

# Spatio-temporal estimates of HIV risk group proportions for adolescent girls and young women across 13 priority countries in sub-Saharan Africa

3rd Waterloo Student Conference in Statistics, Actuarial Science and Finance

Adam Howes

Imperial College London

October 2022

# Background

- In sub-Saharan Africa, adolescent girls and young women (AGYW) aged 15-29 are disproportionately at risk of HIV infection
  - 28% of population and 44% of new infections
- This disparity is because of:
  1. Younger age at first sex
  2. Age patterns of sexual mixing
  3. Structural vulnerabilities and power imbalances
  4. Increased susceptibility to HIV infection



...

An adolescent girl or young woman acquired HIV every two minutes in 2021.

Read about the devastating impact HIV has on women and girls in the new Global AIDS Update 2022 [indanger.unaids.org](https://indanger.unaids.org)

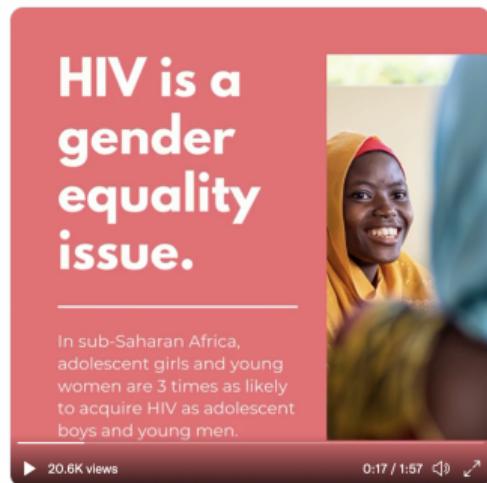


Figure 1: Tweet from UNAIDS

# Prevention packages

- Prevention options can be divided into two:
  1. Core package
  2. Intensified interventions
- Insufficient resources to offer the more costly intensified interventions to all AGYW

⇒ How should interventions be prioritised to have the greatest impact?

# Stratified prevention

- Global AIDS strategy 2021-2026 proposed stratifying HIV prevention for AGYW based on
  1. Population-level HIV incidence
  2. Individual-level sexual risk behaviour
- The two most proximal drivers of sexual transmission



Figure 2: Global AIDS strategy

# Scope for our work

## Goals

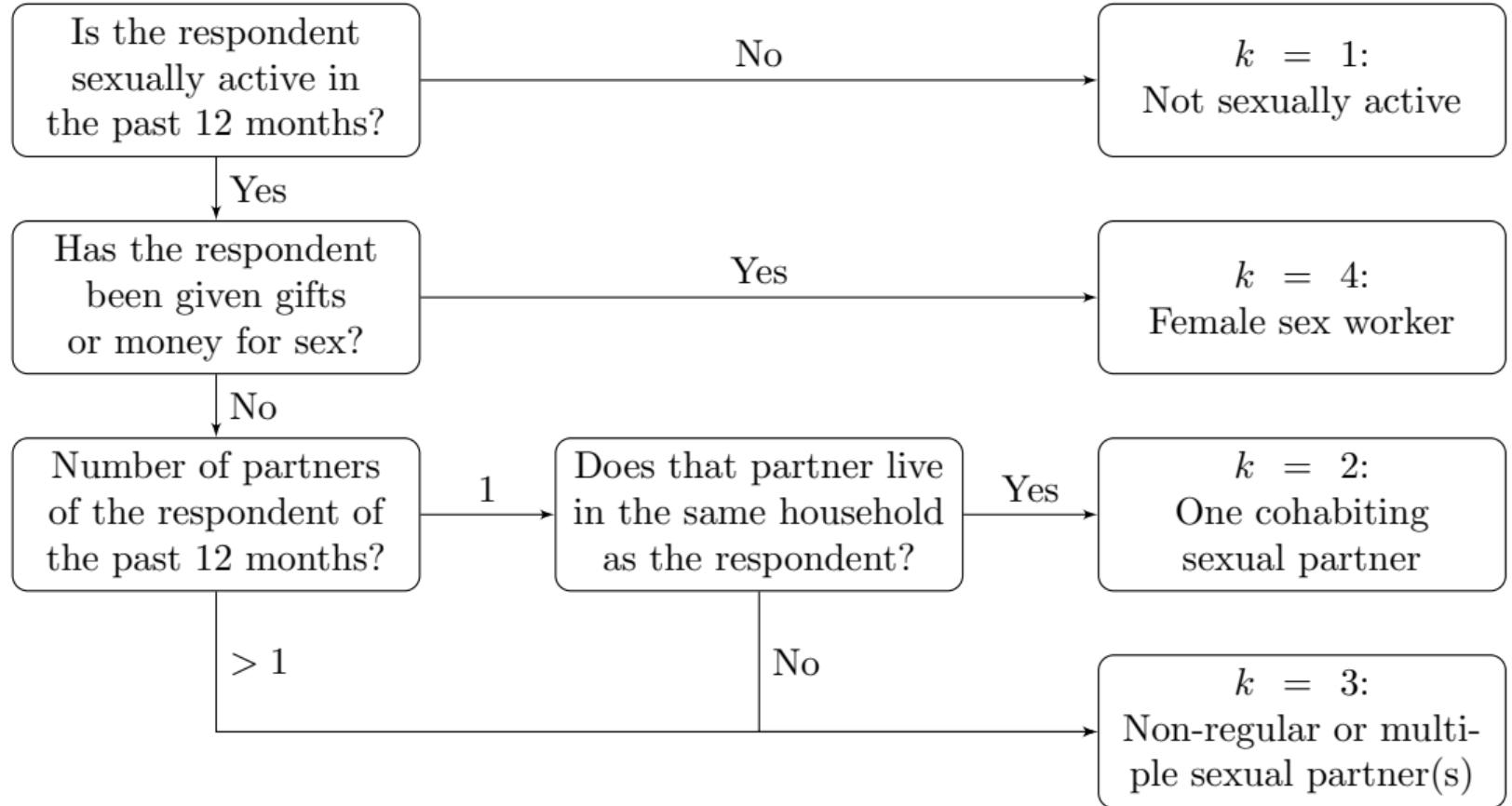
1. Enable implementation of prevention stratified by incidence and behaviour
2. Assess the benefits of such approaches: is it worth it?

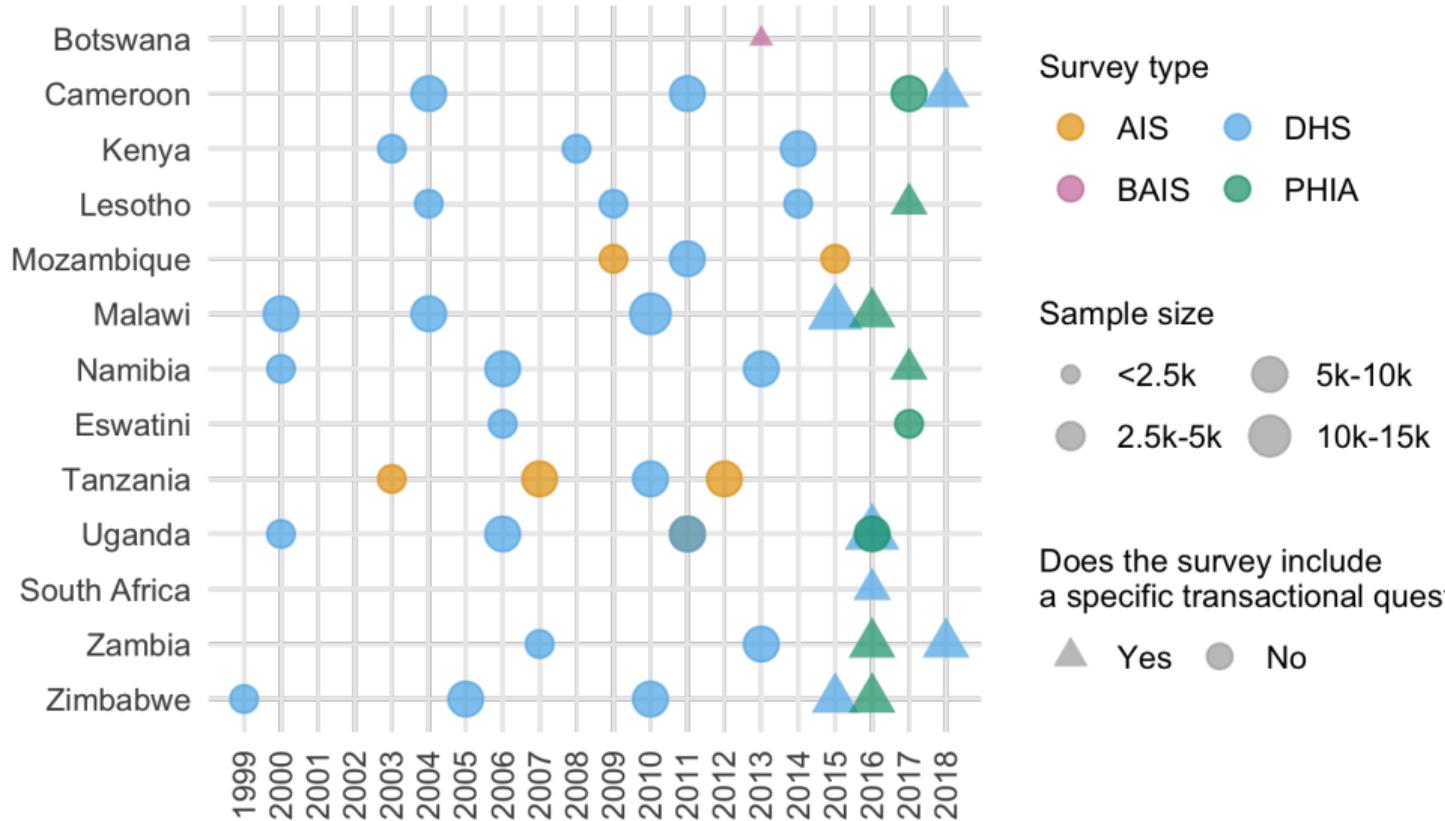
## Approach

1. Estimate the proportion of AGYW in four behavioural risk groups at a district level (in 13 countries identified as priority by The Global Fund)
2. Analyze the new infections which could be reached by different stratified prevention strategies

# Data

- Sexual behaviour data from household surveys to place respondents into  $K = 4$  risk groups:
  1.  $k = 1$  Not sexually active
  2.  $k = 2$  One cohabiting sexual partner
  3.  $k = 3$  Non-regular sexual partner(s)
  4.  $k = 4$  Female sex workers
- District-level HIV incidence, prevalence, population size estimates from the Naomi model (Eaton et al. 2021)
  - Combines survey and programmatic data to estimate indicators in the general population
- Risk ratios from ALPHA network analysis (Slaymaker et al. 2020) and UNAIDS analysis led by Keith Sabin
  - 10 longitudinal studies: <https://alpha.lshtm.ac.uk/>





## Two-stage model for risk group proportions

How do we deal with the fact that only some of the surveys included a transactional sex question? This question is required to differentiate between the  $k = 3$  and  $k = 4$  risk groups

- Our approach was to fit a two-stage model
  1. Spatio-temporal multinomial logistic regression model for the proportion of AGYW in the  $k = 1, 2, 3^+$  risk groups, using all 47 surveys
  2. Spatial logistic regression model for the proportion of those in the  $k = 3^+ = \{3, 4\}$  risk groups who are in the  $k = 4$  risk group, using only the 13 surveys with a specific transactional sex question

## Notation

- $k \in \{1, \dots, 4\}$ : risk groups
- $i \in \{1, \dots, n\}$ : districts
- $c[i] \in \{\text{Botswana}, \dots, \text{Zimbabwe}\}$ : country containing district  $i$
- $t \in \{1999, \dots, 2018\}$ : years
- $a \in \{15-19, 20-24, 25-29\}$ : age groups
- $y_{ita}^* = (y_{ita1}^*, y_{ita2}^*, y_{ita3}^*, y_{ita4}^*)$ : survey weighted multinomial observations
- $m_{ita}^*$ : survey weighted multinomial sample size

## Multinomial logistic regression model

- Would like to use integrated nested Laplace approximations for fast, accurate inference, but R-INLA is not compatible with multinomial likelihoods because they depend on multiple elements of the latent field
- Instead, use that multinomial logistic regression models can be recast as a Poisson log-linear models using the Poisson trick
- This works because conditional on their sum, Poisson counts are jointly multinomially distributed
- If you include observation-specific random effects  $\theta_{ita} \sim \mathcal{N}(0, 1000^2)$  in the linear predictor  $\eta_{itak} = \theta_{ita} + \dots$  then the sample sizes  $m_{ita}^*$  can be exactly recovered, ensuring the models are actually the same

# Multinomial logistic regression model

$$y_{itak}^* \sim \text{Poisson}(\lambda_{itak})$$

$$\log(\lambda_{itak}) = \theta_{ita} + \beta_k + \zeta_{c[i]k} + \alpha_{ac[i]k} + \phi_{ik} + \gamma_{tk} + \delta_{itk}.$$

- The terms are
  - $\theta_{ita}$ : observation (IID)
  - $\beta_k$ : category (IID)
  - $\zeta_{ck}$ : country-category (IID x IID)
  - $\alpha_{ack}$ : age-country-category (IID x IID x IID)
  - $\phi_{ik}$ : space-category ( $\{\text{IID, Besag}\} \times \text{IID}$ )
  - $\gamma_{tk}$ : year-category ( $\{\text{IID, AR1}\} \times \text{IID}$ )
  - $\delta_{itk}$ : space-year-category (Implemented but crashing on cluster at the moment...)
- Without category interactions, any effect will be washed out when we apply softmax to get the probabilities!

Geography



Graph

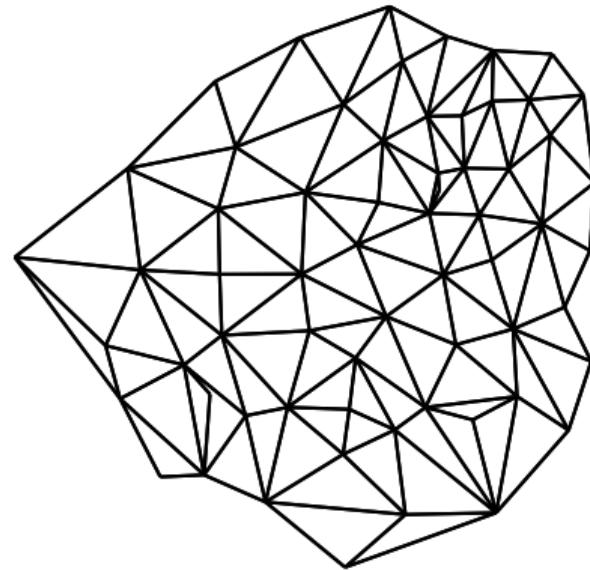


Figure 3: The Besag model,  $\phi_i | \phi_{-i} \sim \mathcal{N} \left( \frac{1}{n_{\delta_i}} \sum_{j:j \sim i} \phi_j, \frac{1}{n_{\delta_i} \tau_{\phi}} \right)$ , operates on the neighbourhood graph.

## Multinomial logistic regression model

- Independent penalised complexity (Simpson et al. 2017) priors on all standard deviation parameters with  $\sigma = 0$  and  $\mathbb{P}(\sigma > 2.5 = 0.01)$ 
  - Sidenote, I'm interested as to if joint priors might be more suitable
- Possible (but tricky) to define all these interactions in R-INLA by combination of the group and replicate options
  - Check out the aaa\_fit\_multi-sexbehav-sae report if you're interested!
- Used sum-to-zero constraints to make posterior inferences interpretable
  - Because we're interested in the contribution of each random effect to total variance
- Model comparison via CPO statistic

# Logistic regression model

$$y_{ia4}^* \sim \text{Binomial}(y_{ia3}^* + y_{ia4}^*, q_{ia}),$$

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

$$\eta_{ia} = \beta_0 + \zeta_{c[i]} + \alpha_{ac[i]} + \phi_i + \beta_{\text{cfsw}} \text{cfsw}_{c[i]}.$$

- The terms are
  - $\beta_0$ : intercept
  - $\zeta_{c[i]}$ : country effects (IID)
  - $\alpha_{ac[i]}$  age-country effects (IID)
  - $\phi_i$ : spatial effects (IID, **Besag**)
  - Clients of FSW covariates (cfswever, cfswrecent) (Hodgins et al. 2022)

## Combination and FSW adjustment

- Take 1000 samples from each model, then manipulate suitably to generate estimates for all four risk groups
- We adjusted the samples from the  $k = 4$  risk group to match age-country FSW estimates, reallocating into non-regular partner(s)
  - Obtained these by disaggregating Stevens et al. (2022) by age using estimates of sexually active population from Nguyen and Eaton (2022)

⇒ Estimates of risk group proportions  $p_{itak}$  by district, year and age group

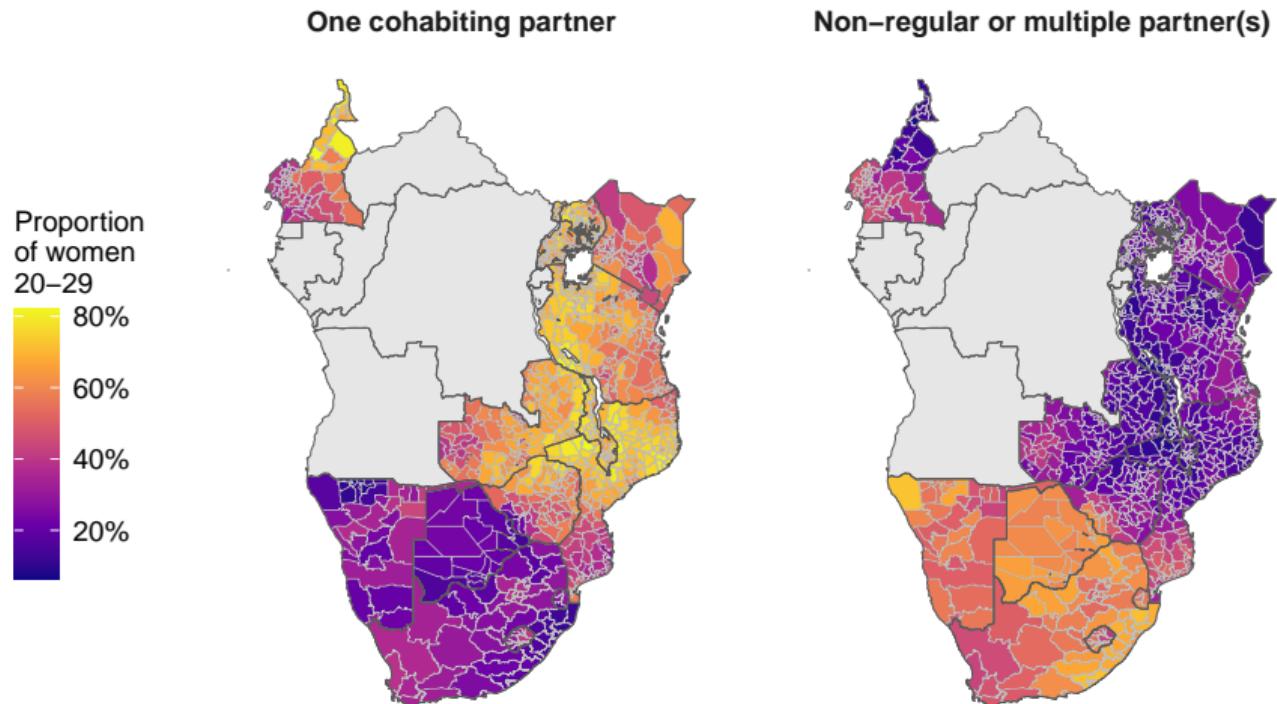
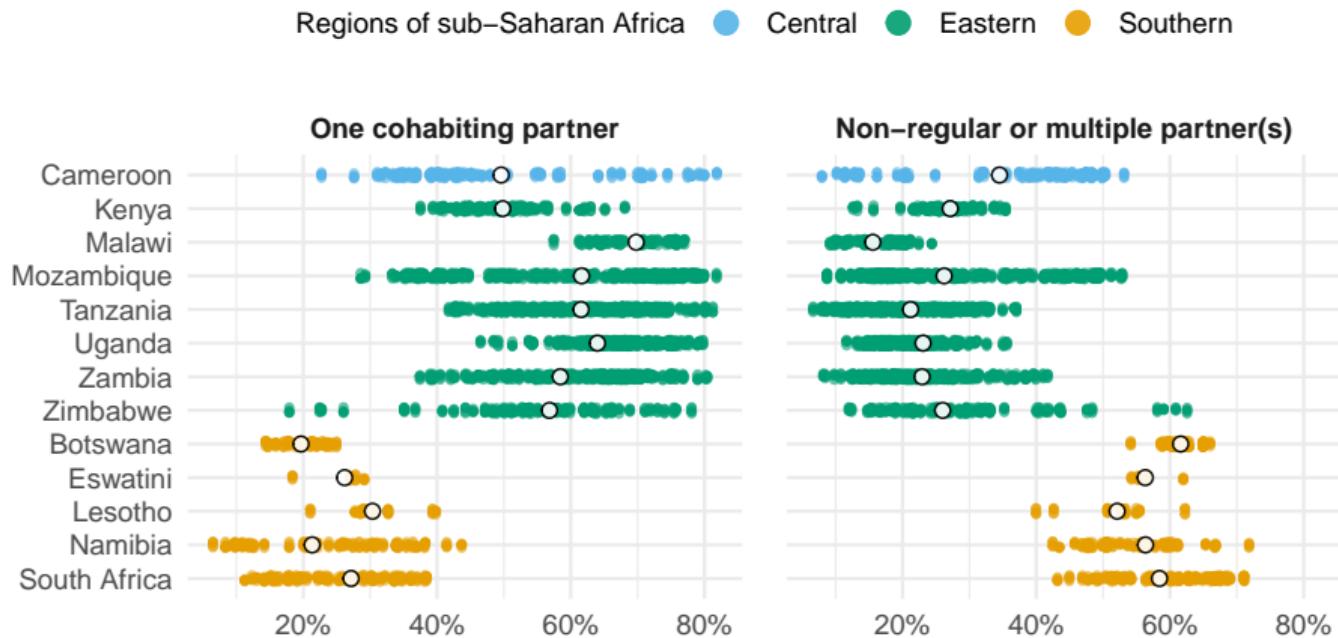


Figure 4: We found a geographic discontinuity in behaviour between Southern and Eastern Africa.



**Not sexually active (not shown) + Cohabiting partner + Nonregular partner(s) + FSW (not shown) = 100%**

Figure 5: Viewing the discontinuity another way.

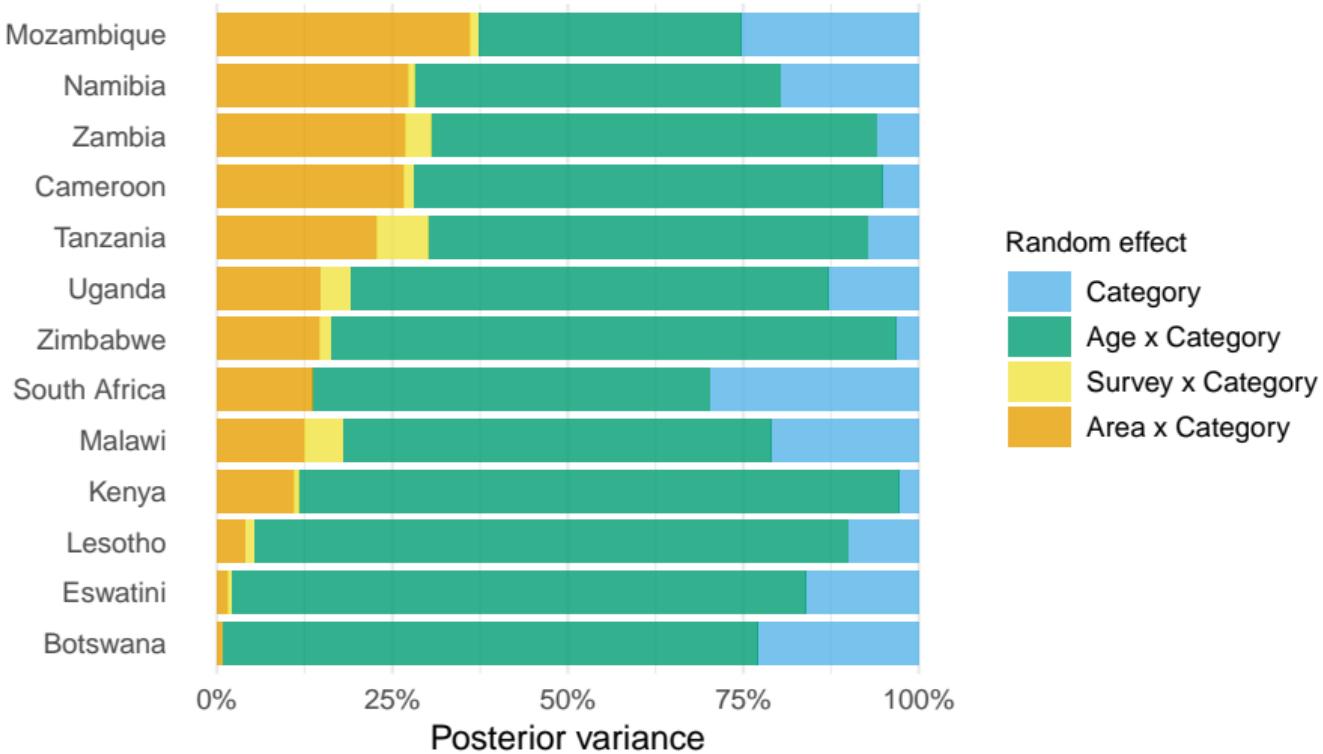


Figure 6: Proportions of variance explained.

# Benefits of our modelled risk group estimates

- Integration of all relevant surveys
  - Two-stage approach allowed estimating FSW proportion even for surveys without a specific transactional sex question
- Alleviating small-sample sizes by borrowing information
  - We borrowed information across space, between countries and over surveys so that our estimates more plausibly reflect spatial heterogeneity
- Estimates where there isn't direct data
  - Although some people think of this as "making up data", the data almost never "speak for themselves" (everything is a model)
  - Uncertainty should be higher in regions with infilling: important to transparently communicate this

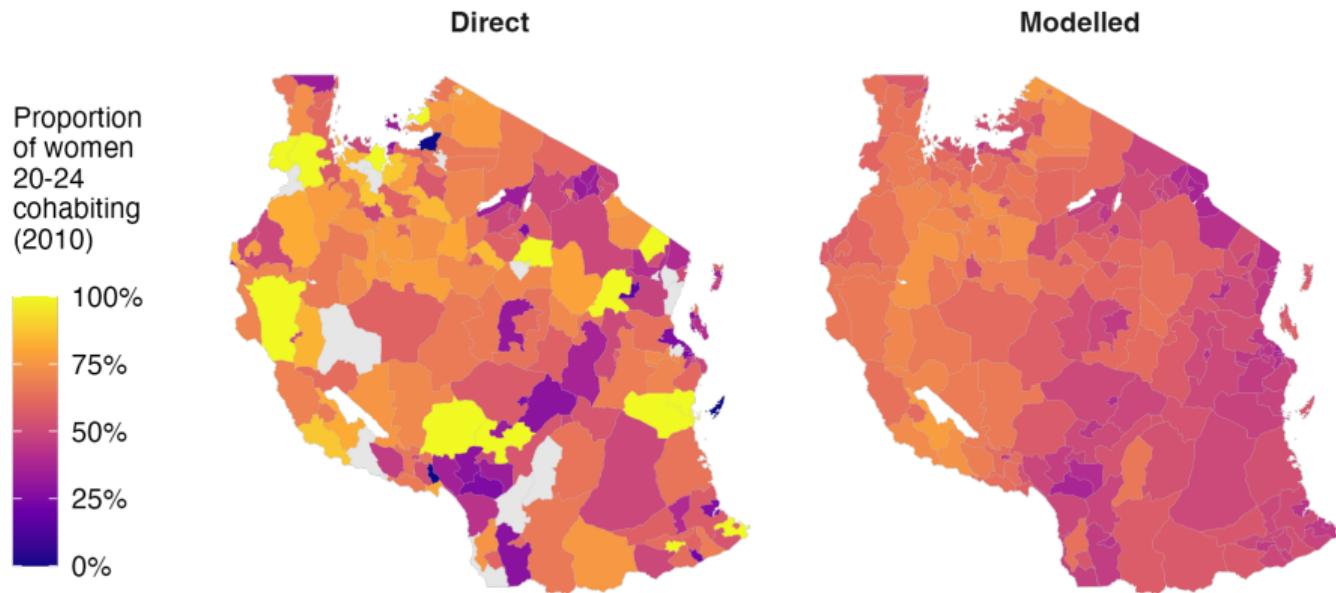


Figure 7: Illustration of the problem with direct survey estimates.

## HIV prevalence and incidence by risk group

- We used our risk group proportion estimates together with incidence relative risk ratios and prevalence ratios to disaggregate general population HIV estimates in the most recent year
- Disaggregated number of new infections on a linear scale, and people living with HIV (PLHIV) on a logit scale
  - Using a linear scale for PLHIV resulted in prevalences outside [0, 1]

⇒ Estimates of HIV incidence  $\lambda_{iak}$ , number of new HIV infections  $I_{iak}$ , HIV prevalence  $\rho_{iak}$  and PLHIV  $H_{iak}$  by district, age group and risk group

## Prioritisation with risk group information

- Suppose we have all of the information (district, age, and risk group)
- Which are the strata with highest incidence?

area_id	age_group	category	population	incidence
ZMB_2_16	Y015_019	sexpaid12m	119.03	0.19
ZAF_2_MAN	Y015_019	sexpaid12m	152.77	0.17
ZAF_2_DC29	Y015_019	sexpaid12m	150.13	0.17
ZAF_2_DC27	Y015_019	sexpaid12m	158.38	0.17
SWZ_1_3	Y015_019	sexpaid12m	262.68	0.16
TZA_4_161rz	Y015_019	sexpaid12m	44.27	0.16

## Prioritisation without risk group information

- What about if we lost the risk group information? Now what are the strata with the highest incidence?

area_id	age_group	population	incidence
SWZ_1_2	Y025_029	8395.92	0.03
MOZ_3_0820	Y020_024	6517.29	0.03
SWZ_1_2	Y020_024	9915.55	0.03
MOZ_3_0803	Y020_024	4278.59	0.03
MOZ_3_0816	Y020_024	11857.78	0.03
SWZ_1_3	Y025_029	17643.13	0.03

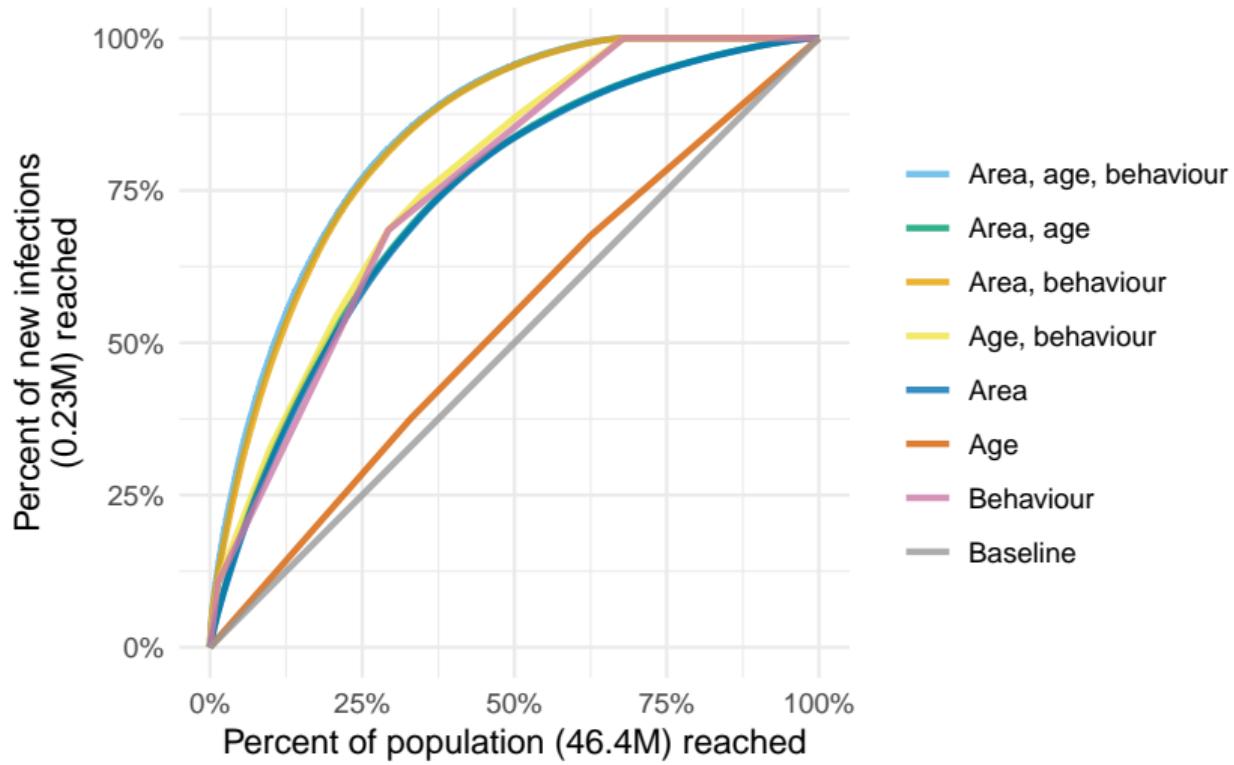


Figure 8: New infections reached prioritising according to different stratifications.

# Limitations

- Simplistic infections reached analysis
  - No accounting for difficulty in reaching each strata
  - Variable intervention effectiveness
  - Change in strata membership
- Under-reporting of high risk sexual behaviours
  - Variation in under-reporting (likely by age, foremost, and location, less so) particularly concerning
- Risk groups definition justification not clear
  - Didn't consider other important characteristics that may determine risk e.g. condom usage
- Only focused on AGYW 15-29
  - Could be extended to adults of both sexes aged 15-49

## Takeaways

1. Risk group estimates can help implement the Global AIDS Strategy; tool and user guide currently being rolled out!
2. Importance of reaching FSW
3. Countries have different epidemic profiles

# Acknowledgements

- Joint work with members of the HIV inference group ([hiv-inference.org](http://hiv-inference.org)) particularly Katie Risher and Jeff Eaton
- The code for this project is at [github.com/athowes/multi-agyw](https://github.com/athowes/multi-agyw)

## About me

- At Waterloo this term to work with Alex Stringer on improving the approximate inference method we use for the Naomi model
- Interested in Bayesian statistics, and applications in global health and biosecurity
- Recently worked on early detection of biological threats with metagenomic sequencing at the Nucleic Acid Observatory
- You can find me online at [athowes.github.io](https://athowes.github.io)
- Come say hi if you have any similar interests!

# Supplement

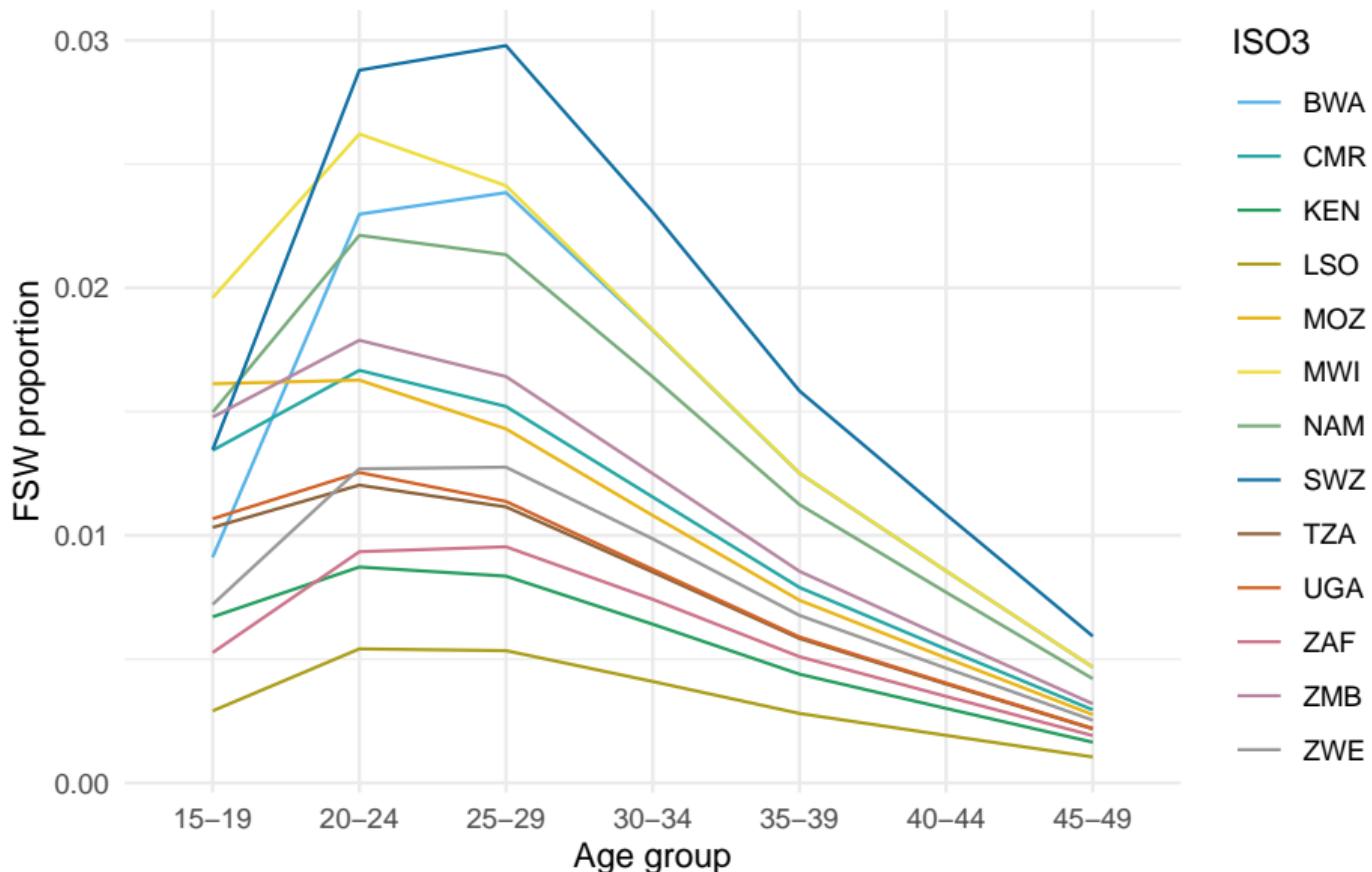


Figure 9: Results of FSW age disaggregation.

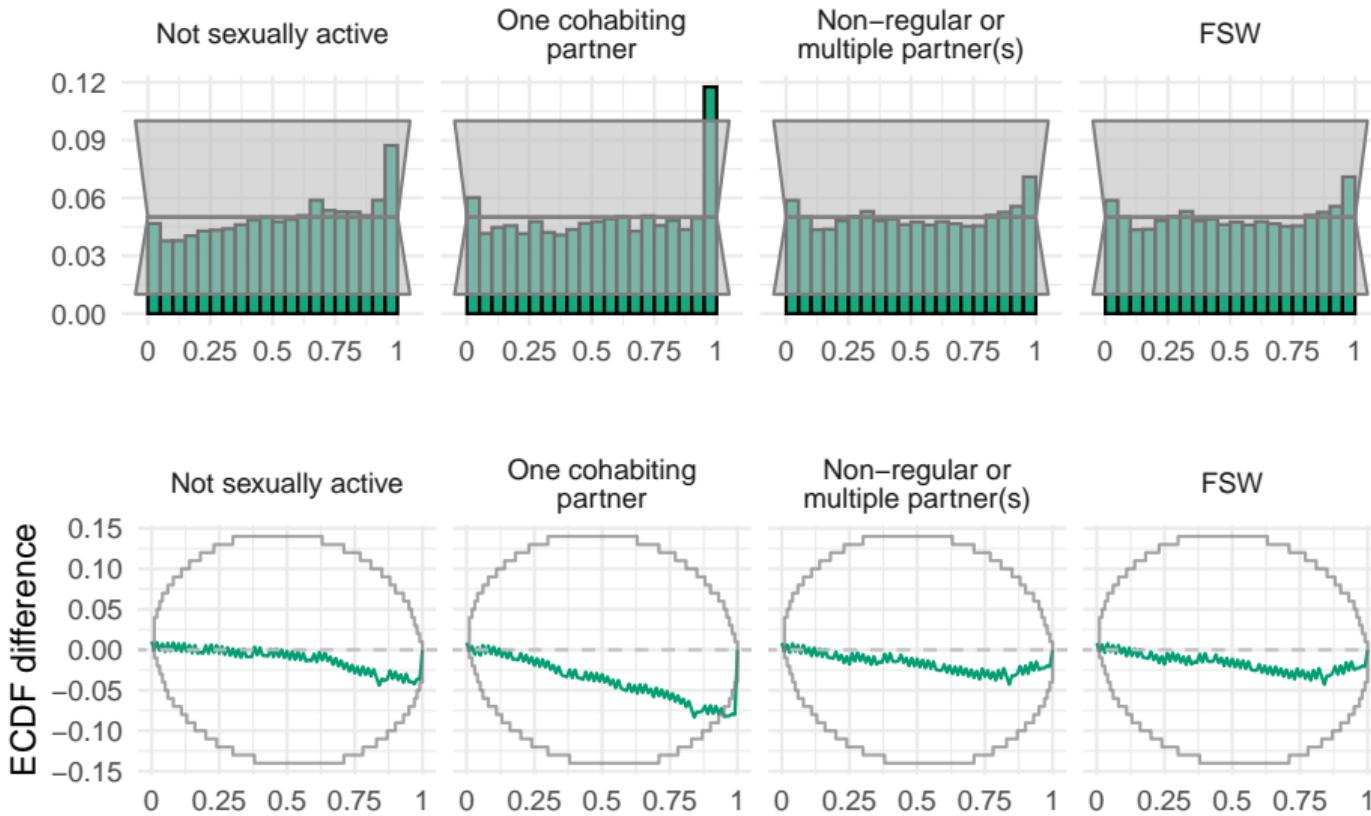


Figure 10: PIT histograms and ECDF difference plots.

# Bibliography I

Eaton, Jeffrey W., Laura Dwyer-Lindgren, Steve Gutreuter, Megan O'Driscoll, Oliver Stevens, Sumali Bajaj, Rob Ashton, et al. 2021. "Naomi: a new modelling tool for estimating HIV epidemic indicators at the district level in sub-Saharan Africa." *Journal of the International AIDS Society* 24 (S5): e25788.

Hodgins, Caroline, James Stannah, Salome Kuchukhidze, Lycias Zembe, Jeffrey W Eaton, Marie-Claude Boily, and Mathieu Maheu-Giroux. 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." *PLoS Medicine* 19 (1): e1003861.

## Bibliography II

- 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." *BMC Public Health* 22 (1): 1120. <https://doi.org/10.1186/s12889-022-13451-y>.
- Simpson, Daniel, Håvard Rue, Andrea Riebler, Thiago G Martins, and Sigrunn H Sørbye. 2017. "Penalising Model Component Complexity: A Principled, Practical Approach to Constructing Priors." *Statistical Science* 32 (1): 1–28.
- Slaymaker, Emma, Kathryn A. Risher, Ramadhani Abdul, Milly Marston, Keith Tomlin, Robert Newton, Anthony Ndyanabo, et al. 2020. "Risk factors for new HIV infections in the general population in sub-Saharan Africa."

## Bibliography III

Stevens, Oliver, Keith Sabin, Sonia Arias Garcia, Kalai Willis, Abu Abdul-Quader, Anne McIntyre, Frances Cowan, et al. 2022. "Estimating key population size, HIV prevalence, and ART coverage for sub-Saharan Africa at the national level."