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

Health Data Science Lab, UWaterloo

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
- 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

Prevention packages

- Prevention options can be divided into two:
  1. Core package
  2. Intensified interventions
- There are not enough resources to offer the more costly intensified interventions to all AGYW, so it's important to prioritise those at highest risk

Stratified prevention

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

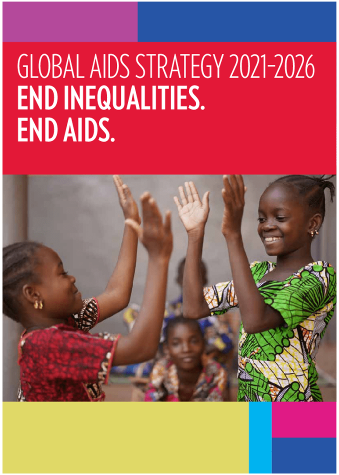


Figure 1: Global AIDS strategy

Scope for our work

Goals

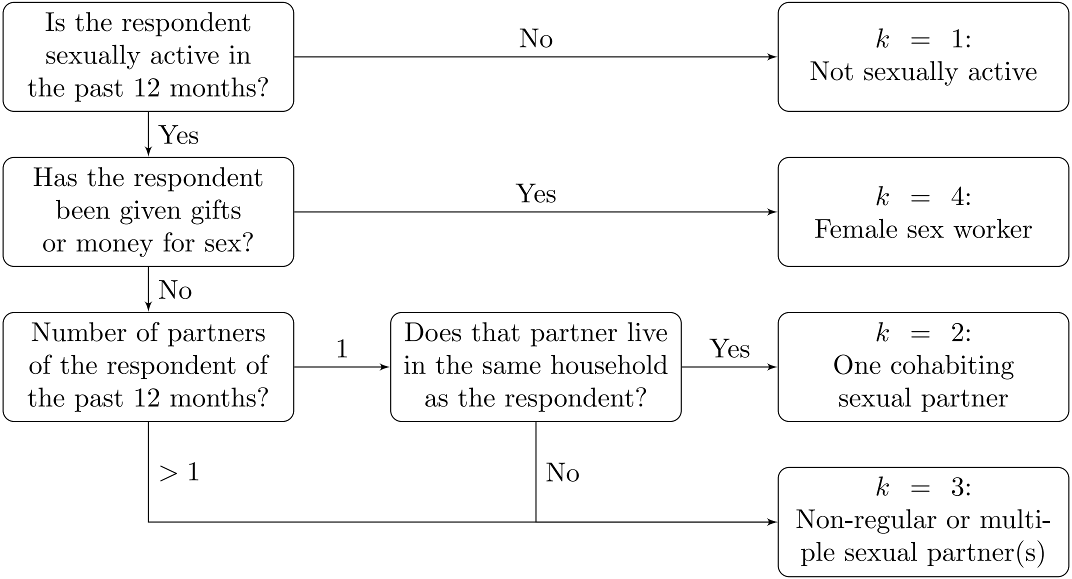
1. Enable implementation of prevention stratified by incidence and behaviour
2. Assess the benefits of such approaches: is it worth it?

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

- We used sexual behaviour data from AIS, BAIS, DHS and PHIA 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)
- Risk ratios from ALPHA network analysis (Slaymaker et al. 2020) and UNAIDS analysis led by Keith Sabin


Is the respondent sexually active in the past 12 months?
No
$k = 1$: Not sexually active
Yes
Has the respondent been given gifts or money for sex?
Yes
$k = 4$: Female sex worker
No
Number of partners of the respondent of the past 12 months?
1
Does that partner live in the same household as the respondent?
Yes
$k = 2$: One cohabiting sexual partner
> 1
No
$k = 3$: Non-regular or multiple sexual partner(s)


Botswana
Cameroon
Kenya
Lesotho
Mozambique
Malawi
Namibia
Eswatini
Tanzania
Uganda
South Africa
Zambia
Zimbabwe
1999
2000
2001
2002
2003
2004
2005
2006
2007
2008
2009
2010
2011
2012
2013
2014
2015
2016
2017
2018
Survey type
AIS
DHS
BAIS
PHIA
Sample size
<2.5k
5k-10k
2.5k-5k
10k-15k
Does the survey include
a specific transactional question?
Yes
No

Two-stage model for risk group proportions

- Only some of the surveys included a transactional sex question, 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
- $a \in \{15\text{-}19, 20\text{-}24, 25\text{-}29\}$: age groups
- $y^\star_{ita} = (y^\star_{ita1}, y^\star_{ita2}, y^\star_{ita3}, y^\star_{ita4})$: survey weighted multinomial observations
- $m^\star_{ita}$: 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
- Include observation-specific random effects $\theta_{ita} \sim \mathcal{N}(0, 1000^2)$ in any linear predictor $\eta_{itak} = \theta_{ita} + \cdots$ to ensure exact reproduction of $m^\star_{ita}$

Multinomial logistic regression model

- Consider models of the form

$$
\begin{aligned}
y^{\star}_{itak} &\sim \text{xPoisson}(\lambda_{itak}) \\
\log(\lambda_{itak}) &= \theta_{ita} + \beta_k + \zeta_{c[i]k} + \alpha_{ac[i]k} + \phi_{ik} + \gamma_{tk} + \delta_{itk}.
\end{aligned}
$$

- The terms are
  - θ_{ita}: observation (IID)
  - β_k: category (IID)
  - ζ_{ck}: country-category (IID x IID)
  - α_{ack}: age-country-category (IID x IID x IID)
  - ϕ_{ik}: space-category ({IID, **Besag**} x IID)
  - γ_{tk}: year-category ({**IID**, AR1} x IID)
  - δ_{itk}: space-year-category (Implemented but crashing on cluster at the moment...)

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
- 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

- Consider models of the form

$$
\begin{aligned}
y^\star_{ia4} &\sim \text{Binomial}\left(y^\star_{ia3} + y^\star_{ia4}, q_{ia}\right), \\
q_{ia} &= \text{logit}^{-1}\left(\eta_{ia}\right), \\
\eta_{ia} &= \beta_0 + \zeta_{c[i]} + \alpha_{ac[i]} + \phi_i + \beta_{\texttt{cfsw}} \texttt{cfsw}_{c[i]}.
\end{aligned}
$$

- The terms are
  - β_0: intercept
  - $\zeta_{c[i]}$: country effects (IID)
  - $\alpha_{ac[i]}$ age-country effects (IID)
  - ϕ_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)

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

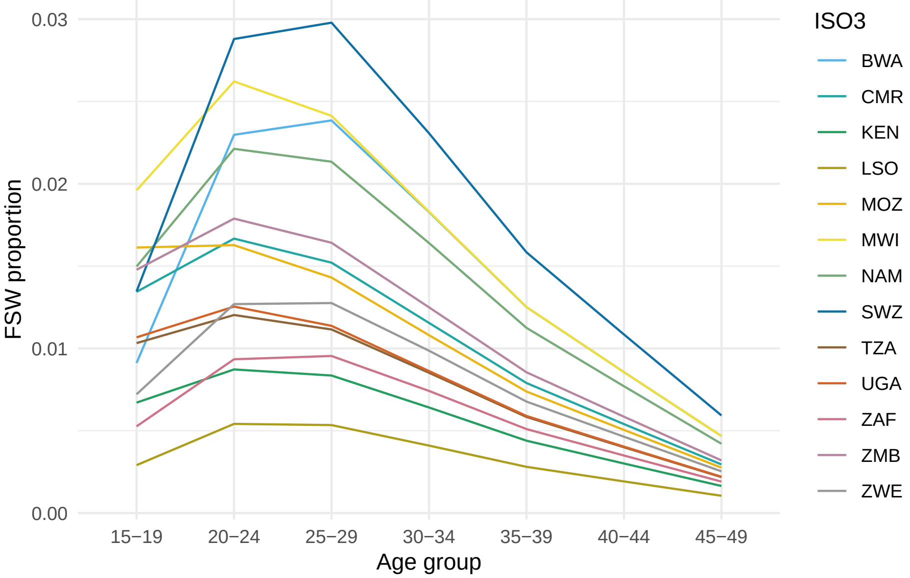


Figure 2: Results of FSW age disaggregation.

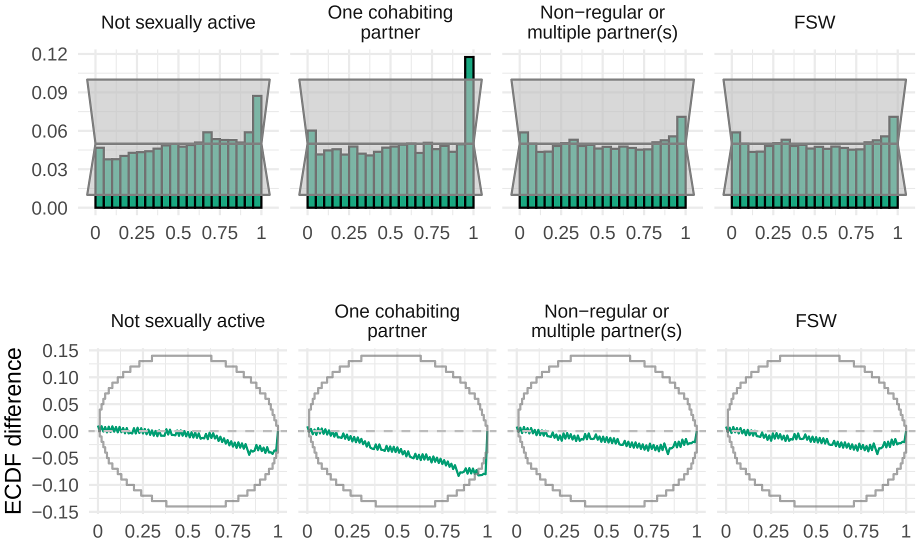


Figure 3: PIT histograms and ECDF difference plots.

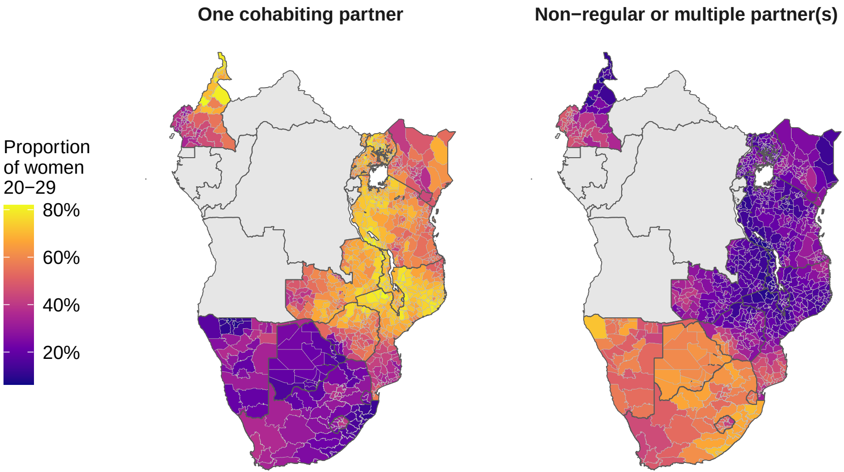


Figure 4: We found a geographic discontinuity in behaviour between Southern and Eastern Africa.

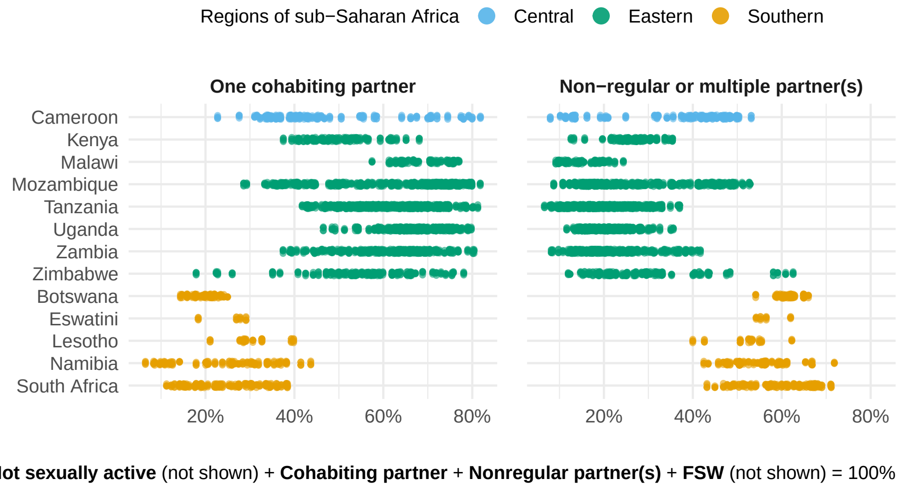


Figure 5: Viewing the discontinuity another way.

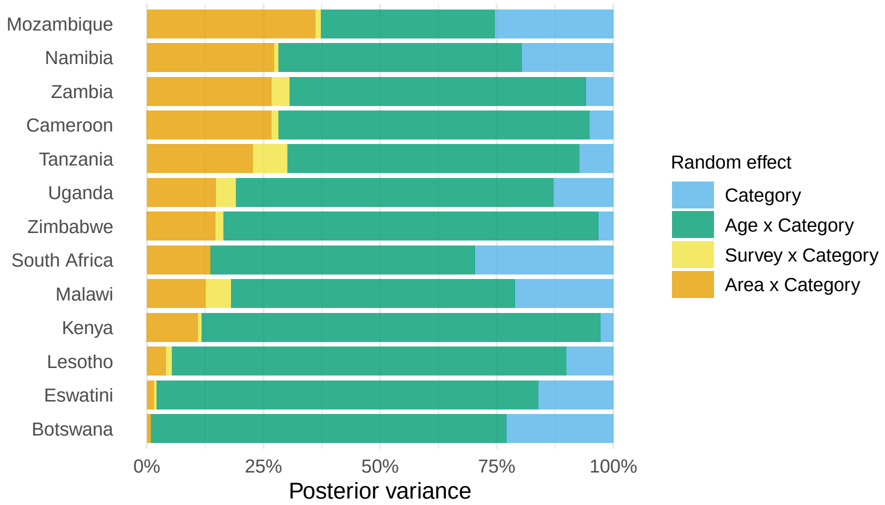


Figure 6: Proportions of variance explained.

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
- Alleivating 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

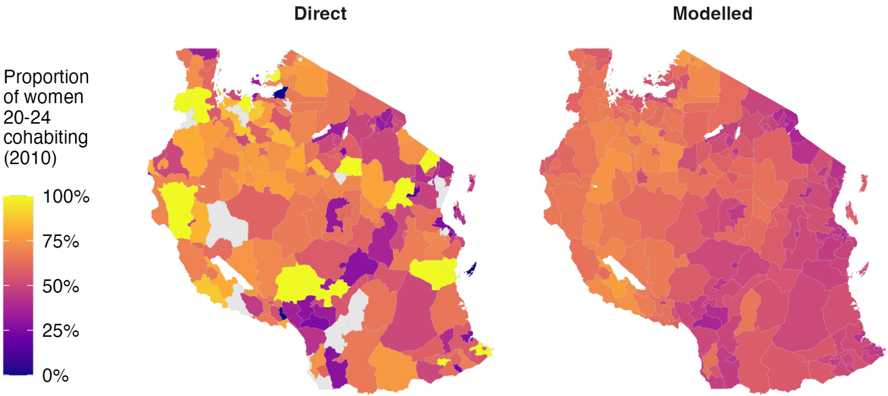


Figure 7: Illustration of the problem with direct survey estimates.

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 $[0, 1]$

$\implies$ Estimates of HIV incidence λ_{iak}, number of new HIV infections I_{iak}, HIV prevalence ρ_{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

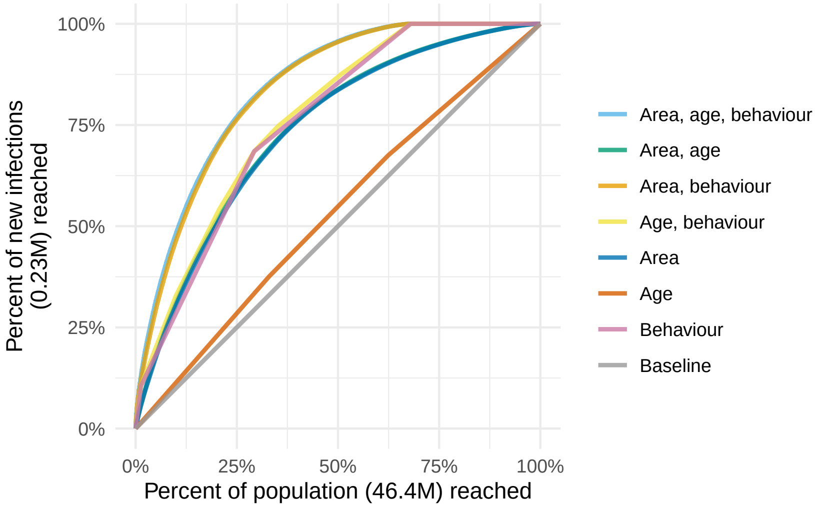


Figure 8: New infections reached prioritising according to different stratifications.

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

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

Thanks for listening!

- Joint work with members of the HIV inference group (hiv-inference.org) particularly Katie Risher and Jeff Eaton
- The code for this project is at github.com/athowes/multi-agyw
- You can find me online at athowes.github.io

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."