave for the Zambia HIV Estimates Technical Working Group, available from `athowes/zambia-unaids`.

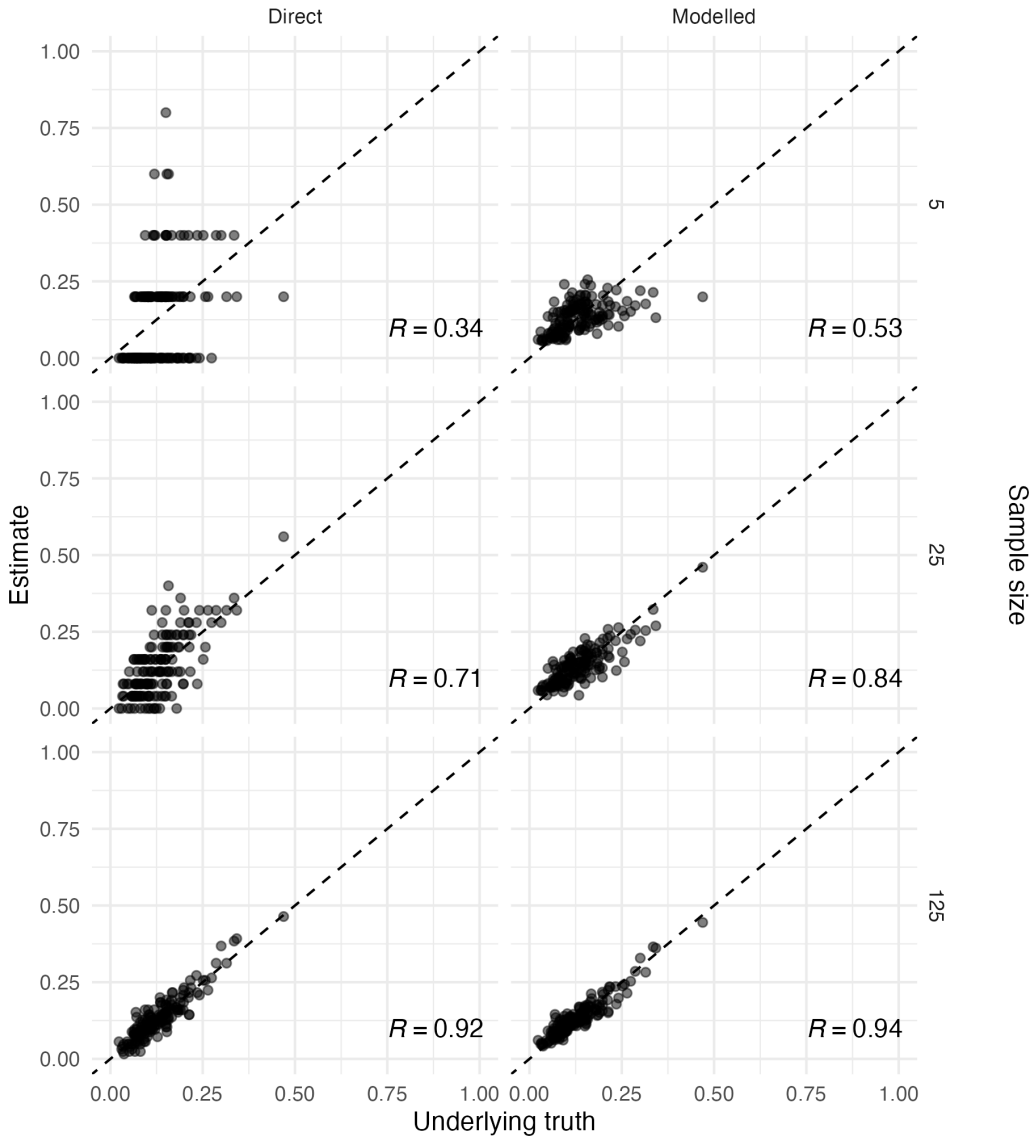

Figure 3.3: The estimates from surveys with higher sample size have higher Pearson correlation coefficient (R) with the underlying truth. For a fixed sample size, correlation can be improved by using modelled estimates to borrow information across spatial units, rather than using the higher variance direct estimates. Points along the dashed diagonal line correspond to agreement between the estimate obtained from the survey and the underlying truth used to generate the data. The setting matches that of Figure 3.2.

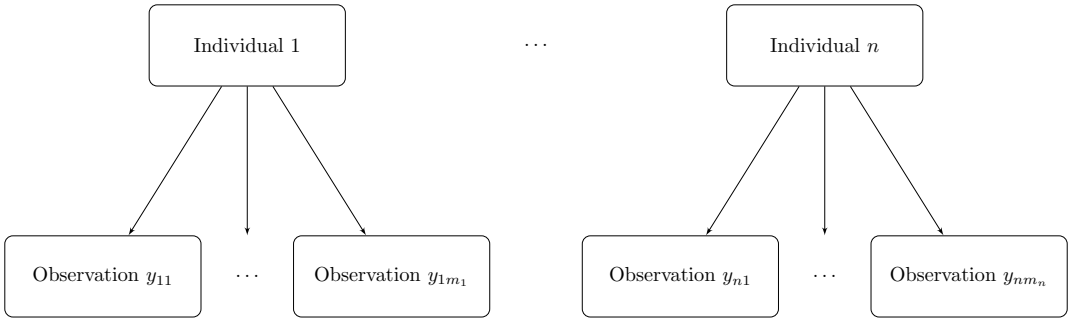

Figure 3.4: A simple example of group structure within data. Each individual $i = 1, \ldots, n$ is associated to m_i observations $y_{i1}, \ldots, y_{im_i}$.

hierarchical model, we partition the parameters so that $\boldsymbol{\phi} = (\mathbf{x}, \boldsymbol{\theta})$. I refer to $\mathbf{x} = (x_1, \ldots, x_n)$ as the latent field, and $\boldsymbol{\theta} = (\theta_1, \ldots, \theta_m)$ as the hyperparameters. The generative model for data $\mathbf{y}$ is then

$$\mathbf{y} \sim p(\mathbf{y} \,|\, \mathbf{x}, \boldsymbol{\theta}), \tag{3.3}$$

$$\mathbf{x} \sim p(\mathbf{x} \,|\, \boldsymbol{\theta}), \tag{3.4}$$

$$\boldsymbol{\theta} \sim p(\boldsymbol{\theta}), \tag{3.5}$$

with posterior distribution proportional to $p(\mathbf{x}, \boldsymbol{\theta} \,|\, \mathbf{y}) \propto p(\mathbf{y} \,|\, \mathbf{x}, \boldsymbol{\theta}) p(\mathbf{x} \,|\, \boldsymbol{\theta}) p(\boldsymbol{\theta})$.

For example, Figure 3.4 illustrates a case where each individual $i = 1, \ldots, n$ in a study is observed m_i times. Observations $y_{i1}, \ldots y_{im_i}$ of the same individual are grouped together, and are more likely to have more similar properties than observations of different individuals. Hierarchical models often control over if and how information is shared between groups.

1. **Complete pooling** In this model, the group structure is ignored and all $\sum_{i=1}^{n} m_i$ observations are treated as independent and identically distributed

$$y_{ij} \sim \mathcal{N}(\mu, \sigma), \tag{3.6}$$

$$(\mu, \sigma) \sim p(\mu, \sigma). \tag{3.7}$$

2. **No pooling** Alternatively, the groups can be modelled entirely separately with group specific mean μ_i and standard deviation σ_i parameters

$$y_{ij} \sim \mathcal{N}(\mu_i, \sigma_i), \tag{3.8}$$

$$(\mu_i, \sigma_i) \sim p(\mu_i, \sigma_i). \tag{3.9}$$

3. **Partial pooling** In this model, some amount of information is shared between the groups

$$y_{ij} \sim \mathcal{N}(\mu_i, \sigma), \tag{3.10}$$

$$\mu_i = \beta + u_i, \tag{3.11}$$

$$\beta \sim p(\beta), \tag{3.12}$$

$$\mathbf{u} \sim p(\mathbf{u}), \tag{3.13}$$

$$\sigma \sim p(\sigma), \tag{3.14}$$

where the vector $\mathbf{u} = (u_1, \ldots, u_n)$.

3.3.2 Mixed effects models

Fixed effects refer to those elements of the latent field which are constant across groups. Random effects refer to those elements of the latent field which vary across groups. These terms have notoriously many different, and incompatible, definitions which can cause confusion (Gelman 2005). I nonetheless find them useful to introduce here.

For concreteness, in the partial pooling model above, the latent field is $\mathbf{x} = (\beta, u_1, \ldots, u_n)$. The scalar β is a fixed effect which applies to all n groups. The vector $\mathbf{u}$ are random effects which alter the mean differently for each group. The only hyperparameter is the standard deviation $\theta = \sigma$.

Random effects can also be structured to share information between some groups more than others. In spatio-temporal statistics, structured spatial and temporal random effects are often used to impose smoothness. Spatial random effects are the subject of Chapter 4.

3.3.3 Latent Gaussian models

Latent Gaussian models [LGMs; Rue, Martino, et al. (2009)] are a class of three-stage Bayesian hierarchical models in which the middle layer is Gaussian. To be

more precise, in an LGM, the likelihood is given by

$$\begin{aligned} y_i &\sim p(y_i \,|\, \eta_i, \boldsymbol{\theta}_1), \quad i = 1, \ldots, n, \\ \mu_i &= \mathbb{E}(y_i \,|\, \eta_i) = g(\eta_i), \\ \eta_i &= \beta_0 + \sum_{l=1}^{p} \beta_j z_{ji} + \sum_{k=1}^{r} f_k(u_{ki}). \end{aligned}$$

The likelihood is given by a product $p(\mathbf{y} \,|\, \boldsymbol{\eta}, \boldsymbol{\theta}_1) = \prod_{i=1}^{n} p(y_i \,|\, \eta_i, \boldsymbol{\theta}_1)$, where $\boldsymbol{\eta} = (\eta_1, \ldots, \eta_n)$. Each response has conditional mean μ_i with inverse link function $g : \mathbb{R} \to \mathbb{R}$ such that $\mu_i = g(\eta_i)$. The vector $\boldsymbol{\theta}_1 \in \mathbb{R}^{s_1}$, with s_1 assumed small, are additional parameters of the likelihood. The structured additive predictor η_i may include an intercept β_0, fixed effects β_j of the covariates z_{ji}, and random effects $f_k(\cdot)$ of the covariates u_{ki}. The parameters β_0, $\{\beta_j\}$, $\{f_k(\cdot)\}$ are each assigned Gaussian prior distributions, and can be collected into a vector $\mathbf{x} \in \mathbb{R}^N$ such that $\mathbf{x} \sim \mathcal{N}(\mathbf{0}, \mathbf{Q}(\boldsymbol{\theta}_2)^{-1})$ where $\boldsymbol{\theta}_2 \in \mathbb{R}^{s_2}$ are further hyperparameters, again with s_2 assumed small. Let $\boldsymbol{\theta} = (\boldsymbol{\theta}_1, \boldsymbol{\theta}_2) \in \mathbb{R}^m$ with $m = s_1 + s_2$ be all hyperparameters, with prior distribution $p(\boldsymbol{\theta})$.

3.3.4 Extended latent Gaussian models

Extended latent Gaussian models [ELGMs; Stringer et al. (2021)] facilitate modelling of LGMs with greater non-linearities. In particular, the structured additive predictor is redefined as $\boldsymbol{\eta} = (\eta_1, \ldots \eta_{N_n})$, where $N_n \in \mathbb{N}$ is a function of n, and it is possible that $N_n \neq n$. Each mean response μ_i now depends on some subset $\mathcal{J}_i \subseteq [N_n]$ of indices of $\boldsymbol{\eta}$, with $\cup_{i=1}^{n} \mathcal{J}_i = [N_n]$ and $1 \leq |\mathcal{J}_i| \leq N_n$, where $[N_n] = \{1, \ldots, N_n\}$. The inverse link function $g(\cdot)$ is redefined for each observation to be a possibly many-to-one mapping $g_i : \mathbb{R}^{|\mathcal{J}_i|} \to \mathbb{R}$, such that $\mu_i = g_i(\boldsymbol{\eta}_{\mathcal{J}_i})$. Put together, ELGMs are of the form

$$\begin{aligned} y_i &\sim p(y_i \,|\, \boldsymbol{\eta}_{\mathcal{J}_i}, \boldsymbol{\theta}_1), \quad i = 1, \ldots, n, \\ \mu_i &= \mathbb{E}(y_i \,|\, \boldsymbol{\eta}_{\mathcal{J}_i}) = g_i(\boldsymbol{\eta}_{\mathcal{J}_i}), \\ \eta_j &= \beta_0 + \sum_{l=1}^{p} \beta_j z_{ji} + \sum_{k=1}^{r} f_k(u_{ki}), \quad j = 1, \ldots, N_n, \end{aligned}$$

with latent field and hyperparameter prior distributions as in the LGM case.

The ELGM class is well suited to small-area estimation of HIV indicators. Indeed, this class of models is used throughout the thesis. While it can be transformed to an LGM using the Poisson-multinomial transformation (Baker 1994) the multinomial logistic regression model used in Chapter 5 is naturally written as an ELGM where each observation depends on the set of structured additive predictors corresponding to the set of multinomial observations. In Chapter 6, the Naomi small-area estimation model used to produce estimates of HIV indicators is shown to have the features of an ELGM.

3.4 Survey methods

Large national household surveys provide the highest quality population-level information about HIV indicators in SSA. Demographic and Health Surveys (DHS) are funded by the United States Agency for International Development (USAID) and run every three to five years in most countries. Population-based HIV Impact Assessment (PHIA) surveys are funded by PEPFAR and run every two to three years in high HIV burden countries.

3.4.1 Survey notation and key terms

Consider a population of individuals $i = 1, \ldots, N$ with outcomes of interest y_i. A census is a type of survey where all individuals are sampled. Supposing responses from all individuals were recorded, then any population means can be calculated directly. For example, if $G_i = G(y_i)$ then the population mean of G is

$$\bar{G} = \frac{1}{N} \sum_{i=1}^{N} G(y_i). \tag{3.15}$$

In practice, it is usually too expensive to run a census. Instead, only a subset of the individuals are sampled. Furthermore, only a subset of those sampled have their outcome recorded, due to nonresponse or otherwise. Let S_i be an indicator for whether or not individual i is sampled, and R_i be an indicator for whether

or not y_i is recorded. If $S_i = 0$ then $R_i = 0$. If $S_i = 1$ then individual i may not respond such that $R_i = 0$. The population mean may be estimated directly based on the recorded subset of the population by

$$\bar{G}_R = \frac{\sum_{i=1}^{N} R_i G(y_i)}{\sum_{i=1}^{N} R_i}, \tag{3.16}$$

where $m_R = \sum_{i=1}^{N} R_i$ is the recorded sample size.

A probability sample refers to the case when individuals are selected to be included in the survey at random. In a non-probability sample, inclusion or exclusion from the survey is deterministic. A simple random sample (SRS) is a probability sample where the sampling probability for each individual is equal, so that $P(S_i = 1) = 1/N$. The survey design is called complex when the sampling probabilities for each individual vary, such that $P(S_i = 1) = \pi_i$ with $\sum_{i=1}^{N} \pi_i = 1$ and $\pi_i > 0$.

Complex survey designs can offer both greater practicality and statistical efficiency than a SRS. However, particular care is required in analysing data collected using complex survey designs. Under a complex design, failing to take into account the unequal sampling probabilities will result in bias. That said, even for a SRS, nonresponse can cause analogous bias.

3.4.2 Survey design

The DHS (DHS 2012) employs a two-stage sampling procedure. In the first stage, ennumeration areas (EAs) from a recently conducted census are typically used as the primary sampling unit (PSU). The EAs are then stratified by region, as well as urban-rural. After appropriate sample sizes are determined, EAs sampled with probability proportional to size (PPS) measured In the second stage, the secondary sampling units (SSUs) are households. All households in the selected EAs are listed, before being sampled systematically. Finally, each selected household is visited, and all adults are interviewed.

The probability an individual is sampled is equal to the probability their household is sampled. The first-stage sampling probability of the jth cluster

in stratum h given by

$$\pi_{1hj} = n_h \times \frac{N_{hj}}{\sum_j N_{hj}}, \tag{3.17}$$

where N_{hj} is the number of households and n_h be the number of clusters selected in stratum h. The second-stage sampling probability each household within the ith cluster in stratum h is

$$\pi_{1hj} = \frac{n_{hj}}{N_{hj}}, \tag{3.18}$$

where n_{hj} is the numer of households selected in cluster j and stratum h. That is, each household in the cluster has equal selection probability. The overall selection probability of each household in cluster j of stratum h is $\pi_{hi} = \pi_{1hj} \times \pi_{2hj}$.

3.4.3 Survey analysis

Suppose a complex survey is run with sampling probabilities π_i. The most popular method for taking into account that some individuals are more likely to be included in the survey than others is to overweight the responses of those unlikely to be included, and underweight the responses of those likely to be included. This can be achieved using design weights $\delta_i = 1/\pi_i$, which can be thought of as the number of individuals in the population represented by the ith sampled individual. Let $P(R_i = 1 \,|\, S_i = 1) = \upsilon_i$ be the probability of response for sampled individual i. The problem of nonresponse can be treated in the same way using nonresponse weights $\gamma_i = 1/\upsilon_i$, which analogously can be thought of as the number of sampled individuals represented by the ith recorded individual. Multiplying the design and nonresponse weights gives survey weights $\omega_i = \delta_i \times \gamma_i$.

A weighted estimate (Hájek 1971) of the population mean using the survey weights ω_i is given by

$$\bar{G}_\omega = \frac{\sum_{i=1}^N \omega_j R_i G(y_i)}{\sum_{i=1}^N \omega_i R_i}. \tag{3.19}$$

Decomposing the additive error of this estimate provides useful intuition as to the potential benefits of survey weighting. Following Meng (2018) then under SRS

$$\bar{G}_\omega - \bar{G} = \frac{\mathbb{E}(\omega_i R_i G_i)}{\mathbb{E}(\omega_i R_i)} - \mathbb{E}(G_i) = \frac{\mathbb{C}(\omega_i R_i G_i)}{\mathbb{E}(\omega_i R_i)} \tag{3.20}$$

$$= \rho_{R_\omega, G} \times \sqrt{\frac{N - m_{R_\omega}}{m_{R_\omega}}} \times \sigma_G, \tag{3.21}$$

where $R_\omega = \omega R$. The data defect correlation (DDC) $\rho_{R_\omega, G}$ measures the correlation between the weighted recording mechanism and given function of the outcome of interest. To minimise the DDC then $G \perp\!\!\!\perp R_\omega$. The data scarcity $\sigma_{R_\omega} = \sqrt{(N - m_{R_\omega})/m_{R_\omega}}$ measures the effective proportion of the population who have been recorded. The problem difficultly σ_G measures the intrinsic difficulty of the estimation problem, and is independent of the sampling or analysis method.

For simplicity, let $G(y_i) = y_i$ and each $y_i \in \{0, 1\}$. We weight then model following Chen et al. (2014). While this approach acknowledges the survey design, it has some important limitations. We ignore clustering structure. All of this isn't great and that someone should figure this out (Gelman 2007).

4

Models for spatial structure

In this chapter, I present an investigation of spatial random effects specifications. My investigation was motivated by a fundamental question encountered during model construction. Namely, should the model be augmented to capture a conjectured feature of the data? The results are presented in Howes, Eaton, et al. (2023+). Code for the analysis in this chapter is available from `athowes/beyond-borders`.

4.1 Background

4.2 Models based on adjacency

4.2.1 The Besag model

Spatial structure can be encoded using a symmetric relation between areas. Let $i \sim j$ if the areas A_i and A_j are adjacent or neighbouring. Adjacency is often defined by a shared border, though other choices are possible (Paciorek et al. 2013). The Besag model (Besag et al. 1991) is an improper conditional auto-regressive (ICAR) model where the full conditional distribution of the ith spatial random effect is given by

$$u_i \,|\, \mathbf{u}_{-i} \sim \mathcal{N}\left(\frac{1}{n_{\delta i}} \sum_{j:j\sim i} u_j, \frac{1}{n_{\delta i}\tau_u}\right), \tag{4.1}$$

where δi is the set of neighbours of A_i with cardinality $n_{\delta i} = |\delta i|$ and $\mathbf{u}_{-i}$ is the vector of spatial random effects with the ith entry removed. The conditional mean of the random effect u_i is the average of its neighbours $\{u_j\}_{j\sim i}$ and the precision $n_{\delta i}\tau_u$ is proportional to the number of neighbours $n_{\delta i}$. By Brook's lemma (Rue and Held 2005) the set of full conditionals of the Besag model are equivalent to the Gaussian Markov random field (GMRF) given by

$$\mathbf{u} \sim \mathcal{N}(\mathbf{0}, \tau_u^{-1}\mathbf{R}^-), \tag{4.2}$$

where $\mathbf{R}^-$ is the generalised inverse of the rank-deficient structure matrix $\mathbf{R}$, so-called because it defines the structure of the precision matrix, with entries

$$R_{ij} = \begin{cases} n_{\delta i}, & i = j \\ -1, & i \sim j \\ 0, & \text{otherwise.} \end{cases} \tag{4.3}$$

The Markov property arises due to the conditional independence structure $p(u_i \,|\, \mathbf{u}_{-i}) = p(u_i \,|\, \mathbf{u}_{\delta i})$ whereby each area only depends on its neighbours. This is reflected in the sparsity of $\mathbf{R}$ such that $u_i \perp u_j \,|\, \mathbf{u}_{-ij}$ if and only if $R_{ij} = 0$. The structure matrix $\mathbf{R}$ may also be expressed as the Laplacian of the adjacency graph $\mathcal{G} = (\mathcal{V}, \mathcal{E})$ with vertices $v \in \mathcal{V}$ corresponding to each area and edges $e \in \mathcal{E}$ between vertices i and j when $i \sim j$.

Rewriting Equation (4.2), the probability density function of $\mathbf{u}$ is

$$p(\mathbf{u}) \propto \exp\left(-\frac{\tau_u}{2}\mathbf{u}^\top \mathbf{R}\mathbf{u}\right) \propto \exp\left(-\frac{\tau_u}{2}\sum_{i\sim j}(u_i - u_j)^2\right). \tag{4.4}$$

This density is a function of the pairwise differences $u_i - u_j$ and so is invariant to the addition of a constant c to each entry $p(\mathbf{u}) = p(\mathbf{u} + c\mathbf{1})$, leading to an improper uniform distribution on the average of the u_i. If $\mathcal{G}$ is connected, in that by traversing the edges, any vertex can be reached from any other vertex, then there is only one impropriety in the model and $\text{rank}(\mathbf{R}) = n - 1$, while if $\mathcal{G}$ is disconnected, and composed of $n_c \geq 2$ connected components with index sets $I_1, \ldots, I_{n_c}$, then the corresponding structure matrix $\mathbf{R}$ has rank $n - n_c$ and the density is invariant to the addition of a constant to each of the connected components $p(\mathbf{u}_I) = p(\mathbf{u}_I + c\mathbf{1})$ where $I = I_1, \ldots, I_{n_c}$.

4.2.2 Best practises for the Besag model

Freni-Sterrantino et al. (2018) recommended three best practices:

1. The structure matrix $\mathbf{R}$ should be rescaled to have generalised variance, defined by the geometric mean of the diagonal elements of its generalised inverse
$$\sigma^2_{\mathrm{GV}}(\mathbf{R}) = \prod_{i=1}^{n} (\mathbf{R}^-_{ii})^{1/n} = \exp\left(\frac{1}{n}\sum_{i=1}^{n} \log(\mathbf{R}^-_{ii})\right), \tag{4.5}$$
equal to one, by replacing $\mathbf{R}$ with $\mathbf{R}^\star = \mathbf{R}/\sigma^2_{\mathrm{GV}}(\mathbf{R})$. As the diagonal elements R^-_{ii} correspond to marginal variances, the generalised variance gives a measure of the average marginal variance. However, this measure, introduced by Sørbye and Rue (2014), ignores off-diagonal entries and more broadly any measure of typical variance could be used. Scaling mitigates the influence of the adjacency graph on the variance of $\mathbf{u}$. Allowing the variance to be controlled by τ_u alone is important as it allows for consistent, interpretable prior selection.

When the adjacency graph is disconnected it is not appropriate to scale the structure matrix $\mathbf{R}$ uniformly since for a given precision τ_u, local smoothing operates on each connected component independently. As such, each connected component should be scaled independently to have generalised variance one giving $\mathbf{R}^\star_I = \mathbf{R}_I/\sigma^2_{\mathrm{GV}}(\mathbf{R}_I)$ where $\mathbf{R}_I$ is the sub-matrix of the structure matrix corresponding to index set I.

2. When one of the connected components is a single area, known either as a singleton or an island, the probability density $\exp\left(-\frac{\tau_u}{2}\sum_{i\sim j}(u_i - u_j)^2\right)$ has no dependence on u_i. This is equivalent to using an improper prior $p(u_i) \propto 1$ and can be avoided by setting each singleton to have independent Gaussian noise $p(u_i) \sim \mathcal{N}(0, 1)$.
3. To avoid confounding of the spatial random effects with the intercept, it is recommended to place a sum-to-zero constraint on each non-singleton connected component. In other words, for each $|I| > 1$ that $\sum_{i\in I} u_i = 0$.

4.2.3 The reparameterised Besag-York-Mollié model

Often, as well as spatial structure, there exists IID over-dispersion in the residuals and it is inappropriate to use purely spatially structured random effects in the model. The Besag-York-Mollié (BYM) model of Besag et al. (1991) accounts for this in a natural way by decomposing the spatial random effect $\mathbf{u} = \mathbf{v} + \mathbf{w}$ into a sum of an unstructured IID component $\mathbf{v}$ and a spatially structured Besag component $\mathbf{w}$, each of which with their own respective precision parameters τ_v and τ_w. The resulting distribution is

$$\mathbf{u} \sim \mathcal{N}(0, \tau_v^{-1}\mathbf{I} + \tau_w^{-1}\mathbf{R}^-). \tag{4.6}$$

Including both $\mathbf{v}$ and $\mathbf{w}$ is intended to enable the model to learn the relative extent of the unstructured and structured components via τ_v and τ_w. However, in this specification scaling of the Besag precision matrix $\mathbf{Q}$ is not taken into account despite this issue being particularly pertinent when dealing with multiple sources of noise. In particular, placing a joint prior $(\tau_v, \tau_w) \sim p(\tau_v, \tau_w)$ which doesn't privilege either component is more easily accomplished if $\mathbf{Q}$ and $\mathbf{I}$ have the same scale. Additionally, supposing we have a prior belief that the over-dispersion is primarily IID and $\mathbf{v}$ accounts for the majority of the dispersion, then it is not immediately obvious how to represent this belief using $p(\tau_v, \tau_w)$, without inadvertently altering the prior about the overall variation. This highlights identifiability issues of the parameters (τ_v, τ_w) resulting from them not being orthogonal. Building on the models of Leroux et al. (2000) and Dean et al. (2001) which tackle this identifiability problem, but do not scale the spatially structured noise, Simpson et al. (2017) propose a reparameterisation $(\tau_v, \tau_w) \mapsto (\tau_u, \phi)$ of the BYM model known as the BYM2 model and given by

$$\mathbf{u} = \frac{1}{\tau_u}\left(\sqrt{1-\phi}\,\mathbf{v} + \sqrt{\phi}\,\mathbf{w}^\star\right), \tag{4.7}$$

where τ_u is the marginal precision of $\mathbf{u}$, $\phi \in [0, 1]$ gives the proportion of the marginal variance explained by each component, and $\mathbf{w}^\star$ is a scaled version of $\mathbf{w}$ with precision matrix given by the scaled structure matrix $\mathbf{R}^\star$. When $\phi = 0$ the

random effects are IID, and when $\phi = 1$ the random effects follow the Besag model. To borrow an analogy (Rue 2020) the parameterisation (τ_v, τ_w) is like having one hot water and one cold water tap, whereas the parameterisation (τ_u, ϕ) is like a mixer tap where the amount of water and its temperature can be adjusted separately.

4.2.4 Concerns about the Besag model's representation of space

The Besag model was originally proposed for use in image analysis, where areas correspond to pixels arranged in a regular lattice structure. Since then, it has seen wider use, including in situations, like small-area estimation of HIV, where the spatial structure is less regular. As such, I have a number of concerns about the model's applicability to this broader setting. This discussion is closely linked to the modifiable areal unit problem (Openshaw and Taylor 1979), whereby statistical conclusions change as a result of seemingly arbitrary changes in data aggregation, as well as the challenge of ecological inference and the ecological fallacy (Wakefield and Lyons 2010).

Adjacency compression

Summarising a geometry by an adjacency graph represents a loss of information. Many geometries share the same adjacency graph, and are as such isomorphic identical under the Besag model (Figure). This is not in itself a problem, but does prompt consideration as to whether the class of geometries with the same adjacency graph is sufficiently similar to merit identical models.[1] Intuitively, the more regular the spatial structure, the less information is lost in compression to an adjacency graph. In image analysis, very little spatial information is lost in compression of a lattice structure to an adjacency graph. On the other hand, the regions of a country, determined by political and geographic forces, tend to display

[1]The regularity of realistic geometries may help to constrain each class to be more similar than it strictly has to be. In other words, although pathological geometries can be constructed, they are implausible in statistical practise and so not of great concern to us here.

greater irregularity. The appropriateness of adjacency compression therefore varies by the type of geometry common to the application setting.

Mean structure

In the Besag model is all adjacent areas count equally. This assumption is unsatisfying: for most geometries, we expect different amounts of correlation between neighbours. Figure illustrates a number of heuristic features for neighbour importance, including length of shared border, and the proximity of centers of mass.

Variance structure

In Equation (4.1) the precision of u_i is proportional to its number of neighbours $n_{\delta i}$. It follow that as $n_{\delta i} \to \infty$ then $\text{Var}(u_i) \to 0$. This is illustrated by Figure where the area on the right is repeatedly divided such that its number of neighbours increases. This property is a consequence of averaging the conditional mean over a greater number of areas, which, in certain situations, can correspond to a greater amount of information. However, if the amount of information in the shaded area remains fixed, it is inappropriate that $\text{Var}(u_1)$ should tend to zero as a result of drawing additional, arbitrary, boundaries. In the image analysis setting this modelling assumption is reasonable: each pixel represents a fixed amount of information and a higher pixel density represents a greater amount of information. On the other hand, in public health and epidemiology, drawing boundaries to create additional areas is not expected to correspond to a greater amount of information.

Suppose we fit a Besag model upon identical data using each of the two geometries in Figure. If the spatial variation is relatively smooth, dividing the shaded areas into two will result in a lower estimated variance σ_u^2 in geometry (ii) as compared with geometry (i) because there will appear to be less variation between neighbouring areas. This problem does not only apply locally: since the effect of σ_u^2 applies everywhere, the smoothing will change even in unaltered parts of the study region.

4.3 Models using kernels

4.4 Simulation study

4.5 HIV prevalence study

4.6 Discussion

5

A model for risk group proportions

In this chapter I describe an application of Bayesian spatio-temporal statistics to small-area estimation of HIV risk group proportions. This work was conducted in collaboration with colleagues from the MRC Centre for Global Infectious Disease Analysis and UNAIDS. I developed the statistical model, building upon an earlier version of the analysis conducted by Kathryn Risher. The results are presented in Howes, Risher, et al. (2023). Kathryn has also created a spreadsheet tool (`https://hivtools.unaids.org/pse/`) using the estimates. This tool is being used by countries, and continues to be developed. Code for the analysis in this chapter is available from `athowes/multi-agyw`.

5.1 Background

In SSA, adolescent girls and young women (AGYW) aged 15-29 are a demographic group at increased risk of HIV infection. Though AGYW are only 28% of the population, they comprise 44% of new infections (UNAIDS 2021a). HIV incidence for AGYW is 2.4 times higher than for similarly aged males. The social and biological reasons for this disparity include structural vulnerabilites and power imbalances, age patterns of sexual mixing, younger age at first sex, and increased susceptibility to HIV infection. On this basis, AGYW have been identified as a

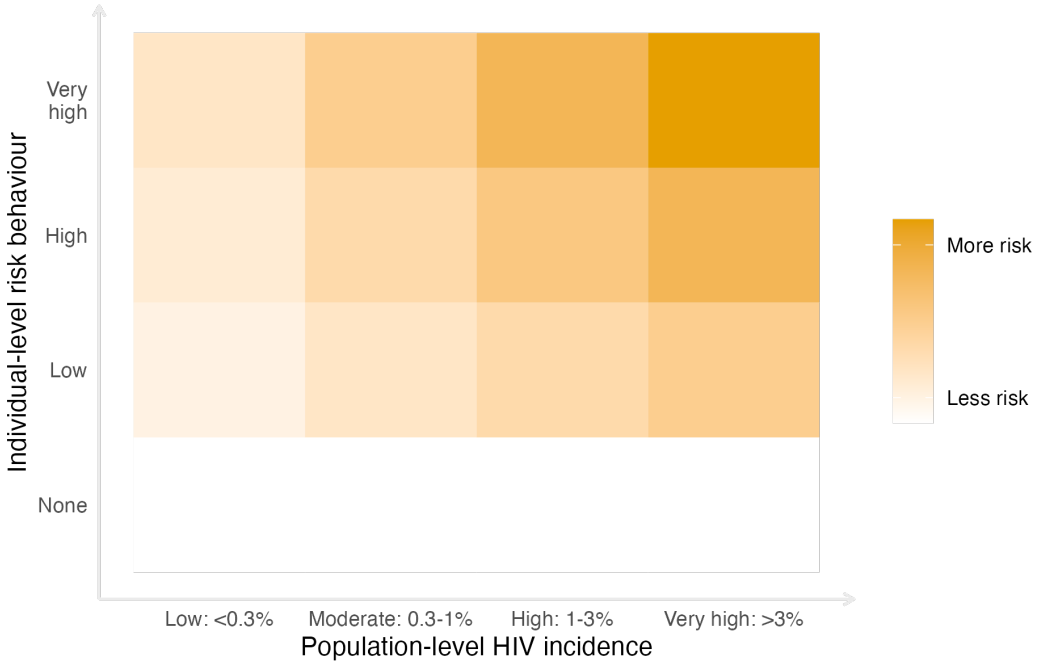

Figure 5.1: Risk of acquiring HIV depends on both individual-level risk behaviour and population-level HIV incidence. I assume that with no individual-level risk behaviour, there is no risk of acquiring HIV, independent of the population-level HIV incidence. This figure is adapted from a presentation I gave for High Impact Medicine, and is intended to be illustrative, rather than interpreted quantitatively.

priority population for HIV prevention services, and significant investments have been made in prevention programming (Saul et al. 2018; The Global Fund 2018).

The Global AIDS Strategy 2021-2026 (UNAIDS 2021b) was adopted by the United Nations (UN) General Assembly in June 2021. It proposed stratifying HIV prevention packages to AGYW based on 1) local population-level HIV incidence and 2) individual-level sexual risk behaviour. Risk of acquiring HIV depends on both factors. As such, prioritisation of prevention services is more efficient if both are taken into account (Figure 5.1). The strategy encourages programmes to define targets for the proportion of AGYW to be reached with a range of interventions. Implementation of the strategy by national HIV programmes and stakeholders requires data on the population size and HIV incidence in each risk group by location.

5.2 Data

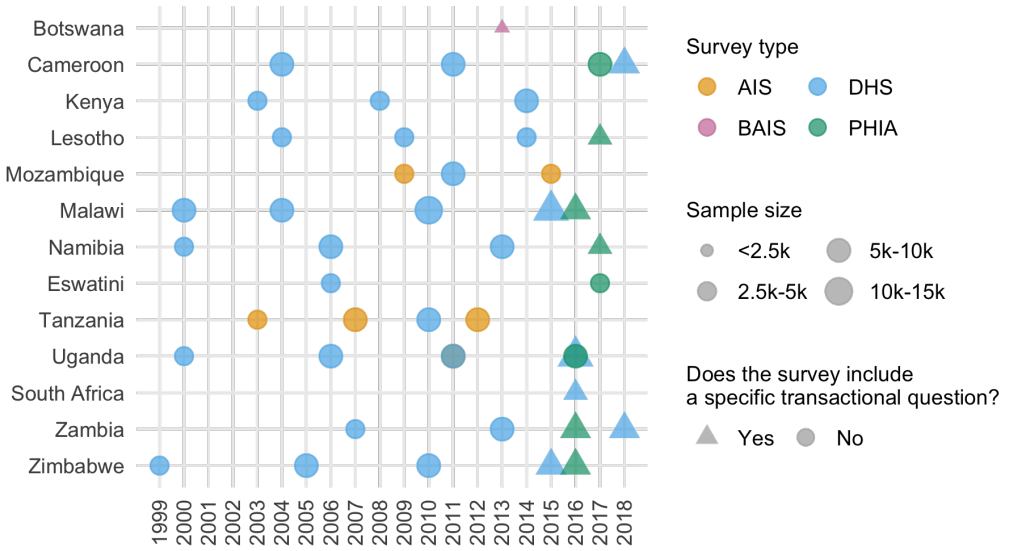

Figure 5.2: Surveys conducted 1999-2018 that were used in the analysis by year, survey type, sample size, and whether the survey included a specific question about transactional sex.

Table 5.1: HIV risk groups and HIV incidence rate ratios relative to AGYW with one cohabiting sexual partner. The incidence rate ratio for women with non-regular or multiple sexual partner(s) was derived from analysis of longitudinal data (Slaymaker et al. 2020). Among FSW, the incidence rate ratio (25.0, 13.0, 9.0, 6.0, 3.0) depended on the level of HIV incidence among the general population (<0.1%, 0.1-0.3%, 0.3-1.0%, 1.0-3.0%, >3.0%), such that higher local HIV incidence in the general population corresponds to a lower incidence rate ratio for FSW. These estimates were derived by UNAIDS based on patterns of relative HIV prevalence among FSW compared to general population prevalence.

Risk group	Description	Incidence rate ratio
None	Not sexually active	0.0
Low	One cohabiting sexual partner	1.0 (baseline)
High	Non-regular or multiple partner(s)	1.72
Very High	Reporting transactional sex (later adjusted to correspond to FSW)	3.0-25.0 (varies depending on local HIV incidence)

I used household survey data from 13 countries identified by the GFATM as priority countries for implementation of AGYW HIV prevention: Botswana, Cameroon, Kenya, Lesotho, Malawi, Mozambique, Namibia, South Africa, Eswatini,

Tanzania, Uganda, Zambia and Zimbabwe. Surveys conducted in these countries between 1999 and 2018 in which women were interviewed about their sexual behaviour and sufficient geographic information was available to locate survey clusters to health districts were included. There were 46 suitable surveys (Figure 5.2), with a total sample size of 274,970 women aged 15-29 years. Of the respondents, 103,063 were aged 15-19 years, 92,173 were aged 20-24 years, and 79,734 were aged 25-29 years. The median number of surveys per country was four, ranging from one in Botswana and South Africa to six in Uganda. Included surveys comprise Demographic and Health Surveys (DHS), AIDS Indicator Surveys (AIS), Population-based HIV Impact Assessment (PHIA) surveys, and the Botswana AIDS Impact Survey 2013 (BAIS).

For each survey, respondents were classified into one of four behavioural risk groups according to reported sexual risk behaviour in the past 12 months. These risk groups were (Table 5.1):

1. Not sexually active
2. One cohabiting sexual partner
3. Non-regular or multiple sexual partner(s), and
4. Reporting transactional sex.

In the case of inconsistent responses, women were categorised according to the highest risk group they fell into, ensuring that the categories were mutually exclusive. Exact survey questions varied slightly across survey types and between survey phases. Questions captured information about whether the respondent had been sexually active in the past twelve months, and if so how with many partners. For their three most recent partners, respondents were also asked about the type of partnership. Possible partnership types included spouse, cohabiting partner, partner not cohabiting with respondent, friend, sex worker, sex work client, and other. Full survey questions used are provided in Appendix B.4.

Some surveys included a specific question asking if the respondent had received or given money or gifts for sex in the past twelve months. In these surveys, 2.64%

of women reported transactional sex. In surveys without such a question, women almost never (0.01%) answered that one of their three most recent partners was a sex work client. As a result of this incomparability, it was not not appropriate to include surveys without a specific transactional sex question when estimating the proportion of the population who engaged in transactional sex. Of the total 46 surveys included in the analysis, 12 had a specific transactional sex question, with a total sample size of 62,853 (28,753 aged 15-19 years, 26,324 aged 20-24 years, and 7,776 aged 25-29 years. There were 6 DHS surveys which excluded women 25-29 from the transactional sex survey question.

I used estimates of population, PLHIV and new HIV infections stratified by district and age group from HIV estimates published by UNAIDS that were developed using the Naomi model (Eaton et al. 2021). The administrative area hierarchy and geographic boundaries I used correspond to those used for health service planning by countries. Exceptions are Cameroon and Kenya, where I conducted analysis one level higher at the department and county levels, respectively. I used the most recent 2022 estimates for all countries, apart from Mozambique where, due to data accuracy concerns, I used the 2021 estimates (in which the Cabo Delgado province is excluded due to disruption by conflict).

5.3 Model for risk group proportions

I took a two-stage modelling approach to estimate the four risk group proportions. Let the four risk groups be indexed by $k \in \{1, 2, 3, 4\}$, and denote being in either the third or fourth risk group as $k = 3^+$. First, using all the surveys, I used a spatio-temporal multinomial logistic regression model (Section 5.3.1) to estimate the proportion of AGYW in the risk groups $k \in \{1, 2, 3^+\}$. Then, using only those surveys with a specific transactional sex question, I fit a spatial logistic regression model (Section 5.3.2) to estimate the proportion of those in the $k = 3^+$ risk group that were in the $k = 3$ and $k = 4$ risk groups respectively.

5.3.1 Spatio-temporal multinomial logistic regression

Let $i \in \{1, \ldots, n\}$ denote districts partitioning the 13 studied AGYW priority countries $c[i] \in \{1, \ldots, 13\}$. Consider the years 1999-2018 denoted as $t \in \{1, \ldots, T\}$, and age groups $a \in \{15\text{-}19, 20\text{-}24, 25\text{-}29\}$. Let $p_{itak} > 0$ with $\sum_{k=1}^{3^+} p_{itak} = 1$, be the probabilities of membership of risk group k.

Multinomial logistic regression

A standard multinomial logistic regression model (e.g. Gelman, Carlin, et al. 2013) is specified by

$$\mathbf{y}_{ita} = (y_{ita1}, \ldots, y_{ita3^+})^\top \sim \text{Multinomial}(m_{ita};\, p_{ita1}, \ldots, p_{ita3^+}), \tag{5.1}$$

$$\log\left(\frac{p_{itak}}{p_{ita1}}\right) = \eta_{itak}, \quad k = 2, 3^+, \tag{5.2}$$

where the number in risk group k is y_{itak}, the fixed sample size is $m_{ita} = \sum_{k=1}^{3^+} y_{itak}$, and $k = 1$ is chosen as the baseline category. This model is not an LGM. I used the multinomial-Poisson transformation to perform inference using the INLA (Rue, Martino, et al. 2009) algorithm via the `R-INLA` package (Martins et al. 2013). INLA has comparable accuracy and is substantially more computationally tractable than MCMC for GLMMs defined over 940 districts, 20 years, 3 age groups, and 4 risk groups.

The multinomial-Poisson transformation

The multinomial-Poisson transformation reframes a given multinomial logistic regression model as an equivalent Poisson log-linear model. The equivalent model is of the form

$$y_{itak} \sim \text{Poisson}(\kappa_{itak}), \tag{5.3}$$

$$\log(\kappa_{itak}) = \eta_{itak}. \tag{5.4}$$

The basis of the transformation is that conditional on their sum Poisson counts are jointly multinomially distributed (McCullagh and Nelder 1989) as follows

$$\mathbf{y}_{ita} \,|\, m_{ita} \sim \text{Multinomial}\left(m_{ita};\, \frac{\kappa_{ita1}}{\kappa_{ita}}, \ldots, \frac{\kappa_{ita3^+}}{\kappa_{ita}}\right), \tag{5.5}$$

where $\kappa_{ita} = \sum_{k=1}^{3+} \kappa_{itak}$. Category probabilities may then be obtained using the softmax function

$$p_{itak} = \frac{\exp(\eta_{itak})}{\sum_{k=1}^{3+} \exp(\eta_{itak})} = \frac{\kappa_{itak}}{\sum_{k=1}^{3+} \kappa_{itak}} = \frac{\kappa_{itak}}{\kappa_{ita}}. \tag{5.6}$$

Under the equivalent model, the sample sizes m_{ita} are treated as random rather than fixed such that

$$m_{ita} = \sum_k y_{itak} \sim \text{Poisson}\left(\sum_k \kappa_{itak}\right) = \text{Poisson}\left(\kappa_{ita}\right). \tag{5.7}$$

The joint distribution of $p(\mathbf{y}_{ita}, m_{ita}) = p(\mathbf{y}_{ita} \,|\, m_{ita})p(m_{ita})$ is then

$$p(\mathbf{y}_{ita}, m_{ita}) = \exp(-\kappa_{ita})\frac{(\kappa_{ita})^{m_{ita}}}{m_{ita}!} \times \frac{m_{ita}!}{\prod_k y_{itak}!} \prod_k \left(\frac{\kappa_{itak}}{\kappa_{ita}}\right)^{y_{itak}} \tag{5.8}$$

$$= \prod_k \left(\frac{\exp(-\kappa_{itak})\,(\kappa_{itak})^{y_{itak}}}{y_{itak}!}\right) \tag{5.9}$$

$$= \prod_k \text{Poisson}\left(y_{itak} \,|\, \kappa_{itak}\right), \tag{5.10}$$

corresponding to the product of independent Poisson likelihoods as in Equation (5.3).

The random sample size model is equivalent to a multinomial logistic regression model only when the normalisation constants m_{ita} are recovered exactly. To ensure that this is the case, observation-specific random effects θ_{ita} can be included in the equation for the linear predictor. Multiplying each of $\{\kappa_{itak}\}_{k=1}^{3+}$ by $\exp(\theta_{ita})$ has no effect on the category probabilities, but does provide the necessary flexibility for κ_{ita} to recover m_{ita} exactly. Although in theory an improper prior distribution $\theta_{ita} \propto 1$ should be used, I found that in practise, by keeping η_{ita} otherwise small using appropriate constraints, so that arbitrarily large values of θ_{ita} are not required, it is sufficient (and practically preferable for inference) to instead use a vague prior distribution.

Model specifications

Table 5.2: Four multinomial regression models were considered. Observation random effects θ_{ita}, included in all models, are omitted from this table.

Category β_k	Country ζ_{ck}	Age α_{ack}	Spatial ϕ_{ik}	Temporal γ_{tk}
M1 IID	IID	IID	IID	IID
M2 IID	IID	IID	Besag	IID
M3 IID	IID	IID	IID	AR1
M4 IID	IID	IID	Besag	AR1

I considered four models (Table 5.2) for η_{ita} in the equivalent Poisson log-linear model (Equation (5.4)) of the form

$$\eta_{ita} = \theta_{ita} + \beta_k + \zeta_{c[i]k} + \alpha_{ac[i]k} + u_{ik} + \gamma_{tk}. \tag{5.11}$$

Observation random effects $\theta_{ita} \sim \mathcal{N}(0, 1000^2)$ with a vague prior distribution were included in all models I considered to ensure the transformation is valid. To capture country-specific proportion estimates for each category, I included category random effects $\beta_k \sim \mathcal{N}(0, \tau_\beta^{-1})$ and country-category random effects $\zeta_{ck} \sim \mathcal{N}(0, \tau_\zeta^{-1})$. Heterogeneity in risk group proportions by age was allowed by including age-country-category random effects $\alpha_{ack} \sim \mathcal{N}(0, \tau_\alpha^{-1})$.

Spatial random effects For the space-category u_{ik} random effects I considered two specifications:

1. Independent and identically distributed (IID) $u_{ik} \sim \mathcal{N}(0, \tau_u^{-1})$,
2. Besag (Besag et al. 1991) grouped by category

$$\mathbf{u} = (u_{11}, \ldots, u_{n1}, \ldots, u_{13^+}, \ldots u_{n3^+})^\top \sim \mathcal{N}(\mathbf{0}, (\tau_u \mathbf{R}_u^\star)^-).$$

   The scaled structure matrix $\mathbf{R}_u^\star = \mathbf{R}_b^\star \otimes \mathbf{I}$ is given by the Kronecker product of the scaled Besag structure matrix $\mathbf{R}_b^\star$ and the identity matrix $\mathbf{I}$, and $-$ denotes the generalised matrix inverse. I followed best practices for the Besag model as described in Chapter 4.

In preliminary testing, I excluded spatial random effects from the model, but found that this negatively effected performance. I also tested using the BYM2 model (Simpson et al. 2017) in place of the Besag, but found that the proportion parameter posteriors tended to be highly peaked at the value one. For simplicity and to avoid numerical issues, by using Besag random effects I effectively decided to fix this proportion to one.

Temporal random effects For the year-category γ_{tk} random effects I considered two specifications:

1. IID $\gamma_{tk} \sim \mathcal{N}(0, \tau_\gamma^{-1})$,
2. First order autoregressive (AR1) grouped by category

$$\boldsymbol{\gamma} = (\gamma_{11}, \ldots, \gamma_{13^+}, \ldots, \gamma_{T1}, \ldots, \gamma_{T3^+})^\top \sim \mathcal{N}(\mathbf{0}, (\tau_\gamma \mathbf{R}_\gamma^\star)^-).$$

The scaled structure matrix $\mathbf{R}_\gamma^\star = \mathbf{R}_r^\star \otimes \mathbf{I}$ is given by the Kronecker product of a scaled AR1 structure matrix $\mathbf{R}_r^\star$ and the identity matrix $\mathbf{I}$. The AR1 structure matrix $\mathbf{R}_r$ is obtained by precision matrix of the random effects $\mathbf{r} = (r_1, \ldots, r_T)^\top$ specified by

$$r_1 \sim \left(0, \frac{1}{1-\rho^2}\right), \tag{5.12}$$

$$r_t = \rho r_{t-1} + \epsilon_t, \quad t = 2, \ldots, T, \tag{5.13}$$

where $\epsilon_t \sim \mathcal{N}(0, 1)$ and $|\rho| < 1$.

Prior distributions All random effect precision parameters $\tau \in \{\tau_\beta, \tau_\zeta, \tau_\alpha, \tau_u, \tau_\gamma\}$ were given independent penalised complexity (PC) prior distributions (Simpson et al. 2017) with base model $\sigma = 0$ given by

$$p(\tau) = 0.5\nu\tau^{-3/2} \exp\left(-\nu\tau^{-1/2}\right) \tag{5.14}$$

where $\nu = -\ln(0.01)/2.5$ such that $\mathbb{P}(\sigma > 2.5) = 0.01$. For the lag-one correlation parameter ρ, I used the PC prior distribution, as derived by Sørbye and Rue (2017), with base model $\rho = 1$ and condition $\mathbb{P}(\rho > 0 = 0.75)$. I chose the base

model $\rho = 1$ corresponding to no change in behaviour over time, rather than the alternative $\rho = 0$ corresponding to no correlation in behaviour over time, as I judged the former to be more plausible a priori.

Constraints

To faciliate interpretability of the posterior inferences, I applied sum-to-zero constraints such that none of the category interaction random effects altered overall category probabilities. For the space-year-category random effects, I applied analogous sum-to-zero constraints to maintain roles of the space-category and year-category random effects (Table 5.3).

Table 5.3: Applying sum-to-zero constraints to interaction effects ensures that the main effect is not interfered with.

Random effects	Constraints
Category	$\sum_k \beta_k = 0$
Country	$\sum_c \zeta_{ck} = 0, \forall\, k$
Age-country	$\sum_a \alpha_{ack} = 0, \forall\, c, k$
Spatial	$\sum_i u_{ik} = 0, \forall\, k$
Temporal	$\sum_t \gamma_{tk} = 0, \forall\, k$

Survey weighted likelihood

I accounted survey design using a weighted pseudo-likelihood where the observed counts y are replaced by effective counts $y^\star$, as described in Section 3.4. These counts may not be integers, and as such the Poisson likelihood given in Equation (5.3) is not appropriate. Instead, I used a generalised Poisson pseudo-likelihood $y^\star \sim \text{xPoisson}(\kappa)$ given by

$$p(y^\star) = \frac{\kappa^{y^\star}}{\lfloor y^\star! \rfloor} \exp\left(-\kappa\right), \tag{5.15}$$

to extend the Poisson distribution to non-integer weighted counts. This working likelihood is implemented by `family = "xPoisson"` in `R-INLA`.

Model selection

I selected the model including Besag spatial random effects and IID temporal random effects based on the conditional predictive ordinate (CPO) criterion (Pettit 1990). For reference, I also computed the deviance information criterion (DIC) (Spiegelhalter, Best, et al. 2002) and widely applicable information criterion (WAIC) (Watanabe 2013).

5.3.2 Spatial logistic regression

To estimate the proportion of those in the $k = 3^+$ risk group that were in the $k = 3$ and $k = 4$ risk groups respectively, I fit logistic regression models of the form

$$y_{ia4} \sim \text{Binomial}\left(y_{ia3} + y_{ia4}, q_{ia}\right), \tag{5.16}$$

$$q_{ia} = \text{logit}^{-1}\left(\eta_{ia}\right), \tag{5.17}$$

where

$$q_{ia} = \frac{p_{ia4}}{p_{ia3} + p_{ia4}} = \frac{p_{ia4}}{p_{ia3^+}}. \tag{5.18}$$

This two-step approach allowed all surveys to be included in the multinomial regression model, but only those surveys with a specific transactional sex question to be included in the logistic regression model. As all such surveys occurred in the years 2013-2018 (Figure 5.2), I assumed q_{ia} to be constant with respect to time.

Model specifications

Table 5.4: Six logistic regression models were considered. The covariate `cfswever` denotes the proportion of men who have ever paid for sex and `cfswrecent` denotes the proportion of men who have paid for sex in the past 12 months.

	Intercept β_0	Country ζ_c	Age α_{ac}	Spatial u_i	Covariates
L1	Constant	IID	IID	IID	None
L2	Constant	IID	IID	Besag	None
L3	Constant	IID	IID	IID	`cfswever`
L4	Constant	IID	IID	Besag	`cfswever`
L5	Constant	IID	IID	IID	`cfswrecent`
L6	Constant	IID	IID	Besag	`cfswrecent`

I considered six logistic regression models (Table 5.4). Each included a constant intercept $\beta_0 \sim \mathcal{N}(-2, 1^2)$, country random effects $\zeta_c \sim \mathcal{N}(0, \tau_\zeta^{-1})$, and age-country random effects $\alpha_{ac} \sim \mathcal{N}(0, \tau_\alpha^{-1})$. The Gaussian prior distribution on β_0 placed 95% prior probability on the range 2-50% for the percentage of those with non-regular or multiple partners who report transactional sex. I considered two specifications (IID, Besag) for the spatial random effects u_i. To aid estimation with sparse data, I also considered national-level covariates for the proportion of men who have paid for sex ever `cfswever` or in the last twelve months `cfswrecent` (Hodgins et al. 2022). For both random effect precision parameters $\tau \in \{\tau_\alpha, \tau_\zeta\}$ I used the PC prior distribution with base model $\sigma = 0$ and $\mathbb{P}(\sigma > 2.5 = 0.01)$. For both regression parameters $\beta \in \{\beta_{\texttt{cfswever}}, \beta_{\texttt{cfswrecent}}\}$ I used the prior distribution $\beta \sim \mathcal{N}(0, 2.5^2)$.

Survey weighted likelihood

As with the multinomial regression model, I used survey weighted counts $y^\star$ and sample sizes $m^\star$. I used a generalised binomial pseudo-likelihood $y^\star \sim \text{xBinomial}(m^\star, q)$ given by

$$p(y^\star \mid m^\star, q) = \binom{\lfloor m^\star \rfloor}{\lfloor y^\star \rfloor} q^{y^\star} (1-q)^{m^\star - y^\star} \tag{5.19}$$

to extend the binomial distribution to non-integer weighted counts and sample sizes. This working likelihood is implemented by `family = "xBinomial"` in `R-INLA`.

Model selection

I selected the model including Besag spatial effects and `cfswrecent` covariates according to the CPO criterion. All results, including DIC and WAIC, are presented in Table. Inclusion of Besag spatial random effects, rather than IID, consistently improved performance. Benefits from inclusion of covariates were more marginal. As some countries had no suitable surveys, inclusion of covariate information is favourable such that the estimates in these countries are based on some country-specific data.

5.3.3 Female sex worker population size adjustment

Having had sex "in return for gifts, cash or anything else in the past 12 months" is not considered sufficient to constitute sex work. As such, I adjusted the estimates obtained based on the transactional sex survey question to match alternative FSW population size estimates.

Stevens, Sabin, Arias Garcia, et al. (2022) used a Bayesian meta-analysis of key population specific data sources to estimate adult (15-49) FSW population size by country. I disaggregated these estimates by age as follows. First, I calculated the total sexually debuted population in each age group, by country. To describe the distribution of age at first sex, I used skew logistic distributions (Nguyen and Eaton 2022) with cumulative distribution function given by

$$F(x) = (1 + \exp(\kappa_c(\mu_c - x)))^{-\gamma_c}, \tag{5.20}$$

where $\kappa_c, \mu_c, \gamma_c > 0$ are country-specific shape, shape and skewness parameters respectively. Next, I used the assumed Gamma($\alpha = 10.4, \beta = 0.36$) FSW age distribution in South Africa from the Thembisa model (Johnson and Dorrington 2020) to calculate the implied ratio between the number of FSW and the sexually debuted population in each age group. I assumed the South African ratios were applicable to every country, allowing calculation of the number of FSW by age group in all 13 countries. The resutling age trends obtained (Figure 5.3) reflect country-level variation in demographics and age-at-first-sex.

5.3.4 Results

Coverage assessment

To assess the calibration of the fitted model, I calculated the quantile q of each observation within the posterior predictive distribution. For calibrated models, these quantiles, known as probability integral transform (PIT) values (Dawid 1984; Bosse et al. 2022), should follow a uniform distribution $q \sim \mathcal{U}[0, 1]$. To generate samples from the posterior predictive distribution, I applied the multinomial likelihood to samples from the latent field, setting the sample size to be the floor of the Kish

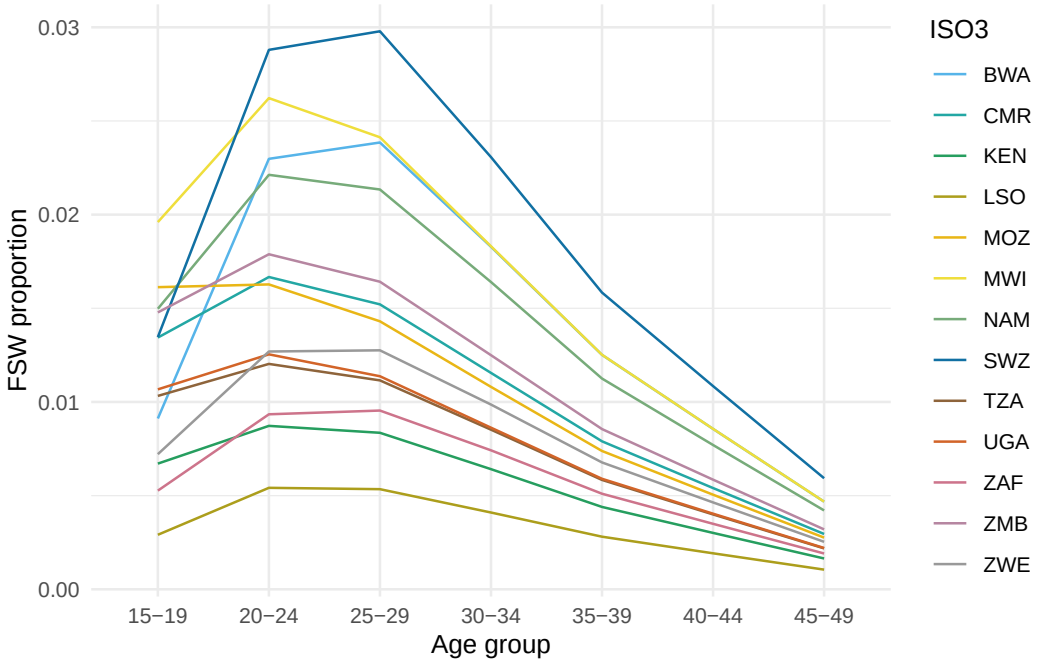

Figure 5.3: The disaggregation procedure I used produces an age distribution for FSW peaking in the 20-24 and 25-29 age groups, and declining for older age groups.

effective sample size. Using the PIT values, it is possible to calculate the empirical coverage of all $(1-\alpha)100\%$ equal-tailed posterior predictive credible intervals. These empirical coverages can be compared to the nominal coverage $(1-\alpha)$ for each value of $\alpha \in [0, 1]$ to give empirical cumulative distribution function (ECDF) difference values. This approach has the advantage of considering all possible confidence values at once. To test for uniformity, I used the binomial distribution based simultaneous confidence bands for ECDF difference values developed by Säilynoja et al. (2021). I found the only significant deviation from uniformity occurred in the right-hand tail of the one cohabiting partner risk group. That is to say, the proportion of the PIT values which were greater than 0.95 was significantly more than would be expected under a calibrated model.

Estimates

Figure 5.7 and Figure 5.6 show posterior mean estimates for the proportion in each risk group for the final model in 2018, the most recent year included

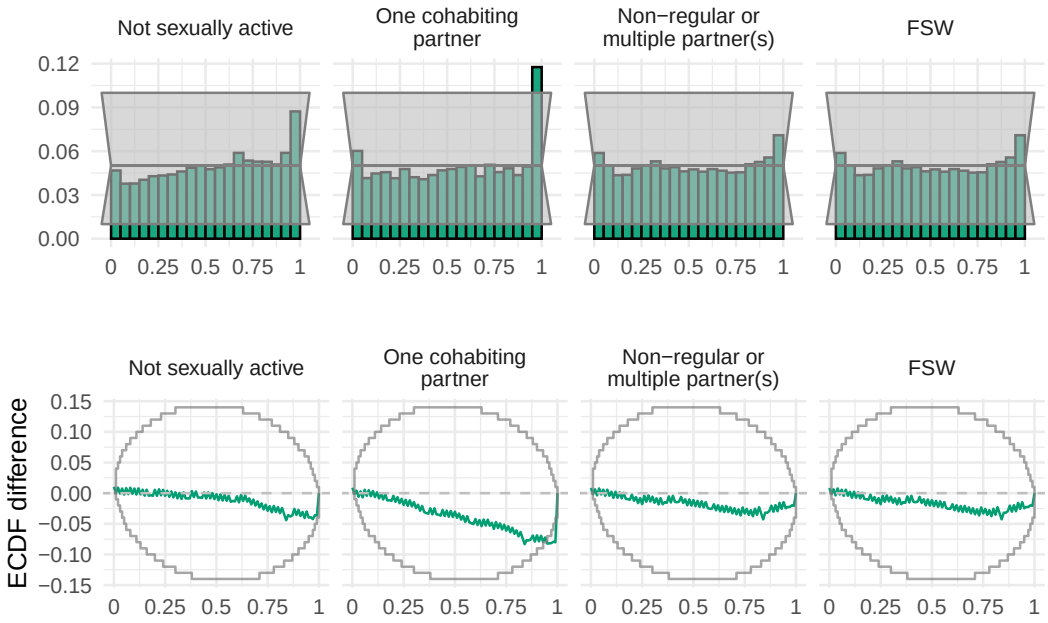

Figure 5.4: Probability integral transform (PIT) histograms (top row) and empirical cumulative distribution function (ECDF) difference plots (bottom row) for the final selected model.

in our analysis. In subsequent results, all estimates refer to this year, unless otherwise indicated.

Variance decomposition

5.4 Prevalence and incidence by risk group

5.4.1 Disaggregation of Naomi estimates

I calculated HIV incidence λ_{iak} and number of new HIV infections I_{iak} stratified according to district, age group and risk group by linear disaggregation

$$I_{ia} = \sum_k I_{iak} = \sum_k \lambda_{iak} N_{iak} \tag{5.21}$$

$$= 0 + \lambda_{ia2} N_{ia2} + \lambda_{ia3} N_{ia3} + \lambda_{ia4} N_{ia4} \tag{5.22}$$

$$= \lambda_{ia2} \left(N_{ia2} + \mathrm{RR}_3 N_{ia3} + \mathrm{RR}_4(\lambda_{ia}) N_{ia4} \right). \tag{5.23}$$

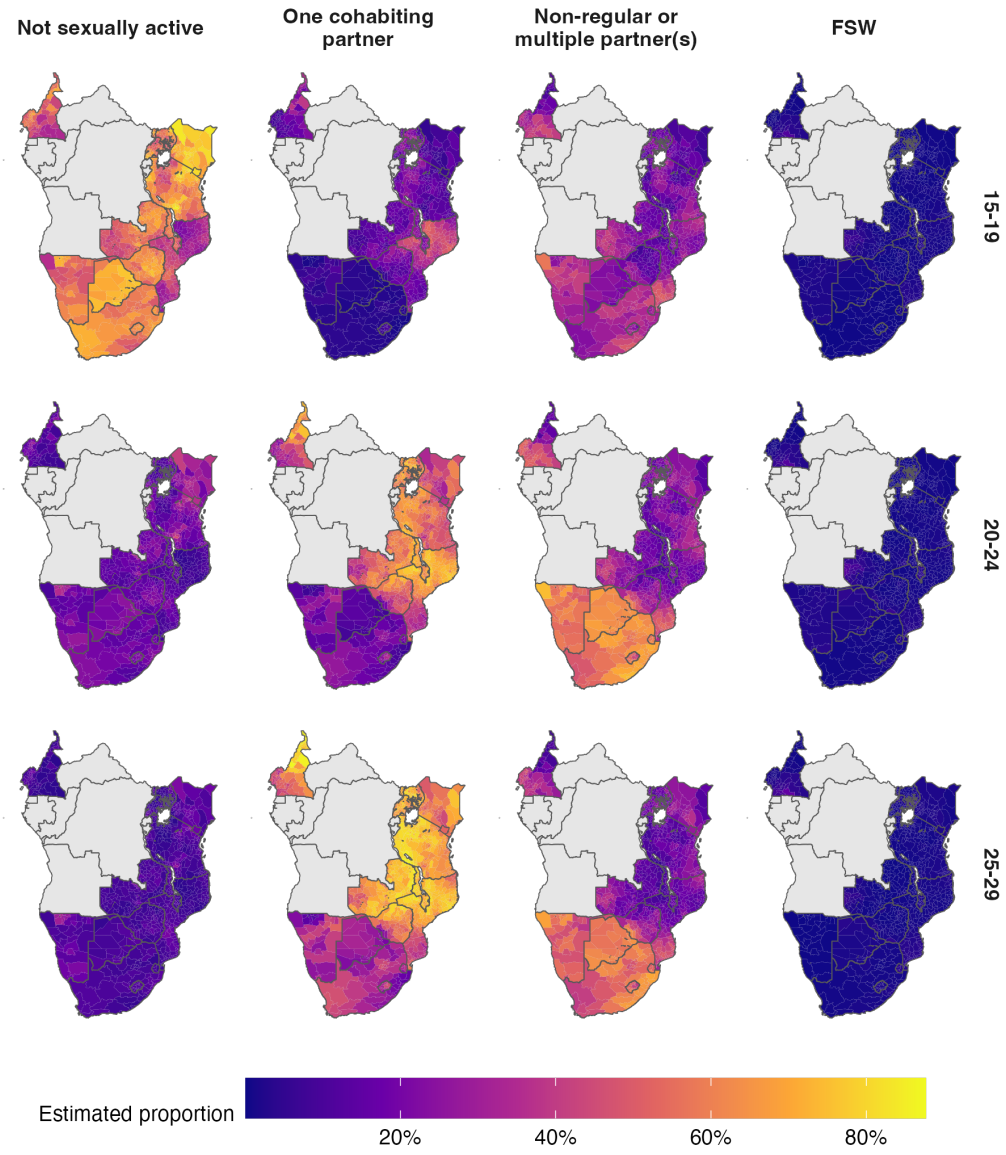

Figure 5.5: The spatial distribution (posterior mean) of the AGYW risk group proportions in 2018. Estimates are stratified by risk group (columns) and five-year age group (rows). Countries in grey were not included in the analysis.

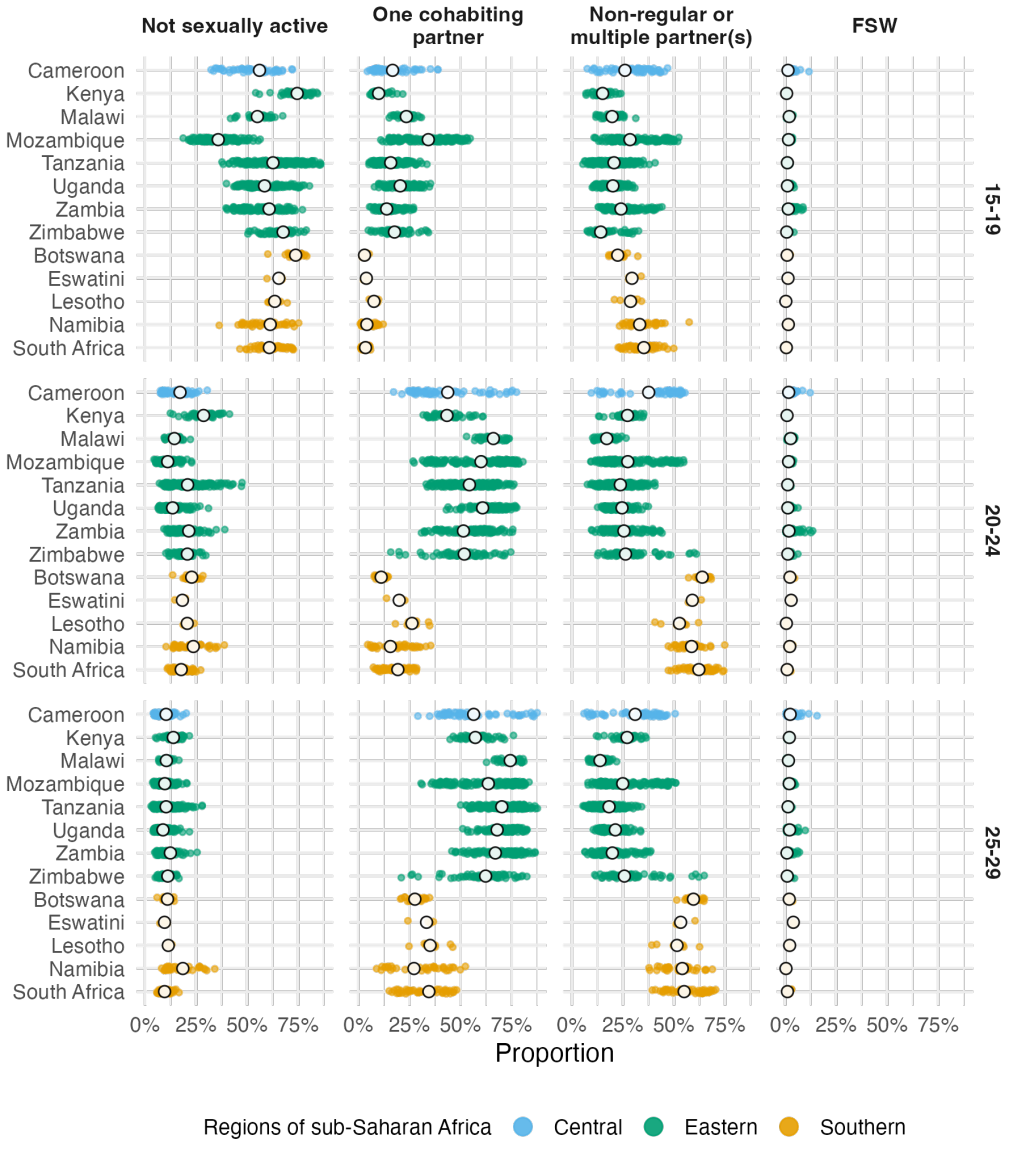

Figure 5.6: National (in white) and subnational (in color) posterior means of the risk group proportions. Estimates are stratified by risk group (columns) and five-year age group (rows).

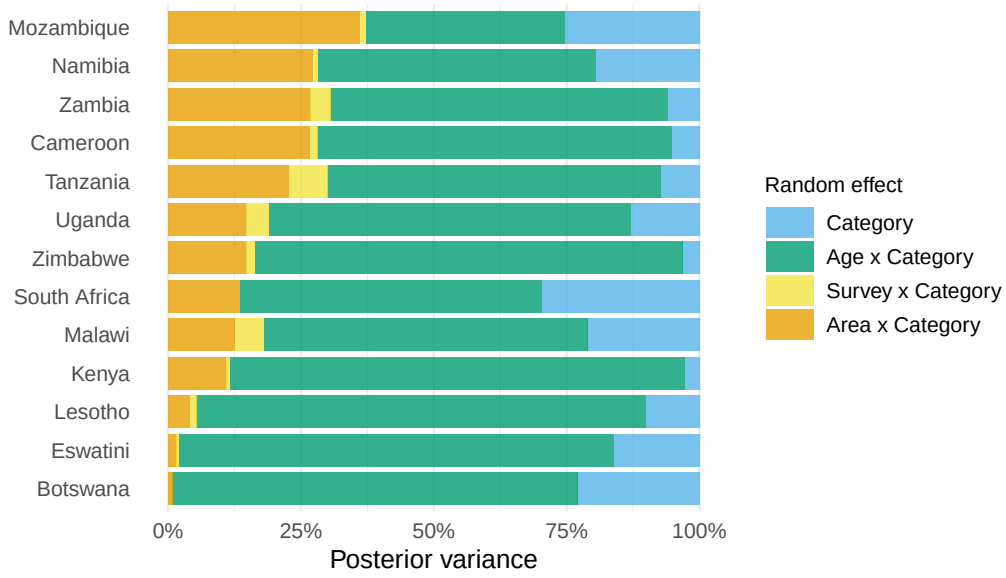

Figure 5.7: Figure caption.

Risk group specific HIV incidence estimates are then given by

$$\lambda_{ia1} = 0, \tag{5.24}$$

$$\lambda_{ia2} = I_{ia} / \left(N_{ia2} + \mathrm{RR}_3 N_{ia3} + \mathrm{RR}_4(\lambda_{ia}) N_{ia4} \right), \tag{5.25}$$

$$\lambda_{ia3} = \mathrm{RR}_3 \lambda_{ia2}, \tag{5.26}$$

$$\lambda_{ia4} = \mathrm{RR}_4(\lambda_{ia}) \lambda_{ia2}. \tag{5.27}$$

which I evaluated using Naomi model estimates of the number of new HIV infections $I_{ia} = \lambda_{ia} N_{ia}$, HIV infection risk ratios $\{\mathrm{RR}_3, \mathrm{RR}_4(\lambda_{ia})\}$, and risk group population sizes as above. The risk ratio $\mathrm{RR}_4(\lambda_{ia})$ was defined as a function of general population incidence. The number of new HIV infections are then $I_{iak} = \lambda_{iak} N_{iak}$.

5.4.2 Expected new infections reached

I calculated the number of new infections that would be reached prioritising according to each possible stratification of the population–that is for all $2^3 = 8$ possible combinations of stratification by location, age, and risk group. As an illustration, for stratification just by age, I aggregated the number of new HIV infections

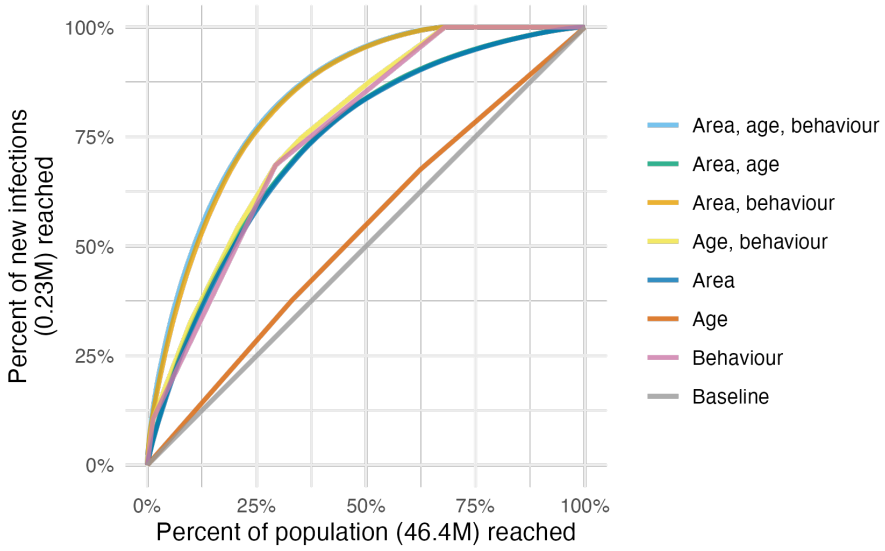

Figure 5.8: Surveys.

and HIV incidence as such

$$I_a = \sum_{ik} I_{iak}, \tag{5.28}$$

$$\lambda_a = I_a / \sum_{ik} N_{iak}. \tag{5.29}$$

Under this stratification, individuals in each age group a are prioritised according to the highest HIV incidence λ_a. By cumulatively summing the expected infections, for each fraction of the total population reached I calculated the fraction of total expected new infections that would be reached.

5.4.3 Results

5.5 Discussion

In this chapter, I estimated the proportion of AGYW who fall into different risk groups at a district level in 13 sub-Saharan African countries. These estimates support consideration of differentiated prevention programming according to geographic locations and risk behaviour, as outlined in the Global AIDS Strategy. Systematic differences in risk by age groups, and variation within and between countries, explained the large majority of variation in risk group proportions. Changes over time were negligible in the overall variation in risk group proportions.

The proportion of 15-19 year olds who are sexually active, and among women aged 20-29 years, norms around cohabitation especially varied across districts and countries. This variation underscores the need for these granular data to implement HIV prevention options aligned to local norms and risk behaviours.

I considered four risk groups based on sexual behaviour, the most proximal determinant of risk. Other factors, such as condom usage or type of sexual act, may account for additional heterogeneity in risk from sexual behaviour. However, I did not include these factors in view of measurement difficulties, concerns about consistency across contexts, and the operational benefits of describing risk parsimoniously. Sexual behaviour confers risk only when AGYW reside in geographic locations where there is unsuppressed viral load among their potential partners. I did not include more distal determinants, such as school attendance, orphanhood, or gender empowerment, as I expect their effects on risk to largely be mediated by more proximal determinants. However, to effectively implement programming, it is crucial to understand these factors, as well as the broader structural barriers and limits to personal agency faced by AGYW. Importantly, programs must ensure that intervention prioritisation occurs without stigmatising or blaming AGYW.

Brugh et al. (2021) previously geographically mapped AGYW HIV risk groups using biomarker and behavioural data from the most recent surveys in Eswatini, Haiti and Mozambique to define and subsequently map risk groups with a range of machine learning techniques. This work builds on Brugh et al. (2021) by including more countries, integrating a greater number of surveys, and connecting risk group proportions with HIV epidemic indicators to help inform programming.

By considering a range of possible risk stratification strategies, I showed that successful implementation of a risk-stratified approach would allow substantially more of those at risk for infections to be identified before infection occurs. A considerable proportion of estimated new infections were among FSW, supporting the case for HIV programming efforts focused on key population groups (Baral et al. 2012). There is substantial variation in the importance of prioritisation by age, location and behaviour within each country. This highlights the importance of

understanding and tailoring HIV prevention efforts to country-specific contexts. By standardising the analysis across all 13 countries, I showed the additional efficiency benefits of resource allocation between countries.

I found a geographic delineation in the proportion of women cohabiting between southern and eastern Africa, calling attention to a divide attributable to many cultural, social, and economic factors. The delineation does not represent a boundary between predominately Christian and Muslim populations, which is further north. I also note that the high numbers of adolescent girls aged 15-19 cohabiting in Mozambique is markedly different from the other countries (UNICEF n.d.).

My modelled estimates of risk group proportions improve upon direct survey results for three reasons. First, by taking a modular modelling approach, I integrated all relevant survey information from multiple years, allowing estimation of the FSW proportion for surveys without a specific transactional sex question. Second, whereas direct estimates exhibit large sampling variability at a district level, I alleviated this issue using spatio-temporal smoothing. Third, I provided estimates in all district-years, including those not directly sampled by surveys, allowing estimates to be consistently fed into further analysis and planning pipelines (such as the analysis of risk group specific prevalence and incidence).

The final surveys included in the risk model model were conducted in 2018. The analysis may be updated with more surveys as they become available. I do not anticipate that the risk group proportions will change substantially, as I found that they did not change significantly over time.

My analysis focused on females aged 15-29 years, and could be extended to consider optimisation of prevention more broadly, accounting for the 0% of new infections among adults 15-49 which occur in women 30-49 and men 15-49. Estimating sexual risk behaviour in adults 15-49 would be a crucial step toward greater understanding of the dynamics of the HIV epidemic in sub-Saharan Africa, and would allow incidence models to include stratification of individuals by sexual risk.

Phylogenetic results from BDI are about transmission rather than incidence. Only age-sex structured not age-sex-behaviour. Does not undermine my work.

5.5.1 Limitations

This analysis was subject to challenges shared by most approaches to monitoring sexual behaviour in the general population (Cleland et al. 2004). In particular, under-reporting of higher risk sexual behaviours among AGYW could affect the validity of my risk group proportion estimates. Due to social stigma or disapproval, respondents may be reluctant to report non-marital partners (Nnko et al. 2004; Helleringer et al. 2011) or may bias their reporting of sexual debut (Zaba et al. 2004; Wringe et al. 2009; Nguyen and Eaton 2022). For guidance of resource allocation, differing rates of under-reporting by country, district, year or age group are particularly concerning to the applicability of my results; and, while it may be reasonable to assume a constant rate over space-time, the same cannot be said for age, where aspects of under-reporting have been shown to decline as respondents age (Glynn et al. 2011), suggesting that the elevated risks I found faced by younger women are likely a conservative estimate. If present, these reporting biases will also have distorted the estimates of infection risk ratios and prevalence ratios I used in my analysis, likely over-attributing risk to higher risk groups.

I have the least confidence in my estimates for the FSW risk group. As well as having the smallest sample sizes, my transactional sex estimates do not overcome the difficulties of sampling hard to reach groups. I inherent any limitations of the national FSW estimates (Stevens, Sabin, Arias Garcia, et al. 2022) which I adjust my estimates of transactional sex to match. Furthermore, I do not consider seasonal migration patterns, which may particularly affect FSW size. More generally, I did not consider covariates potentially predictive of risk group proportions (such as sociodemographic characteristics, education, local economic activity, cultural and religious norms and attitudes), which are typically difficult to measure spatially.

Identifying measurable correlates of risk, or particular settings in which time-concentrated HIV risk occurs, is an important area for further research to improve risk prioritisation and precision HIV programme delivery.

The efficiency of each stratified prevention strategy depends on the ability of programmes to identify and effectively reach those in each strata. My analysis of new infections potentially averted assumed a "best-case" scenario where AGYW of every strata can be reached perfectly, and should therefore be interpreted as illustrating the potentially obtainable benefits rather than benefits which would be obtained from any specific intervention strategy. In practice, stratified prevention strategies are likely to be substantially less efficient than this best-case scenario. Factors I did not consider include the greater administrative burden of more complex strategies, variation in difficulty or feasibility of reaching individuals in each strata, variation in the range or effectiveness of interventions by strata, and changes in strata membership that may occur during the course of a year. Identifying and reaching behavioural strata may be particularly challenging. Empirical evaluations of behavioural risk screening tools have found only moderate discriminatory ability (Jia et al. 2022), and risk behaviour may change rapidly among young populations, increasing the challenge to effectively deliver appropriately timed prevention packages. This consideration may motivate selecting risk groups based on easily observable attributes, such as attendance of a particular service or facility, rather than sexual behaviour.

I did not engage with country experts or civil society organisations. This led to problems, including in Malawi. Future work should be better.

5.5.2 Conclusion

I estimated the proportion of AGYW aged 15-19, 20-24 and 25-29 years in four sexual risk groups at a district-level in 13 priority countries and analyzed the number of infections that could be reached by prioritisation based upon location, age and behaviour. Though subject to limitations, these estimates provide data that national HIV programmes can use to set targets and implement differentiated

HIV prevention strategies as outlined in the Global AIDS Strategy. Successfully implementing this approach would result in more efficiently reaching a greater number of those at risk of infection.

Among AGYW, there was systematic variation in sexual behaviour by age and location, but not over time. Age group variation was primarily attributable to age of sexual debut (ages 15-24). Spatial variation was particularly present between those who reported one cohabiting partner versus non-regular or multiple partners. Risk group proportions did not change substantially over time, indicating that norms relating to sexual behaviour are relatively static. These findings underscore the importance of providing effective HIV prevention options tailored to the needs of particular age groups, as well as local norms around sexual partnerships.

6

Fast approximate Bayesian inference

In this chapter I describe a novel Bayesian inference method I developed. The method is motivated the Naomi small-area estimation model. Over 35 countries have used the Naomi model software (`https://naomi.unaids.org`) to produce subnational estimates of HIV indicators (UNAIDS 2023b). Ensuring fast and accurate Bayesian inferences in this setting is challenging.

I began working on this project at the start of my PhD. However, it was only after I read Stringer et al. (2021) that I began making progress. Alex Stringer later supervised by visit to the University of Waterloo. The results of this work are presented in Howes, Stringer, et al. (2023+). Code for the analysis in this chapter is available from `athowes/naomi-aghq`.

6.1 Inference methods

In a Bayesian analysis, the primary goal is to perform inference. That is, to obtain the posterior distribution

$$p(\boldsymbol{\phi} \,|\, \mathbf{y}) = \frac{p(\boldsymbol{\phi}, \mathbf{y})}{p(\mathbf{y})}. \tag{6.1}$$

Inference is a reasonable goal because the posterior distribution is sufficient for use in decision making. Given a loss function, the posterior loss of a decision depends

on the data only via the posterior distribution.

It is usually intractable to directly obtain the posterior distribution. This is because the denominator contains an intractable integral called the posterior normalising constant

$$p(\mathbf{y}) = \int_{\mathbb{R}^d} p(\mathbf{y}, \boldsymbol{\phi}) \mathrm{d}\boldsymbol{\phi} \tag{6.2}$$

For this reason, approximations to the posterior distribution $\tilde{p}(\boldsymbol{\phi} \,|\, \mathbf{y})$ are typically used in place of the exact posterior distribution. Some approximate Bayesian inference methods avoid directly calculate the posterior normalising constant, instead working with the unnormalised posterior distribution

$$p(\boldsymbol{\phi} \,|\, \mathbf{y}) \propto p(\boldsymbol{\phi}, \mathbf{y}). \tag{6.3}$$

Other approximate Bayesian inference methods can more directly be thought of as ways to estimate the posterior normalising constant.

6.1.1 The Laplace approximation

Laplace's method (Laplace 1774) is a technique used to approximate integrals of the form

$$\int \exp(Ch(z)) \mathrm{d}z, \tag{6.4}$$

where $C > 0$ is a large constant and h is a function which is twice-differentiable. The Laplace approximation (Tierney and Kadane 1986) is obtained by application of Laplace's method to calculate the posterior normalising constant. Let $h(\boldsymbol{\phi}) = \log p(\boldsymbol{\phi}, \mathbf{y})$ such that

$$p(\mathbf{y}) = \int_{\mathbb{R}^d} p(\mathbf{y}, \boldsymbol{\phi}) \mathrm{d}\boldsymbol{\phi} = \int_{\mathbb{R}^d} \exp(h(\boldsymbol{\phi})) \mathrm{d}\boldsymbol{\phi}. \tag{6.5}$$

Laplace's method involves approximating the function h by its second order Taylor expansion evaluated at a maxima of h to eliminate the first order term. Let

$$\hat{\boldsymbol{\phi}} = \arg\max_{\boldsymbol{\phi}} h(\boldsymbol{\phi}) \tag{6.6}$$

be the posterior mode, and

$$\hat{\mathbf{H}} = -\frac{\partial^2}{\partial \boldsymbol{\phi} \partial \boldsymbol{\phi}^\top} h(\boldsymbol{\phi})|_{\boldsymbol{\phi}=\hat{\boldsymbol{\phi}}} \tag{6.7}$$

be the Hessian matrix evaluated at the posterior mode. The Laplace approximation is then

$$\tilde{p}_{\mathtt{LA}}(\mathbf{y}) = \int_{\mathbb{R}^d} \exp\left(h(\hat{\boldsymbol{\phi}}, \mathbf{y}) - \frac{1}{2}(\boldsymbol{\phi} - \hat{\boldsymbol{\phi}})^\top \hat{\mathbf{H}} (\boldsymbol{\phi} - \hat{\boldsymbol{\phi}})\right) \mathrm{d}\boldsymbol{\phi} \tag{6.8}$$

$$= p(\hat{\boldsymbol{\phi}}, \mathbf{y}) \cdot \frac{(2\pi)^{d/2}}{|\hat{\mathbf{H}}|^{1/2}}. \tag{6.9}$$

Equation (6.8) is calculated using the known normalising constant of the Gaussian distribution

$$p_{\mathtt{G}}(\boldsymbol{\phi} \,|\, \mathbf{y}) = \mathcal{N}(\boldsymbol{\phi} \,|\, \hat{\boldsymbol{\phi}}, \hat{\mathbf{H}}^{-1}) = \frac{|\hat{\mathbf{H}}|^{1/2}}{(2\pi)^{d/2}} \exp\left(-\frac{1}{2}(\boldsymbol{\phi} - \hat{\boldsymbol{\phi}})^\top \hat{\mathbf{H}} (\boldsymbol{\phi} - \hat{\boldsymbol{\phi}})\right). \tag{6.10}$$

It follows that the Laplace approximation may be thought of as approximating the posterior distribution by a Gaussian distribution $p(\boldsymbol{\phi} \,|\, \mathbf{y}) \approx p_{\mathtt{G}}(\boldsymbol{\phi} \,|\, \mathbf{y})$ such that

$$\tilde{p}_{\mathtt{LA}}(\mathbf{y}) = \left.\frac{p(\boldsymbol{\phi}, \mathbf{y})}{p_{\mathtt{G}}(\boldsymbol{\phi} \,|\, \mathbf{y})}\right|_{\boldsymbol{\phi}=\hat{\boldsymbol{\phi}}}. \tag{6.11}$$

The marginal Laplace approximation

It may be inaccurate to approximate the full joint posterior distribution using a Gaussian distribution. An alternative is to instead approximate the marginal posterior distribution of some subset of the parameters. As before, let $\boldsymbol{\phi} = (\mathbf{x}, \boldsymbol{\theta})$ where $\mathbf{x}$ is the latent field, and $\boldsymbol{\theta}$ are the hyperparameters. Applying an equivalent Laplace approximation to the latent field, we have $h(\mathbf{x}, \boldsymbol{\theta}) = \log p(\mathbf{x}, \boldsymbol{\theta}, \mathbf{y})$ with posterior mode

$$\hat{\mathbf{x}}(\boldsymbol{\theta}) = \arg\max_{\mathbf{x}} h(\mathbf{x}, \boldsymbol{\theta}) \tag{6.12}$$

and Hessian matrix evaluated at the posterior mode

$$\hat{\mathbf{H}}(\boldsymbol{\theta}) = -\frac{\partial^2}{\partial \mathbf{x} \partial \mathbf{x}^\top} h(\mathbf{x}, \boldsymbol{\theta})|_{\mathbf{x}=\hat{\mathbf{x}}(\boldsymbol{\theta})}. \tag{6.13}$$

For both quantities cases dependence on the hyperparameters $\boldsymbol{\theta}$ is made explicit. The resulting marginal Laplace approximation is then

$$\tilde{p}_{\mathtt{LA}}(\boldsymbol{\theta}, \mathbf{y}) = \int_{\mathbb{R}^N} \exp\left(h(\hat{\mathbf{x}}(\boldsymbol{\theta}), \boldsymbol{\theta}, \mathbf{y}) - \frac{1}{2}(\mathbf{x} - \hat{\mathbf{x}}(\boldsymbol{\theta}))^\top \hat{\mathbf{H}}(\boldsymbol{\theta})(\mathbf{x} - \hat{\mathbf{x}}(\boldsymbol{\theta}))\right) \mathrm{d}\mathbf{x} \tag{6.14}$$

$$= \exp(h(\hat{\mathbf{x}}(\boldsymbol{\theta}), \mathbf{y})) \cdot \frac{(2\pi)^{d/2}}{|\hat{\mathbf{H}}(\boldsymbol{\theta})|^{1/2}} \tag{6.15}$$

$$= \frac{p(\mathbf{x}, \boldsymbol{\theta}, \mathbf{y})}{\tilde{p}_{\mathtt{G}}(\mathbf{x} \,|\, \boldsymbol{\theta}, \mathbf{y})}\Big|_{\mathbf{x}=\hat{\mathbf{x}}(\boldsymbol{\theta})}, \tag{6.16}$$

where $\tilde{p}_{\mathtt{G}}(\mathbf{x} \,|\, \boldsymbol{\theta}, \mathbf{y}) = \mathcal{N}(\mathbf{x} \,|\, \hat{\mathbf{x}}(\boldsymbol{\theta}), \hat{\mathbf{H}}(\boldsymbol{\theta})^{-1})$ is a Gaussian approximation to the marginal posterior of the latent field.

6.1.2 Quadrature

Quadrature is an approach which can be used to approximate integrals like the posterior normalising constant. As with the Laplace approximation, it is deterministic in that the computational procedure is not intrinsically random.

Let $\mathcal{Q}$ be a set of quadrature points $\mathbf{z} \in \mathcal{Q}$ and $\omega : \mathbb{R}^d \to \mathbb{R}$ be a weighting function. Then a quadrature approximation to the posterior normalising constant is given by

$$\tilde{p}_{\mathcal{Q}}(\mathbf{y}) = \sum_{\mathbf{z} \in \mathcal{Q}} p(\mathbf{y}, \mathbf{z}) \omega(\mathbf{z}). \tag{6.17}$$

Quadrature methods are most effective when integrating over small dimensions. The reason why is that exponentially more quadrature points are required to cover each additional dimension. For even moderate dimension, this quickly becomes intractable.

Gauss-Hermite quadrature

Gauss-Hermite quadrature [GHQ; Davis and Rabinowitz (1975)] is a quadrature rule designed to integrate a certain class of functions exactly. These functions are the form $f(z) = \phi(z) P_\alpha(z)$, where $\phi(\cdot)$ is a standard normal density, and $P_\alpha(\cdot)$ is a polynomial of degree α.

Following the notation for GHQ established by Bilodeau et al. (2022) for $z \in \mathbb{R}$, let $H_k(z)$ be the kth Hermite polynomial

$$H_k(z) = (-1)^k \exp(z^2/2) \frac{\mathrm{d}}{\mathrm{d}z^k} \exp(-z^2/2). \tag{6.18}$$

The univariate GHQ rule has nodes $z \in \mathcal{Q}(1, k)$ given by the k zeroes of the kth Hermite polynomial. The corresponding weighting function $\omega : \mathcal{Q}(1, k) \to \mathbb{R}$ is given by

$$\omega(z) = \frac{\phi(z) \cdot k!}{[H_{k+1}(z)]^2}. \tag{6.19}$$

Multivariate GHQ rules are usually constructed using the product rule over identical univariate GHQ rules in each dimension. In d dimensions, the multivariate GHQ nodes $\mathbf{z} \in \mathcal{Q}(d, k)$ are defined by $\mathcal{Q}(d, k) = \mathcal{Q}(1, k)^d = \mathcal{Q}(1, k) \times \cdots \times \mathcal{Q}(1, k)$. The corresponding weighting function $\omega : \mathcal{Q}(d, k) \to \mathbb{R}$ is given by $\omega(\mathbf{z}) = \prod_{j=1}^{d} \omega(z_j)$.

Adaptive quadrature

In adaptive quadrature, the quadrature nodes and weights depend on the specific integrand. Adaptive quadrature rules are particularly important for Bayesian inference problems because the posterior normalising function is a function of the data. No fixed quadrature rule can be expected to perform well in integrating any posterior distributions produced by observation of particular data.

In adaptive GHQ (AGHQ).

6.1.3 Integrated nested Laplace approximation

The integrated nested Laplace approximation (INLA) method (Rue, Martino, et al. 2009) combines marginal Laplace approximations with quadrature to enable approximation of posterior marginal distributions.

6.2 Software

6.2.1 `TMB`

Template Model Builder [TMB, or when referring to the software `TMB`; Kristensen et al. (2016)] is an R package which implements the Laplace approximation. In `TMB`

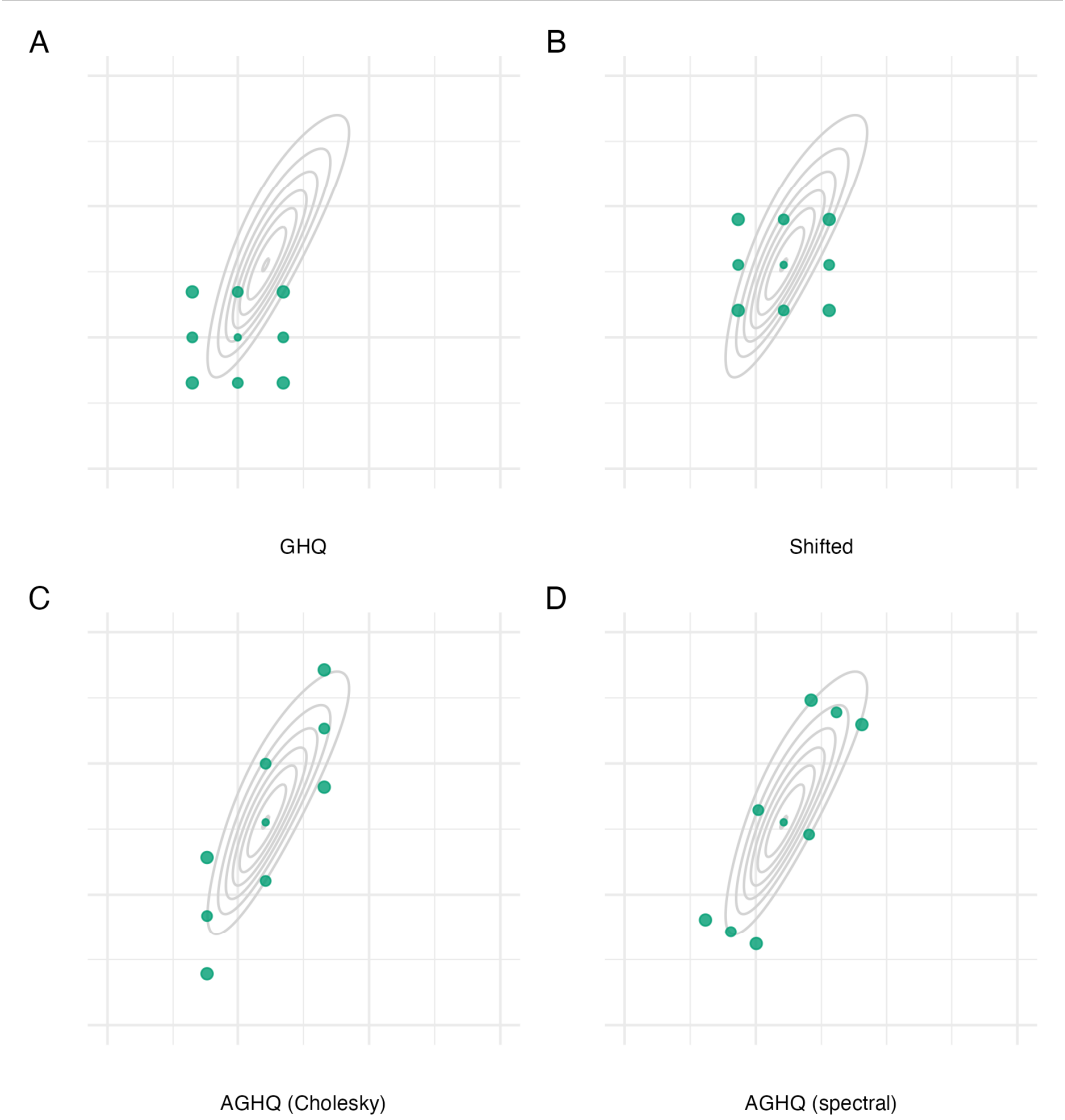

Figure 6.1: The Gauss-Hermite quadrature nodes $\mathbf{z} \in \mathcal{Q}(2, 3)$ for a two dimensional integral with three nodes per dimension (A). Adaption occurs based on the mode (B) and covariance of the integrand via either the Cholesky (C) or spectral (D) decomposition of the inverse curvature at the mode. The integrand is $f(\theta_1, \theta_2) = \text{sn}(0.5\theta_1, \alpha = 2) \cdot \text{sn}(0.8\theta_1 - 0.5\theta_2, \alpha = -2)$, where $\text{sn}(\cdot)$ is the standard skewnormal probability density function with shape parameter $\alpha \in \mathbb{R}$.

derivatives are obtained using automatic differentiation (Baydin et al. 2017).

6.2.2 `R-INLA`

The `R-INLA` software implements the INLA method. `R-INLA` uses a formula interface (e.g. `y ~ 1 + x`) to facilitate use of INLA for common models. This is a beneficial design choice for new users. For more advanced users, the formula interface can impose constraints on model choice.

6.3 A universal INLA implementation

In this section, I implement the INLA method from scratch, using the `TMB` package. The result is universal in that it is compatible with any model with a `TMB` C++ template. This opens the door for application of INLA to models like Naomi which are not compatible with `R-INLA`. Indeed, Martino and Riebler (2019) note that "implementing INLA from scratch is a complex task", and as a result "applications of INLA are limited to the (large class of) models implemented [in `R-INLA`]". The potential benefits of a more flexible INLA implementation based on automatic differentiation were noted by Skaug (2009) in discussion of Rue, Martino, et al. (2009).

6.3.1 Epilepsy example

I use the epilepsy generalised linear mixed model example from Spiegelhalter, Thomas, et al. (1996) to demonstrate the implementation. This model is based on that of Breslow and Clayton (1993), a modification of Thall and Vail (1990), and the data are from an epilespy drug double-blind clinical trial (Leppik et al. 1985). Rue, Martino, et al. (2009) (Section 5.2) demonstrate the INLA method using this example, and find a significant difference in approximation error depending on use of either the Gaussian or Laplace approximation for some parameters.

In the trial, patients $i = 1, \ldots, 59$ were each assigned either the new drug $\texttt{Trt}_i = 1$ or placebo $\texttt{Trt}_i = 0$. Each patient made four visits the clinic $j = 1, \ldots, 4$, and the observations y_{ij} are the number of seizures of the ith person in the two

weeks preceding their jth visit. The covariates used in the model were age $\texttt{Age}_i$, baseline seizure counts $\texttt{Base}_i$ and an indicator for the final clinic visit $\texttt{V}_4$, which were all centered. The observations were modelled using a Poisson distribution $y_{ij} \sim \text{Poisson}(e^{\eta_{ij}})$ with linear predictor

$$\begin{aligned}\eta_{ij} = \beta_0 + \beta_{\texttt{Base}} \log(\texttt{Baseline}_j/4) + \beta_{\texttt{Trt}}\texttt{Trt}_i + \beta_{\texttt{Trt}\times\texttt{Base}}\texttt{Trt}_i \times \log(\texttt{Baseline}_j/4) \\ + \beta_{\texttt{Age}} \log(\texttt{Age}_i) + \beta_{\texttt{V}_4}\texttt{V}_{4j} + \epsilon_i + \nu_{ij}, \quad i \in [59], \quad j \in [4],\end{aligned}$$

where the prior distribution on each of the regression parameters, including the intercept, was $\mathcal{N}(0, 100^2)$. The random effects are IID $\epsilon_i \sim \mathcal{N}(0, 1/\tau_\epsilon)$ and $\nu_{ij} \sim \mathcal{N}(0, 1/\tau_\nu)$ with precision prior distributions $\tau_\epsilon, \tau_\nu \sim \Gamma(0.001, 0.001)$.

6.4 The Naomi model

The Naomi small-area estimation model (Eaton et al. 2021) synthesises data from multiple sources to estimate HIV indicators at a district-level, by age and sex.

6.4.1 Model structure

I consider a simplified version of Naomi defined only at the time of the most recent household survey with HIV testing. This version omits nowcasting and temporal projection. These time points involve limited inferences.

Household survey component

Consider a country in sub-Saharan Africa where a household survey with complex survey design has taken place. Let $x \in \mathcal{X}$ index district, $a \in \mathcal{A}$ index five-year age group, and $s \in \mathcal{S}$ index sex. For ease of notation, let i index the finest district-age-sex division included in the model. Let $I \subseteq \mathcal{X} \times \mathcal{A} \times \mathcal{S}$ be a set of indices i for which an aggregate observation is reported, and $\mathcal{I}$ be the set of all I such that $I \in \mathcal{I}$.

Let $N_i \in \mathbb{N}$ be the known, fixed population size. HIV prevalence $\rho_i \in [0, 1]$, antiretroviral therapy (ART) coverage $\alpha_i \in [0, 1]$, and annual HIV incidence rate $\lambda_i > 0$ are modelled using linked regression equations.

Independent logistic regression models are specified for HIV prevalence and ART coverage in the general population such that $\text{logit}(\rho_i) = \eta_i^\rho$ and $\text{logit}(\alpha_i) = \eta_i^\alpha$. HIV incidence rate is modelled on the log scale as $\log(\lambda_i) = \eta_i^\lambda$, and depends on adult HIV prevalence and adult ART coverage. Let κ_i be the proportion recently infected among HIV positive persons. This proportion is linked to HIV incidence via

$$\kappa_i = 1 - \exp\left(-\lambda_i \cdot \frac{1-\rho_i}{\rho_i} \cdot (\Omega_T - \beta_T) - \beta_T\right), \tag{6.20}$$

where the mean duration of recent infection Ω_T and the proportion of long-term HIV infections misclassified as recent β_T are strongly informed by priors for the particular survey.

These processes are each informed by household survey data. Weighted aggregate survey observations are calculated as

$$\hat{\theta}_I = \frac{\sum_j w_j \cdot \theta_j}{\sum_j w_j},$$

with individual responses $\theta_j \in \{0, 1\}$ and design weights w_j for each of $\theta \in \{\rho, \alpha, \kappa\}$. The design weights are provided by the survey and aim to reduce bias by decreasing possible correlation between response and recording mechanism (Meng 2018). The index j runs across all individuals in strata $i \in I$ within the relevant denominator i.e. for ART coverage, only those individuals who are HIV positive. The weighted observed number of outcomes is $y_I^\theta = m_I^\theta \cdot \hat{\theta}_I$ where

$$m_I^\theta = \frac{\left(\sum_j w_j\right)^2}{\sum_j w_j^2},$$

is the Kish effective sample size (ESS) (Kish 1965). As the Kish ESS is maximised by constant design weights, in exchange for reducing bias the ESS is reduced and hence variance increased. The weighted observed number of outcomes are modelled using a binomial working likelihood (Chen et al. 2014) defined to operate on the reals

$$y_I^\theta \sim \text{xBin}(m_I^\theta, \theta_I),$$

where θ_I are the following weighted aggregates

$$\rho_I = \frac{\sum_{i \in I} N_i \rho_i}{\sum_{i \in I} N_i}, \quad \alpha_I = \frac{\sum_{i \in I} N_i \rho_i \alpha_i}{\sum_{i \in I} N_i \rho_i}, \quad \kappa_I = \frac{\sum_{i \in I} N_i \rho_i \kappa_i}{\sum_{i \in I} N_i \rho_i}.$$

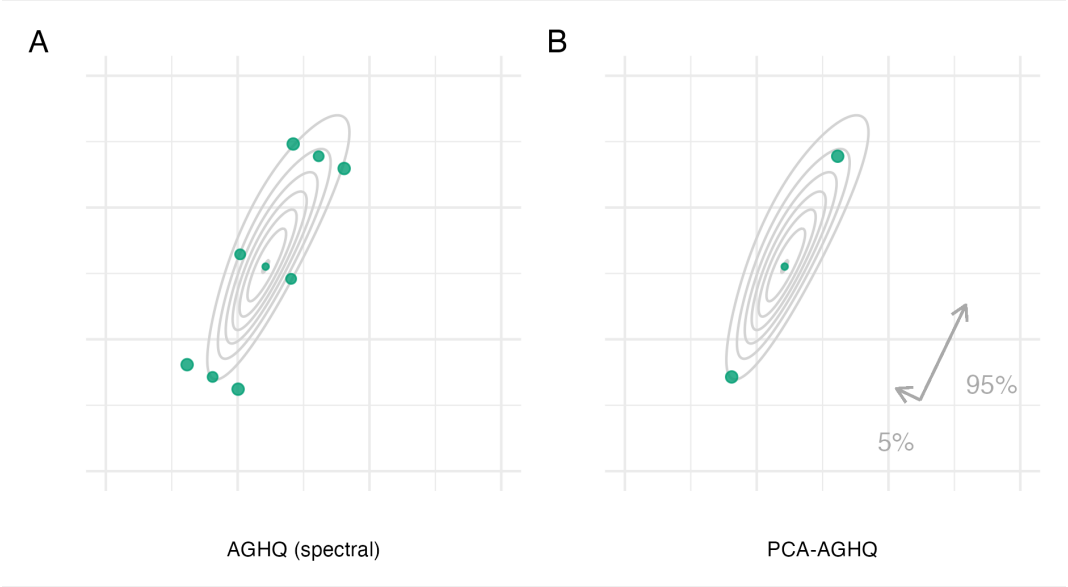

Figure 6.2: See Figure 6.1.

6.4.2 Connection to ELGMs

6.5 Extension of AGHQ to moderate dimensions

The Naomi model has $m = 24$ hyperparameters.

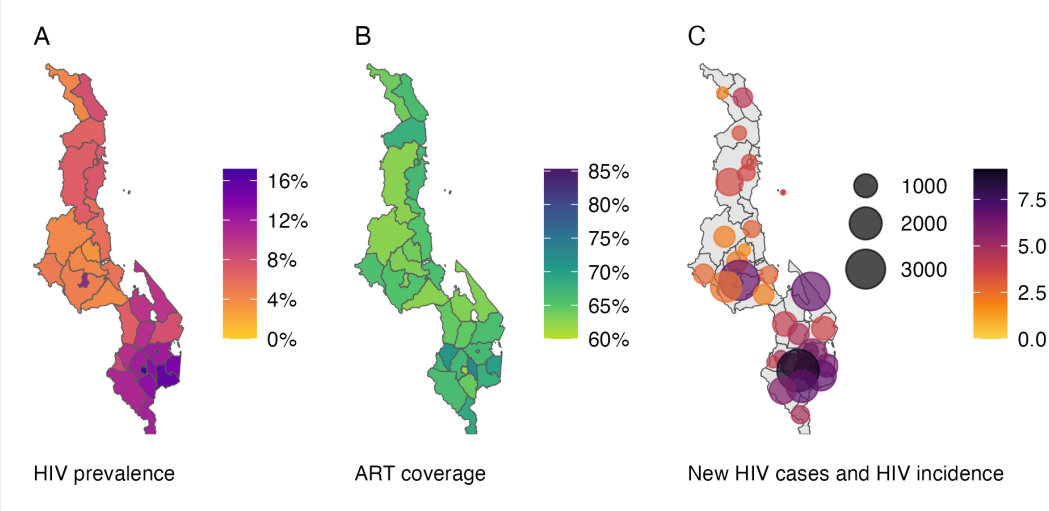

Figure 6.3: Naomi output.

6.5.1 AGHQ with variable levels

6.5.2 Principal components analysis

6.6 Malawi case-study

6.6.1 NUTS convergence

6.6.2 Use of PCA-AGHQ

6.6.3 Model assessment

6.6.4 Inference comparison

6.6.5 Exceedance probabilities

6.7 Discussion

7 Future work and conclusions

7.1 Strengths

7.1.1 Chapter 4

- I designed experiments to thoroughly compare models for spatial structure using tools for model assessment such as proper scoring rules and posterior predictive checks.

7.1.2 Chapter 5

- I estimated HIV risk group proportions for AGYW, enabling countries to prioritise their delivery of HIV prevention services.
- I analysed the number of new infections that might be reached under a variety of risk stratification strategies.
- I used `R-INLA` to specify multinomial spatio-temporal models via the Poisson-multinomial transformation. This includes complex two- and three-way Kronecker product interactions defined using the `group` and `replicate` options.

7.1.3 Chapter 6

- I developed a novel Bayesian inference method, motivated by a challenging and practically important problem in HIV inference.
- The method enables integrated nested Laplace approximations to be fit to and studied on a wider class of models than was previously possible.
- My implementation of the method was straightforward, building on the `TMB` and `aghq` packages, and described completely and accessibly in Howes, Stringer, et al. (2023+).

7.2 Future work

Avenues for future work include:

1. Extending the risk group model described in Chapter 5 to include all adults 15-49. This may involve modelling of age-stratified sexual partnerships (Wolock et al. 2021). Such a model would likely fall out of the scope of `R-INLA`, but would be possible to write with `TMB` and therefore amenable to the methods discussed in Chapter 6.
2. Speeding up the implementation of Laplace marginals using the matrix algebra approximations described in Wood (2020).
3. Evaluating the accuracy of deterministic Bayesian inference methods for a broader variety of extended latent Gaussian models.

7.3 Conclusions

- Modelling complex data, more often than not, pushes the boundaries of the statistical toolkit available.
- A challenge I encountered was the difficulty of implementing identical models across multiple frameworks with the aim of studying the inference method. Or, of a similarly fraught nature, comparing different models implemented in different frameworks with the aim of studying model differences. The

frequently asked questions section of the `R-INLA` website (Rue 2023) notes that "the devil is in the details". I have resolved this challenge by using a given `TMB` model template to fit models using multiple inference methodologies. The benefits of such a ecosystem of packages are noted by Stringer (2021). I particularly highlight the benefit of enabling analysts to easily vary their choice of inference method based on the stage of model development that they are in.

- To the best of my abilities, I have written this thesis, and the work described within it, in keeping with the principles of open science. I hope that doing so allows my work to be scrutinised, and optimistically built upon. This would not have been possible without a range of tools from the R ecosystem such as `rmarkdown` and `rticles`, as well as those developed within the MRC Centre for Global Infectious Disease Analysis such as `orderly` and `didehpc`.

Appendices

A

Spatial structure

B

A model for risk group proportions

B.1 The Global AIDS Strategy

Prioritisation strata	Criterion
Low	0.3-1.0% incidence and low-risk behaviour, or <0.3% incidence and high-risk behaviour
Moderate	1.0-3.0% incidence and low-risk behaviour, or 0.3-1.0% incidence and high-risk behaviour
High	1.0-3.0% incidence and high-risk behaviour
Very high	>3.0% incidence

Table B.1: Prioritisation strata according to HIV incidence in the general population and behavioural risk.

Intervention	Low	Moderate	High	Very High
Condoms and lube for those with non-regular partners(s) with unknown STI status and not on PrEP	50%	70%	95%	95%
STI screening and treatment	10%	10%	80%	80%
Access to PEP	-	-	50%	90%
PrEP use	-	5%	50%	50%
Economic empowerment	-	-	20%	20%

Table B.2: Commitments to be met for each intervention in terms of proportion of the prioritisation strata reached. The symbol "-" represents no commitment.

B.2 Household survey data

	Type	Year	Transactional sex question	Sample size			
				15-19	20-24	25-29	Total
Botswana							
	BAIS	2013	✓	557	588	649	1794
Total				557	588	649	1794
Cameroon							
	DHS	2004	✗	2675	2207	1732	6614
	DHS	2011	✗	3588	3115	2655	9358
	PHIA	2017	✗	2620	2339	2259	7218
	DHS	2018	✓	3349	2463	2345	8157
Total				12232	10124	8991	31347
Kenya							
	DHS	2003	✗	1819	1709	1391	4919
	DHS	2008	✗	1767	1743	1419	4929
	DHS	2014	✗	2861	2534	2858	8253
Total				6447	5986	5668	18101
Lesotho							
	DHS	2004	✗	1761	1455	1026	4242
	DHS	2009	✗	1833	1543	1194	4570
	DHS	2014	✗	1537	1292	1067	3896
	PHIA	2017	✓	1156	1202	1054	3412
Total				6287	5492	4341	16120
Mozambique							
	AIS	2009	✗	1031	1106	987	3124
	DHS	2011	✗	2932	2299	2206	7437
	AIS	2015	✗	1552	1389	1080	4021
Total				5515	4794	4273	14582
Malawi							
	DHS	2000	✗	2914	2998	2358	8270
	DHS	2004	✗	2407	2823	2135	7365
	DHS	2010	✗	5031	4387	4309	13727
	DHS	2015	✓	5273	5094	3976	14343
	PHIA	2016	✓	1646	1934	1511	5091
Total				17271	17236	14289	48796
Namibia							
	DHS	2000	✗	1427	1313	1098	3838

	DHS	2006	✗	2203	1869	1544	5616
	DHS	2013	✗	1852	1709	1481	5042
	PHIA	2017	✓	1491	1525	1370	4386
Total				6973	6416	5493	18882
Eswatini							
	DHS	2006	✗	1265	1027	731	3023
	PHIA	2017	✗	1031	895	811	2737
Total				2296	1922	1542	5760
Tanzania							
	AIS	2003	✗	1466	1377	1270	4113
	AIS	2007	✗	2137	1676	1509	5322
	DHS	2010	✗	2221	1860	1613	5694
	AIS	2012	✗	2474	1923	1815	6212
	PHIA	2016	✓	2999	2845	2521	8365
Total				11297	9681	8728	29706
Uganda							
	DHS	2000	✗	1687	1541	1326	4554
	DHS	2006	✗	1948	1660	1404	5012
	AIS	2011	✗	2451	2164	1921	6536
	DHS	2011	✗	2025	1664	1614	5303
	DHS	2016	✓	4276	3782	3014	11072
	PHIA	2016	✗	3289	3059	2574	8922
Total				15676	13870	11853	41399
South Africa							
	DHS	2016	✓	1505	1408	1397	4310
Total				1505	1408	1397	4310
Zambia							
	DHS	2007	✗	1598	1405	1373	4376
	DHS	2013	✗	3685	3036	2789	9510
	PHIA	2016	✓	2120	2045	1619	5784
	DHS	2018	✓	3112	2687	2166	7965
Total				10515	9173	7947	27635
Zimbabwe							
	DHS	1999	✗	1467	1230	1011	3708
	DHS	2005	✗	2128	1943	1438	5509
	DHS	2010	✗	1963	1796	1679	5438
	DHS	2015	✓	2154	1777	1646	5577
	PHIA	2016	✓	2114	1817	1573	5504
Total				9826	8563	7347	25736

Total	106397	95253	82518	284168

Table B.3: All of the surveys that used in the analysis and their sample sizes, disaggregated by respondent age.

Survey	Exclusion reason
MOZ2003DHS	No GPS coordinates available to place survey clusters within districts.
TZA2015DHS	Insufficient sexual behaviour questions.
UGA2004AIS	Unable to download region boundaries.
ZMB2002DHS	No GPS coordinates available to place survey clusters within districts.

Table B.4: All of that surveys that were excluded from the analysis.

B.3 Spatial analysis levels

Country	Number of areas	Analysis level
Botswana	27	3
Cameroon	58	2
Kenya	47	2
Lesotho	10	1
Mozambique	161	3
Malawi	33	5
Namibia	38	2
Eswatini	4	1
Tanzania	195	4
Uganda	136	3
South Africa	52	2
Zambia	116	2
Zimbabwe	63	2

Table B.5: The numer of areas and analysis levels for each country that were used in the analysis.

B.4 Survey questions and risk group allocation

Variable(s)	Description
`v501`	Current marital status of the respondent.
`v529`	Computed time since last sexual intercourse.
`v531`	Age at first sexual intercourse–imputed.
`v766b`	Number of sexual partners during the last 12 months (including husband).
`v767[a, b, c]`	Relationship with last three sexual partners. Options are: spouse, boyfriend not living with respondent, other friend, casual acquaintance, relative, commercial sex worker, live-in partner, other.
`v791a`	Had sex in return for gifts, cash or anything else in the past 12 months. Asked only to women 15-24 who are not in a union.

Table B.6: The survey questions included in AIDS Indicator Survey (AIS) and Demographic and Health Surveys (DHS).

Variable(s)	Description
`part12monum`	Number of sexual partners during the last 12 months (including husband).
`part12modkr`	Reason for leaving `part12monum` blank.
`partlivew[1, 2, 3]`	Does the person you had sex with live in this household?
`partrelation[1, 2, 3]`	Relationship with last three sexual partners. Options are: husband, live-in partner, partner (not living with), ex-spouse/partner, friend/acquaintance, sex worker, sex worker client, stranger, other, don't know, refused.
`sellsx12mo`	Had sex for money and/or gifts in the last 12 months.
`buysx12mo`	Paid money or given gifts for sex in the last 12 months.

Table B.7: The survey questions included in Population-Based HIV Impact Assessment (PHIA) surveys.

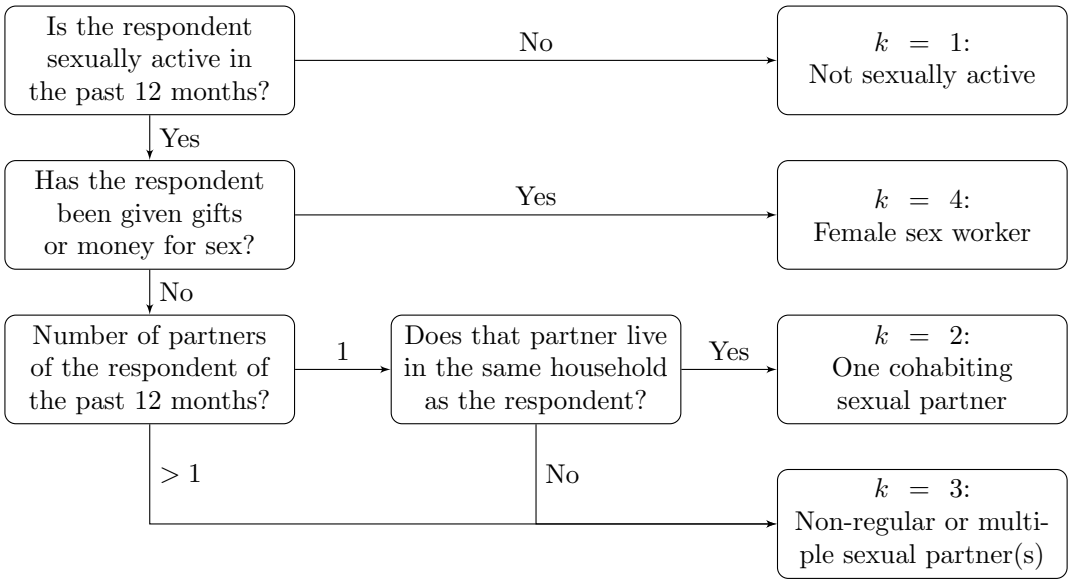

Figure B.1: Flowchart describing allocation of survey respondents to HIV risk groups.

C

Fast approximate Bayesian inference

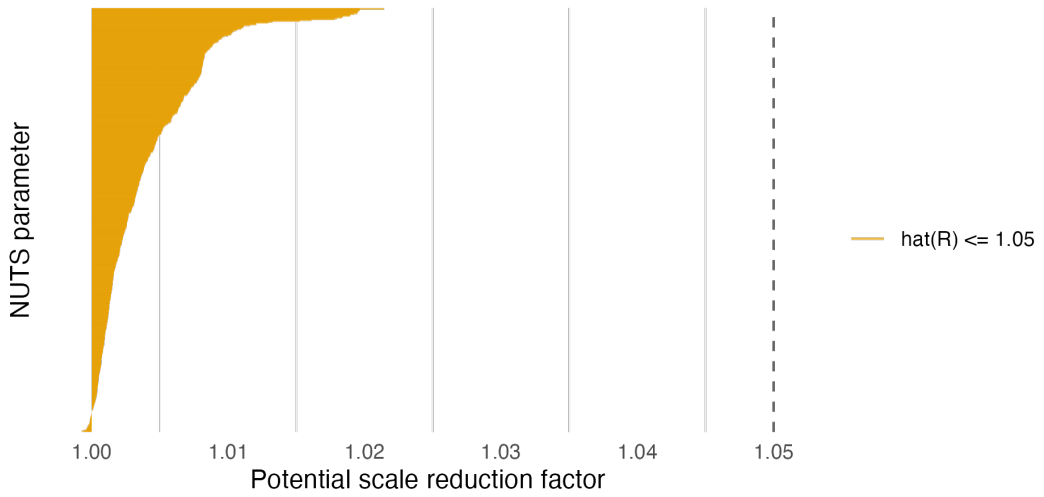

Figure C.1: The potential scale reduction factor compares between- and within- estimates of univariate parameters. It is recommended only to use NUTS results if the value is less than 1.05, which it is for all parameters.

Works Cited

Auvert, Bertran et al. (2005). "Randomized, controlled intervention trial of male circumcision for reduction of HIV infection risk: the ANRS 1265 Trial". In: *PLoS medicine* 2.11, e298.

Bailey, Robert C et al. (2007). "Male circumcision for HIV prevention in young men in Kisumu, Kenya: a randomised controlled trial". In: *The Lancet* 369.9562, pp. 643–656.

Baker, Stuart G (1994). "The multinomial-Poisson transformation". In: *Journal of the Royal Statistical Society: Series D (The Statistician)* 43.4, pp. 495–504.

Baral, Stefan et al. (2012). "Burden of HIV among female sex workers in low-income and middle-income countries: a systematic review and meta-analysis". In: *The Lancet Infectious Diseases* 12.7, pp. 538–549.

Barré-Sinoussi, Françoise et al. (1983). "Isolation of a T-lymphotropic retrovirus from a patient at risk for acquired immune deficiency syndrome (AIDS)". In: *Science* 220.4599, pp. 868–871.

Baydin, Atılım Günes et al. (2017). "Automatic differentiation in machine learning: a survey". In: *The Journal of Machine Learning Research* 18.1, pp. 5595–5637.

Besag, Julian, Jeremy York, and Annie Mollié (1991). "Bayesian image restoration, with two applications in spatial statistics". In: *Annals of the Institute of Statistical Mathematics* 43.1, pp. 1–20.

Bilodeau, Blair, Alex Stringer, and Yanbo Tang (2022). "Stochastic convergence rates and applications of adaptive quadrature in Bayesian inference". In: *Journal of the American Statistical Association*, pp. 1–11.

Bosse, Nikos I. et al. (2022). *Evaluating Forecasts with scoringutils in R*. DOI: 10.48550/ARXIV.2205.07090. URL: https://arxiv.org/abs/2205.07090.

Breslow, Norman E and David G Clayton (1993). "Approximate inference in generalized linear mixed models". In: *Journal of the American statistical Association* 88.421, pp. 9–25.

Brugh, Kristen N et al. (2021). "Characterizing and mapping the spatial variability of HIV risk among adolescent girls and young women: A cross-county analysis of population-based surveys in Eswatini, Haiti, and Mozambique". In: *PLoS One* 16.12, e0261520.

Carpenter, Bob et al. (2017). "Stan: A probabilistic programming language". In: *Journal of Statistical Software* 76.1.

Chen, Cici, Jon Wakefield, and Thomas Lumely (2014). "The use of sampling weights in Bayesian hierarchical models for small area estimation". In: *Spatial and spatio-temporal epidemiology* 11, pp. 33–43.

Cleland, John et al. (2004). "Monitoring sexual behaviour in general populations: a synthesis of lessons of the past decade". In: *Sexually Transmitted Infections* 80.suppl 2, pp. ii1–ii7.

Cohen, Myron S et al. (2011). "Prevention of HIV-1 infection with early antiretroviral therapy". In: *New England journal of medicine* 365.6, pp. 493–505.

Cressie, Noel and Christopher K Wikle (2015). *Statistics for spatio-temporal data*. John Wiley & Sons.

Davis, Philip J and Philip Rabinowitz (1975). *Methods of numerical integration*. Academic Press.

Dawid, A Philip (1984). "Present position and potential developments: Some personal views statistical theory the prequential approach". In: *Journal of the Royal Statistical Society: Series A (General)* 147.2, pp. 278–290.

Dean, CB, MD Ugarte, and AF Militino (2001). "Detecting interaction between random region and fixed age effects in disease mapping". In: *Biometrics* 57.1, pp. 197–202.

DHS (2012). *Sampling and Household Listing Manual: Demographic and Health Surveys Methodology*.

Duane, Simon et al. (1987). "Hybrid Monte Carlo". In: *Physics letters B* 195.2, pp. 216–222.

Eaton, Jeffrey W et al. (2021). "Naomi: A New Modelling Tool for Estimating HIV Epidemic Indicators at the District Level in Sub-Saharan Africa". In.

Economist Impact (2023). "A triple dividend: the health, social and economic gains from financing the HIV response in Africa". In.

Fisher, Ronald Aylmer (1936). "Design of experiments". In: *British Medical Journal* 1.3923, p. 554.

Freni-Sterrantino, Anna, Massimo Ventrucci, and Håvard Rue (2018). "A note on intrinsic conditional autoregressive models for disconnected graphs". In: *Spatial and spatio-temporal epidemiology* 26, pp. 25–34.

Gelman, Andrew (2005). "Analysis of variance—why it is more important than ever". In.

— (2007). "Struggles with survey weighting and regression modeling". In.

Gelman, Andrew, John B Carlin, et al. (2013). *Bayesian data analysis*. CRC press.

Gelman, Andrew, Daniel Simpson, and Michael Betancourt (2017). "The prior can often only be understood in the context of the likelihood". In: *Entropy* 19.10, p. 555.

Global Burden of Disease Collaborative Network (2019). *Global Burden of Disease Study 2019 (GBD 2019) Results*. URL: `https://vizhub.healthdata.org/gbd-results/`.

Glynn, Judith R et al. (2011). "Assessing the validity of sexual behaviour reports in a whole population survey in rural Malawi". In: *PLoS One* 6.7, e22840.

Gottlieb, Michael S et al. (1981). "Pneumocystis pneumonia—Los Angeles". In: *Mmwr* 30.21, pp. 1–3.

Gray, Ronald H et al. (2007). "Male circumcision for HIV prevention in men in Rakai, Uganda: a randomised trial". In: *The Lancet* 369.9562, pp. 657–666.

Hájek, Jaroslav (1971). "Discussion of 'An essay on the logical foundations of survey sampling, part I'". In: *Foundations of Statistical Inference (Proc. Sympos., Univ. Waterloo, Ontario, 1970)*, p. 236.

Helleringer, Stéphane et al. (2011). "The reliability of sexual partnership histories: implications for the measurement of partnership concurrency during surveys". In: *AIDS (London, England)* 25.4, p. 503.

Hodgins, Caroline et al. (2022). "Population sizes, HIV prevalence, and HIV prevention among men who paid for sex in sub-Saharan Africa (2000–2020): A meta-analysis of 87 population-based surveys". In: *PLoS Medicine* 19.1, e1003861.

Hoffman, Matthew D, Andrew Gelman, et al. (2014). "The No-U-Turn sampler: adaptively setting path lengths in Hamiltonian Monte Carlo." In: *J. Mach. Learn. Res.* 15.1, pp. 1593–1623.

Howes, Adam, Jeffrey W. Eaton, and Seth R. Flaxman (2023+). "Beyond borders: evaluating the suitability of spatial adjacency for small-area estimation". In.

Howes, Adam, Kathryn A. Risher, et al. (Apr. 2023). "Spatio-temporal estimates of HIV risk group proportions for adolescent girls and young women across 13 priority countries in sub-Saharan Africa". In: *PLOS Global Public Health* 3.4, pp. 1–14. DOI: 10.1371/journal.pgph.0001731. URL: https://doi.org/10.1371/journal.pgph.0001731.

Howes, Adam, Alex Stringer, et al. (2023+). "Fast approximate Bayesian inference of HIV indicators using PCA adaptive Gauss-Hermite quadrature". In.

Jia, Katherine M et al. (2022). "Risk scores for predicting HIV incidence among adult heterosexual populations in sub-Saharan Africa: a systematic review and meta-analysis". In: *Journal of the International AIDS Society* 25.1, e25861.

Johnson, L and RE Dorrington (2020). "Thembisa version 4.3: A model for evaluating the impact of HIV/AIDS in South Africa". In: *View Article*.

Khoury, Muin J, Michael F Iademarco, and William T Riley (2016). "Precision public health for the era of precision medicine". In: *American journal of preventive medicine* 50.3, pp. 398–401.

Kish, Leslie (1965). *Survey sampling*. 04; HN29, K5.

Kristensen, Kasper et al. (2016). "TMB: Automatic Differentiation and Laplace Approximation". In: *Journal of Statistical Software* 70.i05.

Laplace, P. S. (1774). "Memoire sur la probabilite de causes par les evenements". In: *Memoire de l'Academie Royale des Sciences*.

Leppik, IE et al. (1985). "A double-blind crossover evaluation of progabide in partial seizures". In: *Neurology* 35.4, p. 285.

Leroux, Brian G, Xingye Lei, and Norman Breslow (2000). "Estimation of disease rates in small areas: a new mixed model for spatial dependence". In: *Statistical Models in Epidemiology, the Environment, and Clinical Trials*. Springer, pp. 179–191.

Martino, Sara and Andrea Riebler (2019). "Integrated nested Laplace approximations (INLA)". In: *arXiv preprint arXiv:1907.01248*.

Martins, Thiago G et al. (2013). "Bayesian computing with INLA: new features". In: *Computational Statistics & Data Analysis* 67, pp. 68–83.

McCullagh, Peter and John A Nelder (1989). *Generalized linear models*. Routledge.

McElreath, Richard (2020). *Statistical rethinking: A Bayesian course with examples in R and Stan*. CRC press.

Meng, Xiao-Li (2018). "Statistical paradises and paradoxes in big data (i) law of large populations, big data paradox, and the 2016 us presidential election". In: *The Annals of Applied Statistics* 12.2, pp. 685–726.

Monod, Mélodie et al. (2023). "Growing gender disparity in HIV infection in Africa: sources and policy implications". In: *medRxiv*, pp. 2023–03.

Neal, Radford M et al. (2011). "MCMC using Hamiltonian dynamics". In: *Handbook of Markov chain Monte Carlo* 2.11, p. 2.

Nguyen, Van Kính and Jeffrey W. Eaton (2022). "Trends and country-level variation in age at first sex in sub-Saharan Africa among birth cohorts entering adulthood between 1985 and 2020". In: *BMC Public Health* 22.1, p. 1120. DOI: 10.1186/s12889-022-13451-y. URL: https://doi.org/10.1186/s12889-022-13451-y.

Nnko, Soori et al. (2004). "Secretive females or swaggering males?: An assessment of the quality of sexual partnership reporting in rural Tanzania". In: *Social Science & Medicine* 59.2, pp. 299–310.

Openshaw, S and P.J. Taylor (1979). "A million or so correlation coefficients, three experiments on the modifiable areal unit problem". In: *Statistical Applications in the Spatial Science*, pp. 127–144.

Ord, Toby (2013). "The moral imperative toward cost-effectiveness in global health". In: *Center for Global Development* 12.

Paciorek, Christopher J et al. (2013). "Spatial models for point and areal data using Markov random fields on a fine grid". In: *Electronic Journal of Statistics* 7, pp. 946–972.

Pettit, LI (1990). "The conditional predictive ordinate for the normal distribution". In: *Journal of the Royal Statistical Society: Series B (Methodological)* 52.1, pp. 175–184.

Risher, Kathryn A et al. (2021). "Age patterns of HIV incidence in eastern and southern Africa: a modelling analysis of observational population-based cohort studies". In: *The Lancet HIV* 8.7, e429–e439.

Robert, Christian P and George Casella (2005). *Monte Carlo Statistical Methods (Springer Texts in Statistics)*.

Rue, Håvard (2020). "Comment on R-INLA Discussion Group thread". In.

Rue, Havard (2023). "'R-INLA' Project - FAQ". Accessed 23/01/2023. URL: `https://www.r-inla.org/faq`.

Rue, Havard and Leonhard Held (2005). *Gaussian Markov random fields: theory and applications*. CRC press.

Rue, Håvard, Sara Martino, and Nicolas Chopin (2009). "Approximate Bayesian inference for latent Gaussian models by using integrated nested Laplace approximations". In: *Journal of the Royal Statistical Society: Series B (Statistical Methodology)* 71.2, pp. 319–392.

Säilynoja, Teemu, Paul-Christian Bürkner, and Aki Vehtari (2021). "Graphical Test for Discrete Uniformity and its Applications in Goodness of Fit Evaluation and Multiple Sample Comparison". In: *arXiv preprint arXiv:2103.10522*.

Saul, Janet et al. (2018). "The DREAMS core package of interventions: a comprehensive approach to preventing HIV among adolescent girls and young women". In: *PLoS One* 13.12, e0208167.

Siegfried, Nandi et al. (2011). "Antiretrovirals for reducing the risk of mother-to-child transmission of HIV infection". In: *Cochrane database of systematic reviews* 7.

Simpson, Daniel et al. (2017). "Penalising model component complexity: A principled, practical approach to constructing priors". In: *Statistical Science* 32.1, pp. 1–28.

Skaug, Hans J. (2009). "Discussion of "Approximate Bayesian inference for latent Gaussian models by using integrated nested Laplace approximations"". In: vol. 71. 2. Wiley Online Library, pp. 319–392.

Slaymaker, Emma et al. (2020). "Risk factors for new HIV infections in the general population in sub-Saharan Africa". In.

Sørbye, Sigrunn Holbek and Håvard Rue (2014). "Scaling intrinsic Gaussian Markov random field priors in spatial modelling". In: *Spatial Statistics* 8, pp. 39–51.

— (2017). "Penalised complexity priors for stationary autoregressive processes". In: *Journal of Time Series Analysis* 38.6, pp. 923–935.

Spiegelhalter, David, Andrew Thomas, et al. (1996). "BUGS 0.5 Examples". In: *MRC Biostatistics Unit, Institute of Public health, Cambridge, UK* 256.

Spiegelhalter, David J, Nicola G Best, et al. (2002). "Bayesian measures of model complexity and fit". In: *Journal of the Royal Statistical Society: Series B (statistical Methodology)* 64.4, pp. 583–639.

Stevens, Oliver, Keith Sabin, Sonia Arias Garcia, et al. (2022). "Estimating key population size, HIV prevalence, and ART coverage for sub-Saharan Africa at the national level". In.

Stevens, Oliver, Keith Sabin, Sonia Arias Garcia, et al. (2022). "Key population size, HIV prevalence, and ART coverage in sub-Saharan Africa: systematic collation and synthesis of survey data". In: *medRxiv*, pp. 2022–07.

Stover, John and Yu Teng (2021). "The impact of condom use on the HIV epidemic". In: *Gates Open Research* 5.

Stringer, Alex (2021). "Implementing Approximate Bayesian Inference Using Adaptive Quadrature". Statistics Graduate Student Research Day 2021, The Fields Institute for Research in Mathematical Sciences. URL: `http://www.fields.utoronto.ca/talks/Implementing-Approximate-Bayesian-Inference-Using-Adaptive-Quadrature`.

Stringer, Alex, Patrick Brown, and Jamie Stafford (2021). "Fast, Scalable Approximations to Posterior Distributions in Extended Latent Gaussian Models". In: *arXiv preprint arXiv:2103.07425*.

Tatem, Andrew J (2017). "WorldPop, open data for spatial demography". In: *Scientific data* 4.1, pp. 1–4.

Thall, Peter F and Stephen C Vail (1990). "Some covariance models for longitudinal count data with overdispersion". In: *Biometrics*, pp. 657–671.

The Global Fund (2018). *The Global Fund Measurement Framework for Adolescent Girls and Young Women Programs*. Accessed 30/08/2021. URL: `https://www.theglobalfund.org/media/8076/me_adolescentsgirlsandyoungwomenprograms_frameworkmeasurement_en.pdf`.

Tierney, Luke and Joseph B Kadane (1986). "Accurate approximations for posterior moments and marginal densities". In: *Journal of the American Statistical Association* 81.393, pp. 82–86.

Tobler, Waldo R (1970). "A computer movie simulating urban growth in the Detroit region". In: *Economic geography* 46.sup1, pp. 234–240.

U.S. Department of State (2022). *Latest Global Program Results*. `https://www.state.gov/wp-content/uploads/2022/11/PEPFAR-Latest-Global-Results_December-2022.pdf`. Accessed: 10/08/2023.

UNAIDS (2021a). *2021 UNAIDS Global AIDS Update - Confronting Inequalities - Lessons for pandemic responses from 40 Years of AIDS*. Accessed: June 2023.

— (2021b). "Global AIDS strategy 2021–2026. End inequalities. End AIDS". In: Accessed: June 2023.

— (2022). *In Danger: UNAIDS Global AIDS Update 2022*. `https://www.unaids.org/en/resources/documents/2022/in-danger-global-aids-update`. Accessed: June 2023.

— (2023a). *AIDSinfo: Global data on HIV epidemiology and response*. `https://aidsinfo.unaids.org/`. Accessed: August 2023.

— (2023b). *The path that ends AIDS: UNAIDS Global AIDS Update 2023*. `https://www.unaids.org/en/resources/documents/2023/global-aids-update-2023`. Accessed: August 2023.

UNICEF (n.d.). *Adolescent & social norms situation in Mozambique*. Accessed 25/03/2022. URL: https://www.unicef.org/mozambique/en/adolescent-social-norms.

Wakefield, Jonathan and Hilary Lyons (Mar. 2010). "Spatial Aggregation and the Ecological Fallacy". In: vol. 2010, pp. 541–558. DOI: 10.1201/9781420072884-c30.

Watanabe, Sumio (2013). "A widely applicable Bayesian information criterion". In: *Journal of Machine Learning Research* 14.Mar, pp. 867–897.

Wolock, Timothy M et al. (June 2021). "Evaluating distributional regression strategies for modelling self-reported sexual age-mixing". In: *eLife* 10. Ed. by Eduardo Franco, Talía Malagón, and Adam Akullian, e68318. DOI: 10.7554/eLife.68318. URL: https://doi.org/10.7554/eLife.68318.

Wood, Simon N (2020). "Simplified integrated nested Laplace approximation". In: *Biometrika* 107.1, pp. 223–230.

World Health Organization (2005). "Guidelines for measuring national HIV prevalence in population-based surveys". In.

— (2017). "Consolidated guidelines on person-centred HIV patient monitoring and case surveillance". In.

Wringe, A et al. (2009). "Comparative assessment of the quality of age-at-event reporting in three HIV cohort studies in sub-Saharan Africa". In: *Sexually transmitted infections* 85.Suppl 1, pp. i56–i63.

Yao, Yuling et al. (2018). "Yes, but did it work?: Evaluating variational inference". In: *International Conference on Machine Learning*. PMLR, pp. 5581–5590.

Zaba, Basia et al. (2004). "Age at first sex: understanding recent trends in African demographic surveys". In: *Sexually transmitted infections* 80.suppl 2, pp. ii28–ii35.