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

September 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

Thanks to Seth Flaxman and Jeff Eaton for supervision of this research; staff and students of the StatML CDT at Imperial and Oxford; members of the HIV Inference Group at Imperial; members of the Machine Learning and Global Health Network; the Bill & Melinda Gates Foundation and EPSRC via the StatML CDT for funding this PhD; Mike McLaren, Kevin Esvelt, the Nucleic Acid Observatory team, and the Sculpting Evolution lab for hosting my visit to the MIT Media Lab; Alex Stringer for hosting my visit to Waterloo; the Effective Altruism community; my friends and family.

Adam Howes
Imperial College London
September 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 by the Global AIDS Strategy. Finally, I develop a novel 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 Bayesian in that I use probability models to arrive at conclusions based on data. It is applied and methodological in that I am concerned first with real world questions and second with the means to answer them.

The real world questions relate to surveillance of the human immunodeficiency virus (HIV) epidemic in sub-Saharan Africa (SSA). It is estimated that 66% of the 39 million people living with HIV (PLHIV) reside in SSA. 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 numerous 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. One important feature of this data is its location in space and time. I use modelling techniques from the field of spatio-temporal statistics. While diverse, spatio-temporal data have distinctive commonalities which can can be utilised across settings.

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

1.1 Chapter overview

In this section, I provide a brief overview of the chapters to follow.

- 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.
- 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 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. Motivated by this model, 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: I discuss avenues for future work, and my conclusions regarding the research, as well as its strengths and weaknesses.

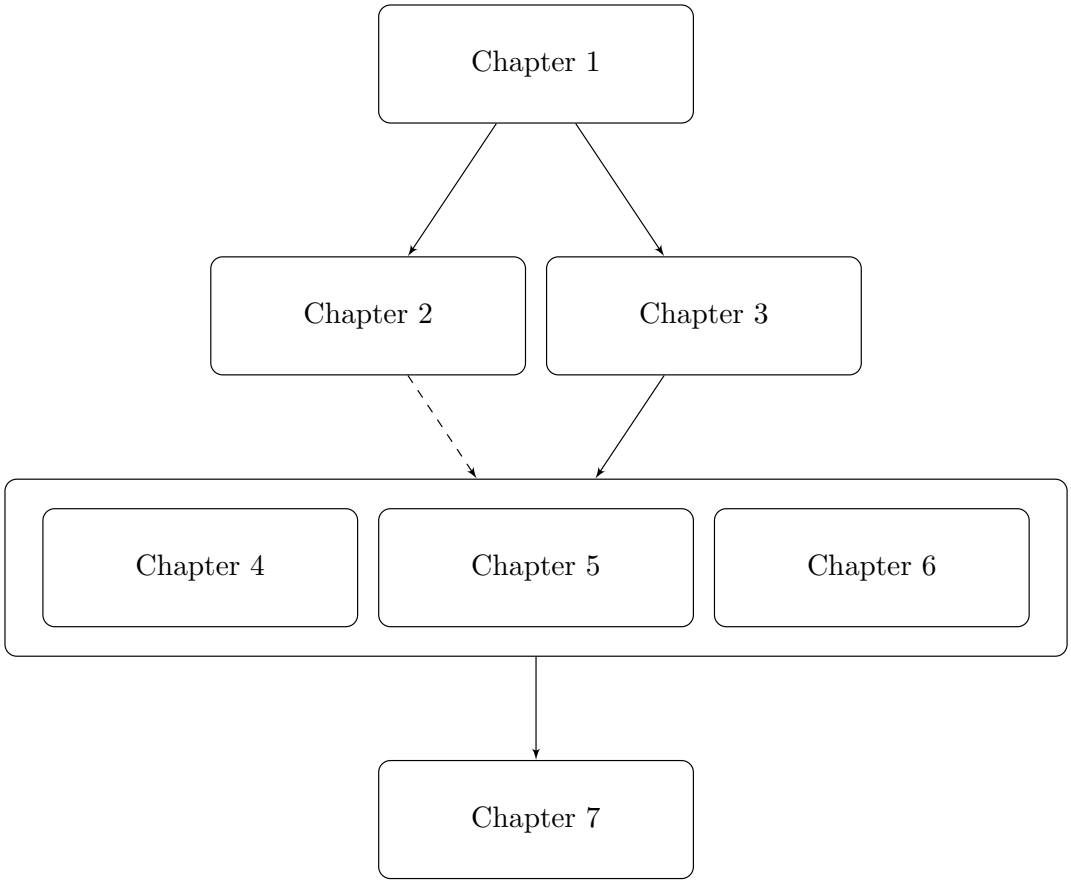

Figure 1.1: Flowchart presenting the recommended chapter dependencies for this thesis. The dashed line represents a recommended, but not required dependency.

Figure 1.1 gives the dependency structure for the chapters. Though chronological order is recommended, Chapter 4, Chapter 5 and Chapter 6 may be read in any order as, for the most part, they correspond to separable research projects.

2

The HIV/AIDS epidemic

2.1 Background

HIV is a retrovirus which infects humans. If untreated, HIV can 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 *Pneumocystis* pneumonia and tuberculosis.

The first AIDS cases were reported in Los Angeles in the early 1980s. 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.

A major international effort has been made to address the epidemic, including by national governments, civil society organisations, pharmaceutical companies, research institutions, international agencies such as the Joint United Nations Programme on HIV/AIDS (UNAIDS), and global health initiatives such as the President's Emergency Plan for AIDS Relief (PEPFAR; $85 billion invested) and the Global Fund to Fight AIDS, Tuberculosis, and Malaria (GFATM; $29.2 billion

invested). A number of interventions have had a substantial impact on the course of the epidemic:

- Antiretroviral therapy (ART) is a drug which stops the virus from replicating in the body. By taking ART daily, a person living with HIV can live a full and healthy life. Of the 39 million people living with HIV (PLHIV) in 2022, around 76 are accessing ART. As a result, ART is estimated to have averted 16.5 million AIDS-related deaths. Furthermore, if the virus is undetectable then it cannot be transmitted sexually. For this reason, treatment also operates as prevention (TaSP). ART has been estimated to have averted this many new infections.
- Condoms are an inexpensive and effective method for prevention of HIV and other sexually transmitted infections (STIs). Scale-up of condom usage since 1990 has been estimated to have averted over 10 million new HIV infections (Stover and Teng 2021).
- Pre-exposure prophylaxis (PrEP) and post-exposure prophylaxis (PEP) are drugs which can be taken before and after exposure to prevent transmission.
- Voluntary medical male circumcision (VMMC) has been found to provide partial protection against female-to-male HIV transmission.

The epidemic response has seen both the yearly number of new HIV cases and number of AIDS-related deaths fall significantly since their respective peaks (Figure 2.1). However, there remains significant distance to go in ending AIDS as a public health threat: in 2022 there were 1.3 million people newly infected with HIV and 630 thousand AIDS-related deaths (UNAIDS 2022). Furthermore progress is halting due to lack of investment, COVID-19 and other reasons.

There is substantial geographic inequality in disease burden. In some countries, the epidemic is concentrated within small populations, and prevalence is low. In others, transmission is sustained in the general population, and prevalence is higher. Most of the countries severely affected by HIV are in sub-Saharan Africa (SSA),

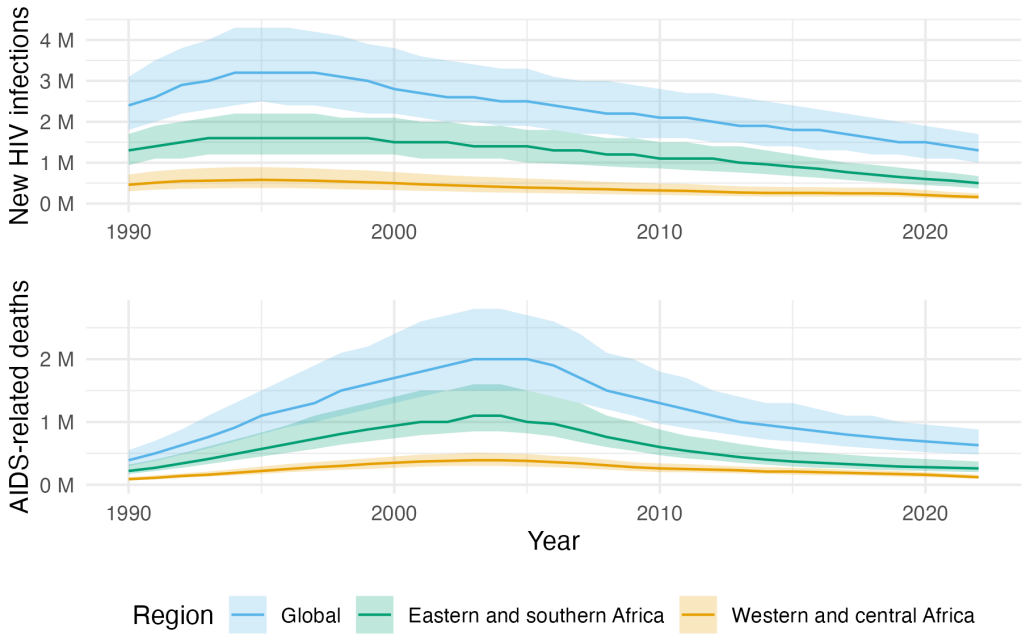

Figure 2.1: Globally, yearly new HIV infections peaked in 1995, and have since decreased by 59% and yearly AIDS-related deaths peaked in 2004, and have since decreased by 68%. Data from UNAIDS epidemiological estimates 2023.

which accounts for 66% of PLHIV worldwide. Even within countries, there is significant geographic variation.

Across all settings, the inequality in disease burden extends to groups. Groups at increased risk of HIV infection are known as key populations (KPs), and include men who have sex with men (MSM), female sex workers (FSW), people who inject drugs (PWID), and transgender people (TGP). KPs are often marginalised, and face legal and social issues.

In SSA, risk is more diffuse than in concentrated settings. Demographic groups, while not classified as KPs, at increased risk of HIV infection include adolescent girls and young women (AGYW), migratory populations, and rural communities.

HIV interventions should be prioritised to have the greatest impact on the epidemic. The precision public health paradigm aims to get the right interventions, to the right populations, in the right place, at the right time. This requires data.

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.

2.2.1 HIV indicators

Important indicators include:

- **HIV prevalence** The proportion of the population who have HIV ρ, typically given as a percentage. Though an important indicator, HIV prevalence is for the most part used indirectly to calculate other indicators rather than directly itself. In some circumstances those other indicators can be difficult to estimate, and HIV prevlance can be a useful proxy. The number of PLHIV is given by $N\rho$, where N is the population size.
- **HIV incidence** The rate of new HIV infections λ, typically given 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.
- **ART coverage** The proportion of PLHIV who are on ART α, typically given as a percentage. Planning of treatment.

2.2.2 Challenges

There are significant difficulties associated with obtaining reliable, timely, estimates of these indicators at an appropriate resolution.

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.
2. **Response biases** Under-reporting. Non-response. Measurement error.

3. **Hard to reach populations** Many of those at greatest risk, including KPs, are difficult to reach. May not appear on household survey sampling frames.
4. **Denominators and demography** Many indicators are rates or proportions, incorporating the population at risk in the denominator. However, accurately estimating these denominators is usually itself a challenging task, and taking a ratio of uncertain quantities only goes to amplify the uncertainty. For example, ART coverage has PLHIV as the denominator, which itself has the population size as its denominator.

2.2.3 Statistical methods

The above challenges make direct interpretation of data misleading, and foreground the importance of careful statistical modelling.

1. **Borrowing information** Although survey responses from women aged 20-24 are not directly informative about HIV prevalence in women 25-29, they are indirectly informative.
2. **Evidence synthesis** Multiple sources of evidence can be combined to overcome limitations of any one data source.
3. **Expert guidance** Expert epidemiological, demographic opinion can be used to improve estimates.

2.2.4 Future trends

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

1. **Greater reliance on routine health system data** Phasing out of nationally representative household surveys.
2. **Integration with other health programs**
3. **Use of mobile technologies**

3

Bayesian spatio-temporal statistics

3.1 Bayesian statistics

Bayesian statistics is a mathematical paradigm for learning from data. It is well suited for attending to the challenges in Section 2.2 because it allows principled, flexible incorporation of 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 $\boldsymbol{\phi} = (\phi_1, \ldots, \phi_d)$, where 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 $p(\boldsymbol{\phi})$ whereby 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. 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 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 the numerator is proportional to the posterior distribution $p(\boldsymbol{\phi} \mid \mathbf{y}) \propto p(\mathbf{y} \mid \boldsymbol{\phi})p(\boldsymbol{\phi})$ and easy to evaluate, it is typically difficult to evaluate the posterior distribution itself.

A great variety of computational methods have been developed to tackle this problem. Briefly, the main categories are:

1. **Sampling algorithms** Also known as Monte Carlo algorithms, these approaches look to generate samples from the posterior distribution. 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), a Hamiltonian Monte Carlo (HMC) algorithm, implemented in the Stan probabilistic programming language (PPL).
2. **Variational Bayes** In variational Bayes, the posterior distribution is assumed to belong to a particular class of functions and use optimisation to choose the best member of that class.

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.

Correlation structure is an important feature of spatio-temporal data.

Bayesian

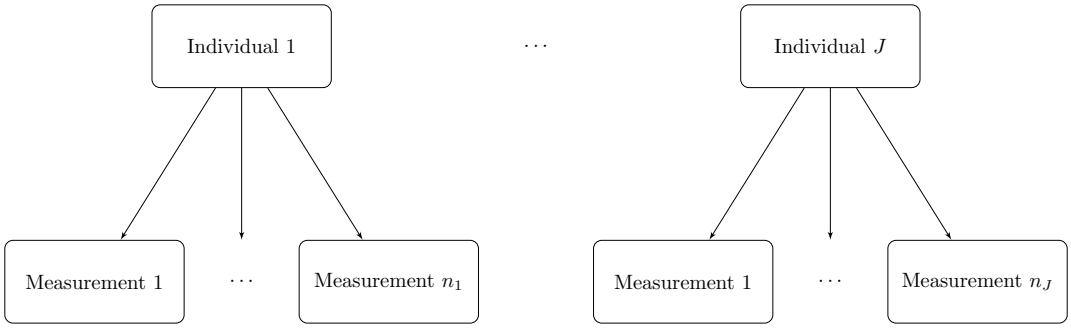

Figure 3.1: Data often has a hierarchical structure.

3.3 Model classes

3.3.1 Hierarchical models

Real world data usually has hierarchical structure. For example, we might have a study (Figure 3.1) where each individual $j = 1, \ldots, J$ has a group of n_j observations. Then, it would be inappropriate to treat all $\sum_j n_j$ observations as being independent and identically distributed (IID). Observations belonging to the same group, that is of the same individual, are likely to be more similar than observations from different groups, which refer to different individuals.

Bayesian hierarchical or multilevel models, comprised of multiple stages, allow handling data of this sort. For example, in a three-stage 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 $p(\mathbf{x}, \boldsymbol{\theta} \,|\, \mathbf{y}) \propto p(\mathbf{y} \,|\, \mathbf{x}, \boldsymbol{\theta}) p(\mathbf{x} \,|\, \boldsymbol{\theta}) p(\boldsymbol{\theta})$. Referring back to the previous example, we may informally think about the latent field as modelling properties of individuals, and the hyperparameters as modelling shared properties of individuals or the observation process.

Bayesian

3.3.2 Mixed effects models

Fixed effects refer to parts of the latent field which are constant across groups. Random effects refer to parts of the latent field which vary across groups. Though these terms have notoriously many different definitions (Gelman 2005), I nonetheless find them useful to introduce here. No pooling, partial pooling, complete pooling. Structured random effects.

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, loosely speaking, the middle layer is Gaussian. More specifically, 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 $p(\boldsymbol{\theta})$.

Spatio-temporal data are well suited to being modelled with LGMs.

3.3.4 Extended latent Gaussian models

Some models used in small-area estimation fall outside the LGM class. Many of these models do fit into the class of extended latent Gaussian models (ELGMs) as proposed by Stringer et al. (2021). By allowing many-to-one link functions, ELGMs facilitate modelling of non-linearities. In particular, the structured additive predictor is redefined as $\boldsymbol{\eta} = (\eta_1, \dots \eta_{N_n})$, 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 $[N_n] = \{1, \dots, N_n\}$. 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 then of the form

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

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 in which each observation depends on the set of structured additive predictors corresponding to each multinomial observation. In Chapter 6, the Naomi small-area estimation model used to produce estimates of HIV indicators across over 35 countries is shown to have the features of an ELGM.

3.4 Survey methods

Survey methods concern the design and analysis of survey data (Lohr 2009). In SSA, nationally representative household surveys, such as the Demographic and Health Surveys (DHS) or Population-based HIV Impact Assessment (PHIA) surveys,

provide the highest quality information about HIV. Talk about the DDC (Meng 2018). With regard to the DDC, discuss how survey weighting is helpful. Discuss how we weight then model following Chen et al. (2014). Discuss how we ignore clustering structure. Discuss how all of this isn't great and that someone should figure this out.

4 Spatial structure

In this chapter, I present an investigation of spatial random effects specifications. This investigation was motivated by a particular instance of a fundamental question encountered during model construction. Namely, should the model be augmented to capture a feature we have prior belief exists? The results are presented in Howes, Eaton, et al. (2023+).

Code for the analysis in this chapter is available from `athowes/beyond-borders` and supported by the R package `arealutils`.

4.1 Background

4.2 Models based on adjacency

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 $I = I_1, \ldots, I_{n_c}$.

4.3 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. My primary role was to develop the statistical model. I built on 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 using the estimates which is now being used by countries to guide policy. Code for the analysis in this chapter is available from `athowes/multi-agyw` and supported by the R package `multi.utils` (Howes 2023).

5.1 Background

Adolescent girls and young women (AGYW, defined here as females aged 15-29) are a demographic group at increased risk of HIV infection. Though AGHYW 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 are given in Table 5.1.

Table 5.1: AYGW are at higher risk of HIV infection for interacting social and biological reasons.

Reason	Description
Structural vulnerability and power imbalances	Gender inequality, lack of agency, economic factors, limited healthcare access, stigma and discrimination
Age patterns of sexual mixing	Older male partners have higher prevalence of HIV
Younger age at first sex	Text
Increased susceptibility to HIV infection	Immature reproductive tract, including a thinner cervical lining and higher levels of immune cells

On this basis, AGYW have been identified as a priority population for HIV prevention services (Saul et al. 2018; The Global Fund 2018). The Global AIDS Strategy 2021-2026 (UNAIDS 2021b) proposed stratifying HIV prevention packages to AGYW based on 1) local population-level HIV incidence and 2) individual-level sexual risk behaviour. As risk depends on both factors, prioritisation of prevention services would be 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. Estimates of the size of each risk group are required.

Table 5.2: Behavioural risk groups.

Risk group	Description	Local HIV incidence	Incidence ratio
None	Not sexually active	–	0.0
Low	One cohabiting partner	–	1.0 (Baseline)
High	Non-regular or multiple partner(s)	–	1.72
Very High	Transactional sex (adjusted to correspond to female sex workers)	0.1-0.3%	13.0
		0.3-1.0%	9.0

Risk group	Description	Local HIV incidence	Incidence ratio
		1.0-3.0%	6.0
		>3.0%	3.0

5.2 Data

I used household survey data from 13 designated AGYW priority countries.: Botswana, Cameroon, Kenya, Lesotho, Malawi, Mozambique, Namibia, South Africa, Eswatini, Tanzania, Uganda, Zambia and Zimbabwe.

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

1. Not sexually active
2. One cohabiting sexual partner
3. Non-regular or multiple sexual partner(s), and
4. AGYW who report 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.

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. Due to this incomparability across surveys, I did not include

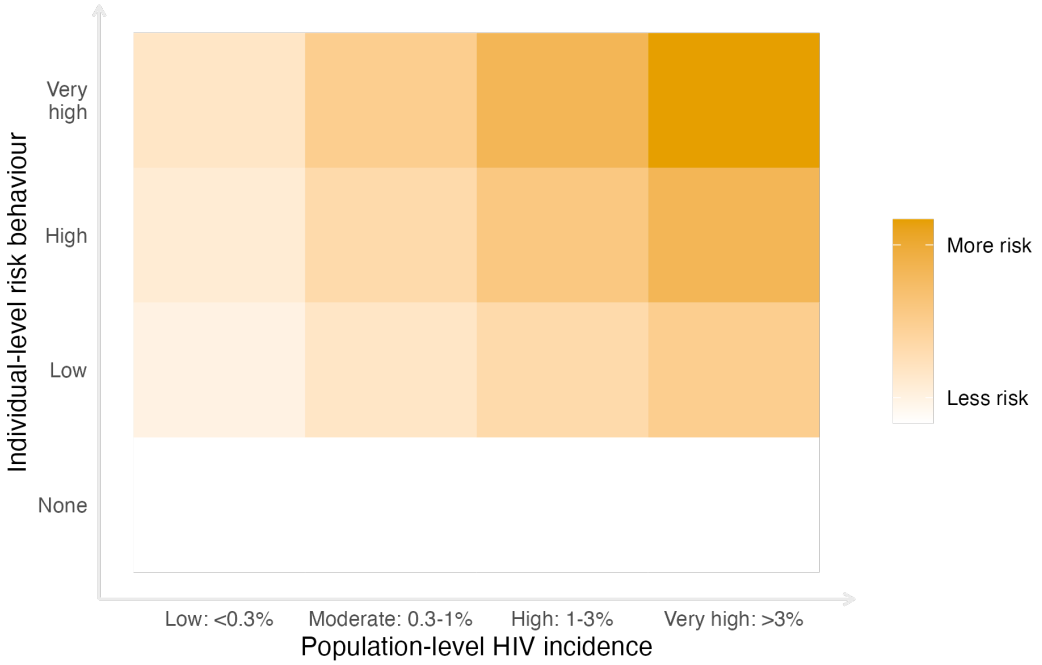

Figure 5.1: Risk depends on both individual-level risk behaviour and population-level HIV incidence.

surveys without a specific transactional sex question when estimating the proportion of the population who engaged in transactional sex. I instead focused on estimating the proportion of women who reported transactional sex at a district level, and subsequently adjusted these proportions to align to national estimates for the number of female sex workers.

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. Index the four risk groups as $k \in \{1, 2, 3, 4\}$, and denote being in either the third or fourth risk group by $k = 3^+$. 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 baseline category multinomial logistic regression model 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, and is not possible to fit in `R-INLA`.

The multinomial-Poisson transformation

I use the multinomial-Poisson transformation to enable inference with `R-INLA`. The transformation reframes a given multinomial logisitic regression model as an

equivalent Poisson log-linear model 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 are then obtained by 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 $m_{ita} = \sum_k y_{itak}$ are treated as random $m_{ita} \sim \text{Poisson}\,(\kappa_{ita})$ rather than fixed. The joint distribution of $p(\mathbf{y}_{ita}, m_{ita}) = p(\mathbf{y}_{ita} \,|\, m_{ita})p(m_{ita})$ is then

$$p(\mathbf{y}_{ita}, m_{ita}) = \exp(-\kappa_{ita}) \frac{(\kappa_{ita})^{m_{ita}}}{m_{ita}!} \times \frac{m_{ita}!}{\prod_k y_{itak}!} \prod_k \left(\frac{\kappa_{itak}}{\kappa_{ita}}\right)^{y_{itak}} \tag{5.7}$$

$$= \prod_k \left(\frac{\exp(-\kappa_{itak})\,(\kappa_{itak})^{y_{itak}}}{y_{itak}!}\right) \tag{5.8}$$

$$= \prod_k \text{Poisson}\,(y_{itak} \,|\, \kappa_{itak})\,. \tag{5.9}$$

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

This model, including random sample sizes, is equivalent to the multinomial logistic regression only when the normalisation constants m_{ita} are recovered exactly. To ensure that this is the case, one approach is to include observation-specific random effects θ_{ita} 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 $\theta_{ita} \propto 1$ 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.

Model specifications

I considered four models for η_{ita} in Equation 5.4 of the form

$$\eta_{ita} = \theta_{ita} + \beta_k + \zeta_{c[i]k} + \alpha_{ac[i]k} + \phi_{ik} + \gamma_{tk}. \tag{5.10}$$

Observation random effects $\theta_{ita} \sim \mathcal{N}(0, 1000^2)$ were included in all models I considered, and are required for the multinomial-Poisson transformation to be 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 ϕ_{ik} random effects I considered two specifications:

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

$$\boldsymbol{\phi} = (\phi_{11}, \ldots, \phi_{n1}, \ldots, \phi_{13^+}, \ldots \phi_{n3^+})^\top \sim \mathcal{N}(\mathbf{0}, (\tau_\phi \mathbf{R}_\phi^\star)^-).$$

   The scaled structure matrix $\mathbf{R}_\phi^\star = \mathbf{R}_b^\star \otimes \mathbf{I}$ 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. Scaling of the structure matrix to have generalised variance one ensures interpretable priors may be placed on the precision parameter (Sørbye and Rue 2014). I followed the further recommendations of Freni-Sterrantino et al. (2018) with regard to disconnected adjacency graphs, singletons and constraints. The Besag structure matrix $\mathbf{R}_b$ was obtained by the precision matrix of the random effects $\mathbf{b} = (b_1, \ldots, b_n)^\top$ with full conditionals

$$b_i \,|\, \mathbf{b}_{-i} \sim \mathcal{N}\left(\frac{\sum_{j:j\sim i} b_j}{n_{\delta i}}, \frac{1}{n_{\delta i}}\right), \tag{5.11}$$

   where $j \sim i$ if the districts A_i and A_j are adjacent, and $n_{\delta i}$ is the number of districts adjacent to A_i.

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

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

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 $|\rho| < 1$.

Priors All random effect precision parameters $\tau \in \{\tau_\beta, \tau_\zeta, \tau_\alpha, \tau_\phi, \tau_\gamma\}$ were given independent penalised complexity (PC) priors (Simpson et al. 2017) with base model $\sigma = 0$ given by $p(\tau) = 0.5\nu\tau^{-3/2}\exp\left(-\nu\tau^{-1/2}\right)$ 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, 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 ensure interpretable posterior inferences relating to the random effects, 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. Together, these were:

1. Category $\sum_k \beta_k = 0$,
2. Country $\sum_c \zeta_{ck} = 0$, $\forall\, k$,
3. Age-country $\sum_a \alpha_{ack} = 0$, $\forall\, c, k$,
4. Spatial $\sum_i \phi_{ik} = 0$, $\forall\, k$,
5. Temporal $\sum_t \gamma_{tk} = 0$, $\forall\, k$.

Survey weighted likelihood

I included surveys which use a complex design, in which each individual has an unequal probability of being included in the sample. For example the DHS often employs a two-stage cluster design, first taking an urban rural stratified sample of ennumeration areas, before selecting households from each ennumeration area using systematic sampling (DHS 2012).

To account for this aspect of survey design, I use a weighted pseudo-likelihood where the observed counts y are replaced by effective counts $y^\star$ calculated using the survey weights w_j of all individuals j in the corresponding strata. I multiplied direct estimates produced using the `survey` package (Lumley 2004) by the Kish effective sample size (Kish 1965)

$$m^\star = \frac{\left(\sum_j w_j\right)^2}{\sum_j {w_j}^2} \tag{5.14}$$

to obtain $y^\star$. These counts may not be integers, and as such the Poisson likelihood I used in Equation 5.3 is not appropriate. Instead, I used a generalised Poisson pseudo-likelihood $y^\star \sim \text{xPoisson}(\kappa)$, given by

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

as implemented by `family = "xPoisson"` in `R-INLA`, which accepts non-integer input.

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 $q_{ia} = p_{ia4}/(p_{ia3} + p_{ia4}) = p_{ia4}/p_{ia3^+}$. Taking this two-step approach allowed me to include all surveys in the multinomial regression model, but only those surveys with a specific transactional sex question in Equation 5.16. As all such surveys occurred in 2013-2018, in the logistic regression model I assumed q_{ia} to be constant over time.

Model specifications

I considered six logistic regression models. Each included a constant $\beta_0 \sim \mathcal{N}(-2, 1^2)$, 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 prior 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 ϕ_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`, available from Hodgins et al. (2022). For both random effect precision parameters $\tau \in \{\tau_\alpha, \tau_\zeta\}$ I used the PC prior with base model $\sigma = 0$ and $\mathbb{P}(\sigma > 2.5 = 0.01)$. For the regression parameters $\beta \in \{\beta_{\texttt{cfswever}}, \beta_{\texttt{cfswrecent}}\}$ I used the prior $\beta \sim \mathcal{N}(0, 2.5^2)$.

Survey weighted likelihood

As with the multinomial regression model, I used survey weighted counts $\{y^\star_{itak}\}$ and sample sizes $\{m^\star_{itak}\}$. I used a generalised binomial pseudo-likelihood $y^\star \sim \text{xBinomial}(m^\star, q)$, as implemented by `family = "xBinomial"` in `R-INLA`, given by

$$p(y^\star \mid m^\star, q) = \binom{\lfloor m^\star \rfloor}{\lfloor y^\star \rfloor} q^{y^\star} (1-q)^{m^\star - y^\star}. \tag{5.18}$$

to extend the binomial distribution to non-integer weighted counts and sample sizes.

5.3.3 Female sex worker population size adjustment

Responding "yes" to the question "have you had sex in return for gifts, cash or anything else in the past 12 months" is not considered sufficient to constitute sex work. Recognising this, I adjusted the estimates obtained based on the survey to match FSW population size estimates obtained via alternative methods.

Stevens 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, in each 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

$$F(x) = \left(1 + \exp(\kappa_c(\mu_c - x))\right)^{-\gamma_c}, \tag{5.19}$$

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 these ratios in South Africa were applicable to every country, allowing calculation of the number of FSW by age group in all 13 countries. The results obtained are shown in Figure 5.2.

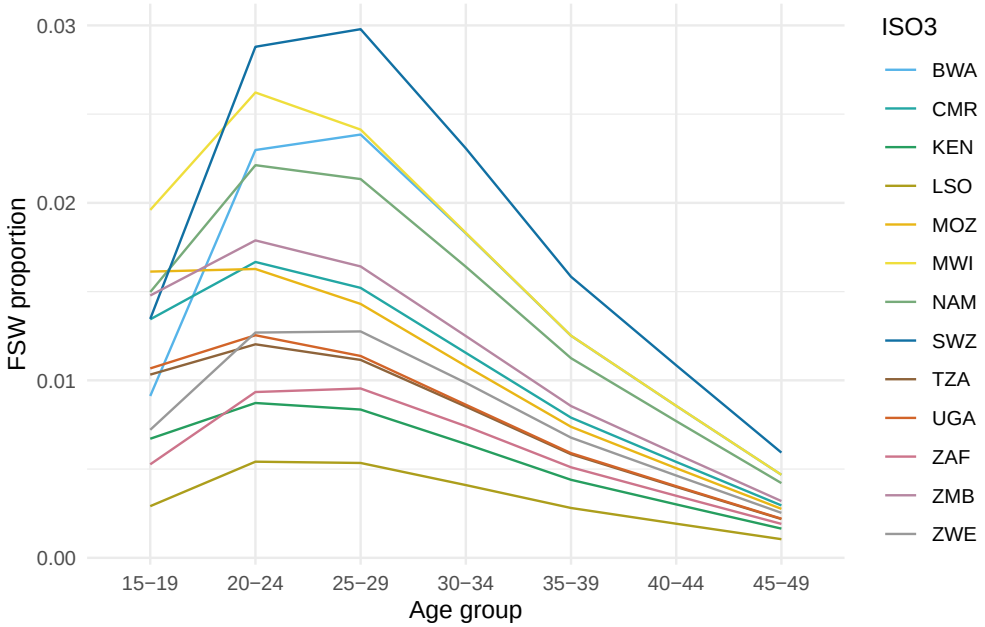

Figure 5.2: The disaggregation procedure I used produces a plausible age distirbution of FSW by country.

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 (**dawid1984present**; **bosse2022scoringutils**), 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 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. **sailynoja2021graphical** develop binomial distribution-based simultaneous confidence bands for ECDF difference values which test uniformity.

Variance decomposition
Estimates

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

$$I_{ia} = \sum_k I_{iak} = \sum_k \lambda_{iak} N_{iak} \tag{5.20}$$

$$= 0 + \lambda_{ia2} N_{ia2} + \lambda_{ia3} N_{ia3} + \lambda_{ia4} N_{ia4} \tag{5.21}$$

$$= \lambda_{ia2} \left(N_{ia2} + \mathrm{RR}_3 N_{ia3} + \mathrm{RR}_4(\lambda_{ia}) N_{ia4} \right). \tag{5.22}$$

Risk group specific HIV incidence estimates are then given by

$$\lambda_{ia1} = 0, \tag{5.23}$$

$$\lambda_{ia2} = I_{ia} / \left(N_{ia2} + \mathrm{RR}_3 N_{ia3} + \mathrm{RR}_4(\lambda_{ia}) N_{ia4} \right), \tag{5.24}$$

$$\lambda_{ia3} = \mathrm{RR}_3 \lambda_{ia2}, \tag{5.25}$$

$$\lambda_{ia4} = \mathrm{RR}_4(\lambda_{ia}) \lambda_{ia2}. \tag{5.26}$$

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 and HIV incidence as such

$$I_a = \sum_{ik} I_{iak}, \tag{5.27}$$

$$\lambda_a = I_a / \sum_{ik} N_{iak}. \tag{5.28}$$

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

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. (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. My work builds on Brugh et al. (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 our 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 our analysis of risk group specific prevalence and incidence).

The final surveys included in our 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

My analysis was subject to challenges shared by most approaches to monitoring sexual behaviour in the general population (**cleland2004monitoring**). 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 (**nnko2004secretive**) or may bias their reporting of sexual debut (**zaba2004age**; **wringe2009comparative**; 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 (**glynn2011assessing**), 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 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 with the aim of facilitating fast and accurate inference for the Naomi small-area estimation model. Naomi is a complex model, used in over 35 countries in sub-Saharan Africa to produce estimates of HIV indicators.

Though I began working on this project in 2020, I only started to make significant progress after reading Stringer et al. (2021). I am grateful to have subsequently collaborate with Alex Stringer on this project, including my visiting the University of Waterloo during the fall term of 2022. 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` and supported by the R package `inf.utils`.

6.1 Background

In a Bayesian analysis, the primary goal is to obtain the posterior distribution $p(\boldsymbol{\phi} \mid \mathbf{y})$. This goal is reasonable because given a loss function, the posterior loss of a decision depends on the data only via the posterior distribution. In this sense, the posterior distribution is sufficient for use in decision making.

It is usually intractable to directly obtain the posterior distribution, because the denominator contains the integral

$$p(\mathbf{y}) = \int_{\mathbb{R}^d} p(\mathbf{y}, \boldsymbol{\phi}) \mathrm{d}\boldsymbol{\phi} \tag{6.1}$$

As such, approximations to the posterior distribution are typically used.

6.2 Inference methods

6.2.1 The Laplace approximation

The posterior normalising constant may be approximated using Laplace's method (Tierney and Kadane 1986). Let $h(\boldsymbol{\phi}) = \log p(\boldsymbol{\phi}, \mathbf{y})$ such that

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

Let

$$\hat{\boldsymbol{\phi}} = \arg\max_{\boldsymbol{\phi}} h(\boldsymbol{\phi}) \tag{6.3}$$

be the posterior mode, and

$$\hat{\mathbf{H}} = -\frac{\partial^2}{\partial \boldsymbol{\phi} \partial \boldsymbol{\phi}^\top} h(\boldsymbol{\phi})|_{\boldsymbol{\phi}=\hat{\boldsymbol{\phi}}} \tag{6.4}$$

be the Hessian matrix evaluated at the posterior mode. Taking a second order Taylor approximation at the posterior mode gives the Laplace approximation to the normalising constant as

$$\tilde{p}_{\mathtt{LA}}(\mathbf{y}) = \int_{\mathbb{R}^d} \exp\left(h(\hat{\boldsymbol{\phi}}) - \frac{1}{2}(\boldsymbol{\phi} - \hat{\boldsymbol{\phi}})^\top \hat{\mathbf{H}} (\boldsymbol{\phi} - \hat{\boldsymbol{\phi}})\right) \mathrm{d}\boldsymbol{\phi} \tag{6.5}$$

$$= \exp(h(\hat{\boldsymbol{\phi}})) \cdot \frac{(2\pi)^{d/2}}{|\hat{\mathbf{H}}|^{1/2}}, \tag{6.6}$$

where Equation 6.5 is calculated using the normalising constant of the Gaussian distribution $\mathcal{N}(\cdot \mid \hat{\boldsymbol{\phi}}, \hat{\mathbf{H}}^{-1})$.

6.2.2 Adaptive Gauss-Hermite quadrature

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

$$\tilde{p}(\mathbf{y}) = \sum_{\boldsymbol{\phi} \in \mathcal{Q}} p(\mathbf{y}, \boldsymbol{\phi}) \omega(\boldsymbol{\phi}). \tag{6.7}$$

6.2.3 Integrated nested Laplace approximation

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 for the log-posterior. The `R-INLA` software uses a formula interface (e.g. `y ~ 1 + x`) which facilitates use of the method for common models. Though beneficial for new users, this imposes constraints on advanced users who would like to use more complicated models. 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`]". Skaug (2009) highlights the potential benefits of a more flexible INLA implementation using automatic differentiation.

I demonstrate the implementation using the epilepsy generalised linear mixed model example from Spiegelhalter et al. (1996). The model is based on that of Breslow and Clayton (1993), itself 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 that there is a difference in approximation error depending on use of either the Gaussian or Laplace approximation for some parameters (Figure 3).

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

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

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 Naomi model

6.5 Malawi case-study

6.6 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 supplement

B

A model for risk group proportions supplement

C

Fast approximate Bayesian inference supplement

Works Cited

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.

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.

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.

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.

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

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

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.

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.

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.

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.

Howes, Adam (2023). *multi.utils: Utility functions for multi-agyw*. R package version 0.1.0.

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*.

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

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

Lohr, Sharon L (2009). *Sampling: Design and Analysis*. Nelson Education.

Lumley, Thomas (2004). "Analysis of Complex Survey Samples". In: *Journal of Statistical Software, Articles* 9.8, pp. 1–19. DOI: 10.18637/jss.v009.i08. URL: https://www.jstatsoft.org/v009/i08.

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

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.

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.

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.

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.

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.

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.

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

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

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*.

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.

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

— (2021b). *Global AIDS Update 2021*. https://www.unaids.org/en/resources/documents/2021/global-aids-update. 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.

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

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.