Made available through Montana State University's ScholarWorks

# Mapping the number of female sex workers in countries across sub-Saharan Africa

Ian Laga, Xiaoyue Niu, Katherine Rucinski, Stefan Baral, Amrita Rao, David Chen, Nikita Viswasam, Nancy Refilwe Phaswana-Mafuya, Daouda Diouf, Keith Sabin, Jinkou Zhao, Jeffrey W. Eaton, Le Bao

© 2023 National Academy of Sciences.

# Main Manuscript for

## Mapping the number of female sex workers in countries across sub-Saharan Africa

Ian Laga[1+], Xiaoyue Niu[2+], Katherine Rucinski[3], Stefan Baral[3], Amrita Rao[3], David Chen[2], Nikita Viswasam[3], Nancy Refilwe Phaswana-Mafuya[4], Daouda Diouf[5], Keith Sabin[6], Jinkou Zhao[7], Jeffrey W. Eaton[8], Le Bao[2*].

[1]Montana State University. [2]Pennsylvania State University. [3]John Hopkins University. [4]Health Sciences at the University of Johannesburg, South Africa. [5]Enda Santé, Senegal. [6]UNAIDS. [7]The Global Fund to Fight AIDS, Tuberculosis and Malaria. [8]MRC Centre for Global Infectious Disease Analysis, School of Public Health, Imperial College London.

[+]These authors contributed equally to the project.

**Corresponding Author information:** *Le Bao, Department of Statistics, Penn State University, University Park, PA 16802

**Email:** lebao@psu.edu

**Author Contributions:** L.B. and X.N. conceived and planned the study. K. S., N. V., and D.C. conducted literature search and extracted the FSW raw data. D.C. extracted and processed nighttime light data. I.L. extracted and processed the remaining covariates. I.L. carried out the statistical analysis with assistance and input from L.B. and X.N.. S.B., K.S., J.E., K.S., J. Z., M.P., and D.D. contributed to the scope of the analysis and discussion points. I.L. prepared figures and tables. I.L. wrote the first draft of the manuscript. J.E., S.B., K.S., J.Z, M.P., and D.D. provided comments and feedback of the drafts. L.B., X.N., and K. R. contributed to the revisions. I.L. and X.N. finalized the manuscript.

**Competing Interest Statement:** We declare no competing interests

**Preprint Servers:** SSRN: https://dx.doi.org/10.2139/ssrn.3919697

**Classification:** Major: Social, Social Science. Minor: Physical, Statistics.

**Keywords:** Key populations, Size estimates, Bayesian hierarchical model,

**This PDF file includes:**

Main Text
Materials and Methods
References
Figures 1 to 2
Tables 1 to 2

## Abstract

Female sex workers (FSW) are affected by individual, network, and structural risks, making them vulnerable to poor health and wellbeing. HIV prevention strategies and local community-based programs can rely on estimates of the number of FSW to plan and implement differentiated HIV prevention and treatment services. However, there are limited systematic assessments of the number of FSW in countries across sub-Saharan Africa to facilitate identification of prevention and treatment gaps. Here we provide estimated population sizes of FSW and the corresponding uncertainties for almost all sub-national areas in sub-Saharan Africa. We first performed a literature review of FSW size estimates and then developed a Bayesian hierarchical model to synthesize these size estimates, resolving competing size estimates in the same area and producing estimates in areas without any data. We estimated that there are 2.5 million (95% uncertainty interval 1.9-3.1) FSW aged 15-49 in sub-Saharan Africa. This represents a proportion as percent of all women of childbearing age of 1.1% (95% uncertainty interval 0.8%-1.3%). The analyses further revealed substantial differences between the proportions of FSW among adult females at the sub-national level and studied the relationship between these heterogeneities and many predictors. Ultimately, achieving the vision of no new HIV infections by 2030 necessitates dramatic improvements in our delivery of evidence-based services for sex workers across sub-Saharan Africa.

## Significance Statement

In 2021, the proportion of sub-national areas in sub-Saharan Africa with FSW size estimates remains low, challenging evidence-based resource allocation for health and HIV services. We produced model-based FSW population size estimates in all areas across sub-Saharan Africa. Ultimately, achieving the vision of no new HIV infections by 2030 necessitates dramatic improvements in our delivery of evidence-based services for sex workers across sub-Saharan Africa.

## Main Text

## Introduction

Female sex workers (FSW) are at increased risk of negative health outcomes and often face stigma and discrimination from their communities, clients, partners, and families. Female sex workers are more likely to acquire and experience ongoing transmission of HIV than other reproductive aged women (1, 2). Global median HIV prevalence among FSW is 3.9% (3, 4), while FSW in sub-Saharan Africa are at much higher risk for HIV with about one in five sex workers estimated to be living with HIV. In addition, FSW have an increased risk of syphilis and herpes simplex virus type 2 (HSV-2) (5). As noted by Strathdee and colleagues, “sex—whether paid for or not—does not cause HIV infection”(6). Rather, higher prevalence of HIV and other STIs is due to increased exposure to infection coupled with limited access to essential sexual and reproductive health services including HIV prevention and treatment. Recent literature calls attention to FSW’s reproductive health and desire to be mothers (7, 8) need to provide financial support to their family and friends (6, 9), increasing number of human rights violations against sex workers (10, 11), higher vulnerability to homicide (10), sex trafficking (12, 13) and the prevalence of sexually exploited children and adolescents (14, 15). The issues FSW face extend beyond HIV, and it is important to better understand FSW populations to address these problems.

Epidemiologic, health, and social risks among FSW have been well documented across sub-Saharan Africa, but the scarcity of comparable data characterizing FSW within and across countries challenges an empiric and equitable HIV response (16). Given the limitedness of robust surveillance systems, rigorous methodologies have been developed to estimate the size of FSW and other hidden populations, including census and enumeration, capture-recapture, multiplier, network scale-up, and geographic mapping methods (16). However, few real-world examples of these rigorous studies exist (17, 18). Additionally, multiple size estimates for FSW within and across settings, including estimates derived from different approaches, make it challenging to directly compare estimates and to fully understand the true variation in the size of FSW across countries in sub-Saharan Africa. Generating consistent FSW size estimates can help inform equitable responses for national HIV programs to effectively reduce new infections (17, 19–21). These size estimates can also guide other programs such as those focused on protecting FSW from exposure to unsafe working conditions, increase condom usage, and improve reproductive health for FSW.

## Results

### Overview

While there have been multiple studies and recent systematic reviews to estimate HIV prevalence across Africa (22), including among key populations (23–26), there is no published comprehensive effort to estimate population size of sex workers in sub-Saharan Africa at the national or sub-national level. One study presented FSW size estimates for a subset of African countries and additional countries around the world, primarily at the national level (26). We searched for existing FSW population-size estimates in sub-Saharan Africa, geo-located these data, and used a Bayesian hierarchical model to meta-analyze the proportion of adult women who are FSW across sub-Saharan Africa, accounting for different study methodologies.

We modelled the proportions of women of reproductive age, aged 15-49 years who are FSW for all mainland sub-Saharan African countries and Madagascar for all sub-national areas (typically the largest geographic units that divide a country, additional details regarding the spatial resolution are found in the Methods section) in the year 2015. Note that sex workers are only recognized as aged 18 and above by the United Nation (27). We present estimates for ages 15-49 because this is one of the most common age ranges considered in our literature review and includes FSW of reproductive age. Additionally, we provide national-level estimates by aggregating the sub-national estimates within each country. Our estimates are based on a Bayesian hierarchical model to leverage all existing estimates in the literature and address the heterogeneity in the existing FSW estimates by controlling for the effects that country and estimation method have on the size estimates and including auxiliary spatial covariates that explain much of the variability. While the FSW population sizes are of main interest, we modelled the proportions to better handle the varying population sizes in sub-national areas and cities. Furthermore, we present the results primarily as FSW proportions to directly compare estimates both within and between countries.

### Female sex worker sizes

We performed a literature review and meta-analysis of FSW size estimates in 44 countries in sub-Saharan Africa (Supplementary Table S3). For this review, we considered all definitions of FSW that involve the exchange of sex services for money or anything of value. We identified 1253 unique FSW population size estimates (PSE) across 38 countries in sub-Saharan Africa at various spatial resolutions between 2000 and 2018. Existing FSW size estimates were not uniformly distributed across countries in sub-Saharan Africa. While 38 of the 44 sub-Saharan African countries in our meta-analysis had at least one FSW size estimate, more than 50% of the 1253 unique FSW PSE belonged to only four countries (Ghana, Malawi, Nigeria, and Uganda). There were fewer than five FSW size estimates in 8 of the 38 countries. We classified observations as national (related to the entire country), sub-national (related to an administrative division, either administrative-one or -two level), city, or city area. After stratifying sub-Saharan Africa into sub-national areas for prediction and modeling, we removed PSE if they corresponded to multiple sub-national areas (127 instances; 100 corresponded to regions in Ghana, 14 to states in Nigeria, and 13 to the 8 former provinces in Kenya before the provinces were subdivided into 46 districts and Nairobi), related to a sub-city area (58 instances, all of which corresponded to cities with other city-level estimates), or missed key information like the name of the city (55 instances). We also excluded 81 national estimates since many represent scaled-up versions of the sub-national estimates and including the national estimates adversely affected out-of-sample model performance. Finally, we removed one unreliable observation that estimated the FSW proportion of Nairobi, Kenya to be larger than 60% of the number of women aged 15-49 and more than 10 times larger than the next highest FSW proportion estimate of the same area. All other estimates, including some with FSW proportions larger than 60%, were kept due to lack of evidence they were wrong. After removing these PSE, we were left with 932 observations, composed of 616 sub-national area estimates and 316 city level estimates across 33 countries. PSE availability at the city and sub-national level and the FSW percents of these observations from 2000 to 2018 are shown in Figure 1. Data availability was generally sparse and varied widely across countries. Very few countries had both sub-national area and city estimates or estimates across multiple years, making longitudinal study of FSW size difficult, if not impossible. We chose 2015 as the representative year for our estimates since most literature estimates were available around 2015. The dataset and analytic files are available in the supplementary information.

The FSW size estimates in sub-Saharan Africa were of varying quality and exhibited large heterogeneity both across different and within the same geographic areas. Within our sub-national dataset, the proportion of women who were FSW varied widely, ranging from 0.1% to 70%. Estimates were often significantly different for the same sub-national area within a country. For example, one FSW size estimate for Mwanza, Tanzania was over three times larger than another. The literature estimates also varied widely with respect to the direct size estimation methods. We grouped the many different reported methods into eight categories (capture-recapture, enumeration, expert opinion, mapping, multiple methods, multiplier, miscellaneous, and not reported). Eighteen countries contained sub-national area or city estimates from more than one method. Zambia reported the largest variety of estimates, containing 47 area or city estimates from six different methods. Conversely, Nigeria contained 96 estimates, all associated with a mapping method. Systematically combining existing FSW size estimates and extrapolating to areas with no estimates must handle the large variation in the data and accurately reflect the uncertainty of the resulting estimates.

Uncertainty estimates of the existing FSW size estimates should vary significantly depending on the sample size and method used in the literature. Unfortunately, very few of the existing size estimates were reported with uncertainty intervals (less than 13%). Ideally, any model-based estimator would utilize the uncertainty of the data to produce more accurate estimates and uncertainty intervals. In our case, this was not possible, but we emphasize that future size estimation studies should report uncertainty intervals in addition to the size estimates, and models based on these estimates should incorporate the uncertainty intervals into the model.

We developed a Bayesian hierarchical random effects model to combine the existing studies and FSW PSE. Specifically, we assumed that the logit FSW proportion followed a Student's t distribution. Heterogeneity was captured using auxiliary variables, independent and identically distributed gaussian random effects for the estimation method effect, and independent and identically distributed gaussian random effects for the country effect. Reference population sizes were calculated using citypopulation.de and WorldPop data to convert FSW PSE to FSW proportion with respect to women aged 15-49 (Supplementary Information). In the case where the reference population size could not be calculated, the reference population size from the literature was used if available. We do not use the reference population from the literature by default because many studies use populations other than women of reproductive age as the reference population.

We conducted all analysis using R (versions 3.6.3–4.0.2) (28).

## Modelled FSW size estimates

We fitted a Bayesian linear hierarchical model with spatially varied economic, social, and environmental predictors and effects for country and estimation method. We present here the estimated sub-national FSW proportions and uncertainties (Figure 2a, 2b) and national FSW proportions and uncertainties (Figure 2c, 2d). Overall, we estimated there are 2.5 million (95% uncertainty interval 1.9-3.1) FSW aged 15-49 in sub-Saharan Africa, corresponding to a proportion as percent of all women of childbearing age of 1.1% (95% uncertainty interval 0.8%-1.3%). FSW proportions varied widely at both the national and sub-national levels. The estimated national FSW proportions ranged from 0.3% in Malawi to 2.8% in Burundi. Within countries, the range of FSW proportions were also substantial. For example, in Nigeria, the proportion estimates ranged from 1.1% in Jakusko to 2.3% in Ogori-Magongo.

For each country, we examined the contribution from the predictors and the country effects by decomposing the final estimates into the predictor effects and the country effects, reported in Supplementary Table S5. This decomposition is useful for better understanding the components that contribute to a country's estimate. For example, the low estimated FSW proportion in Malawi was primarily due to the country effect. The estimated FSW proportion in Malawi from only the predictors was 0.9% (the 30th smallest among all countries), but the relatively large amount of literature estimates (103 PSE for 27 districts) with small FSW proportion resulted in a large negative country-level effect and consequently a small FSW proportion of 0.3%.

## FSW size predictors

To handle poorly behaved predictors (e.g., those that exhibited extreme distributions with little relationship to the response), we screened many covariates based on visual diagnostics of their distribution and their relationship with the response variable. We then performed Bayesian variable selection via a horseshoe prior (Supplementary Information) (29, 30). The penalized regression shrunk non-significant coefficients towards zero and handled covariates with high collinearity. Our final model contained 69 predictors and an interaction term. Note that all predictors remain in the model for Bayesian penalized regression. We present all estimated coefficients, covariate explanations, and credible intervals in Supplementary Table S2 and Supplementary Figure S10. In Table 1 here, we present the estimated coefficients and corresponding 95% credible intervals for the ten largest estimated coefficients. These nine covariates (excluding intercept) in the order of absolute value of coefficient estimate are indicator for city, the interaction between city and log number of women of reproductive age, log number of women of reproductive age, population density, percent urban, G6PD (glucose-6-phosphate dehydrogenase) deficiency allele frequency, population weighted walking time to nearest healthcare, population weighted sickle haemoglobin allele frequency, and isothermality. For the continuous predictors, the estimated coefficients indicate how much the logit-transformed FSW prevalence changes for one standard deviation increase of the predictor. For interpretation, Figure S11 includes a plot which shows how one standard deviation increase of the above predictors affect the estimated FSW percent.

Fixing all other covariates, for sub-national areas, as the number of women of reproductive age increases, the FSW proportion decreases. In cities, FSW proportion decreases even faster, so larger cities tend to have lower FSW proportion than smaller cities. For an average population size, city estimates are higher than sub-national estimates. Additionally, sub-national areas with higher urban density, higher population density, higher G6PD allele frequency, and higher sickle haemoglobin allele frequency have higher FSW proportions. Conversely, areas where people must walk farther to the nearest healthcare facility and areas with higher isothermality have lower FSW proportions.

Uncertainty in estimates was large for sub-national areas in countries with few or no existing FSW estimates. The relationship between data availability and estimate uncertainty is shown in Supplementary Figure S2. Estimates in countries with few data are drawn towards the average sub-Saharan Africa FSW proportion. While the predictors we included in the model improved predictions, there remains a large country-level effect that is unidentified for countries with no usable estimates. Including even a few estimates (as in Sudan and Republic of the Congo, for example) can drastically improve estimates for the remaining sub-national areas in the country. It is unclear, however, whether the significant differences in FSW proportions between countries reflects the actual proportions in the countries or whether the FSW PSE produced in the studies differ dramatically in implementation, obscuring the true FSW proportions.

### Size estimation methods

The measurement methods used for direct size estimation were not strongly associated with the observed population size. All eight 95% credible intervals included zero (full parameter estimates are shown in Table 2). These parameter estimates do not correspond to any methodological biases and are only observations about how different methods compared to each other. One explanation for the lack of significant difference between the estimation methods in our study is that there was not sufficient data within each country to accurately detect systematic variation due to the estimation method. In the literature, countries overwhelmingly used only one or two estimation methods. Notable exceptions included Malawi (29 estimates from capture recapture, 20 from enumeration, and 56 from multiple methods) and Zambia (9 from enumeration, 8 from expert opinion, 13 from multiple methods, 4 from multiplier, 12 miscellaneous methods, and 1 not reported). The observed FSW proportions for each method varied considerably, suggesting that even after accounting for variability using the auxiliary spatial covariates, the uncertainty associated with the estimation method is too large to quantify significant differences in method. There were likely significant differences in the implementation of the estimation methods, obscuring any systematic differences in the methods.

## Discussion

This is the first study to present FSW size estimates inclusive of all countries in sub-Saharan Africa. Through an extensive synthesis of existing FSW size estimates from the literature and the use of geospatial auxiliary covariates, we were able to comprehensively generate FSW size estimate proportions for both national and sub-national areas of the entire sub-Saharan African region. Our findings demonstrate substantial variation in estimates within and across countries. Resulting estimates also had limited precision in settings where direct size estimates were largely absent, affirming the importance of continued primary data collection for FSW population size estimates moving forward.

Consistent with all model-based methods, estimates are reliant on the quality of the data. In our case, this is not only the FSW size estimates from the literature, but also our reference population size calculations as an underwhelming percent of studies identified through the literature reported either the reference population or the corresponding FSW proportion. While our generated reference populations are largely consistent with the more reliable reference populations where available in the literature (Supplementary Figure S12), in general the true reference population is unknown, making it difficult to verify the populations for all estimates. We posit that our generated reference populations are accurate, and any error introduced is small relative to the variability in the size estimates and the modeling, but we emphasize the importance of reporting reference population sizes alongside FSW size estimates. Along these lines, it is also beneficial to report the definition of FSW used in the study, the year the data were collected, the catchment area, what administrative areas are represented, and whether extrapolation has been used for each size estimate. Additionally, many existing estimates do not include quantitative uncertainty assessments making it difficult to distinguish between reliable estimates based on rich information and ambiguous guestimates. To improve future studies that rely on the synthesis of existing estimates, it is also helpful to report the uncertainty in the form of variance or uncertainty intervals. The above recommendations are also applicable to documenting the size estimates of other hard-to-reach populations.

While this study provides the first FSW proportion estimates in areas with no previous estimates, the handling of competing estimates in the same area is equally important. Repeated studies for the same sub-national area typically produce widely different FSW size estimates, and our model provides a way to combine these estimates in a straight-forward and sensible approach. A key assumption of our model is that averaging across all size estimation methods, FSW PSE are unbiased. Given that no single estimation method clearly outperforms the other methods in practice, this assumption is desirable. If a method emerges as the gold standard, our model can be easily modified to center around that method. We believe that overall, our model-based estimates provide reliable FSW estimates, and all estimates and measures of uncertainty are available as an interactive map at https://github.com/ilaga/Mapping-FSW-SSA/blob/main/Figures/FSW_Map_Interactive.zip. In many cases the uncertainty is relatively large, and any future analysis which utilizes our FSW size estimates should incorporate the uncertainty and not rely solely on the point estimates.

Our study is also a call to action for more accurate and extensive estimates of FSW. Importantly, model-based methods that rely on existing FSW size estimates are a complement rather than a substitute for primary data collection and direct size estimation studies. Interventions implemented specifically for FSW and clients have a larger impact on HIV prevention than equivalent resources directed more broadly, including towards low-risk activity groups (31). UNAIDS emphasizes that the strategy to end AIDS as a public health threat by 2030 relies on mitigating inequities, both among and within countries (11). In order to meet new 95-95-95 targets, UNAIDS highlighted that countries need to have updated size estimates of all key populations, including FSW, to allow HIV prevention programs and implementation partners to allocate the correct amount of resources for the true population sizes (11). It is not surprising that the sub-national areas and countries with more FSW data correspond to the estimates with the smallest uncertainty. By increasing both the quality and quantity of FSW size estimation studies, not only does resource allocation become easier and more efficient, but these studies shed light on FSW in neighboring areas.

WHO guidelines on combination HIV prevention programs suggest the need for comprehensive and well specified strategies for sex workers as a means of improving HIV outcomes and quality of life. Historically, the majority of services have focused on individual risks such as condom use (6). While necessary, these programs are not sufficient in terms of overcoming the significant risks for both the acquisition and transmission of STIs and reproductive health services for FSW. Addressing structural risks with community empowerment approaches where sex workers collectively control the HIV programs in their community have been shown to reduce HIV and STI prevalence and increase condom usage (32). Sex work is work, but it is work done often in dangerous conditions, and affected by intersecting stigmas and criminalization of sex workers challenging the reporting of physical and sexual violence (10). In 2020, the world came nowhere close to achieving the goal of 500,000 new HIV infections. Since then, the emergence of COVID-19 has further affected the ability to deliver services. Moving forward necessitates a dramatic shift in how we address the needs of FSW to decrease disease incidence and improve their quality of life. Advocating for services to achieve this and address the structural risks faced by sex workers necessitates improved data on how many sex workers there are and where they live. Our study provides the most comprehensive answer to these questions for all sub-national areas in sub-Saharan Africa available to date.

## Materials and Methods

### Overview.

This analysis estimates female sex worker (FSW) proportion among women of reproductive age (aged 15-49 years) at the sub-national level for most sub-Saharan African countries. The final estimates represent the FSW proportion in 2015 since most literature estimates were available around 2015.

**Population size estimate data.**

We compiled a dataset of 1253 unique estimates across 38 countries in sub-Saharan Africa from public resources. The reference population was women of reproductive age (aged 15-49). Reference population sizes were calculated using citypopulation.de and WorldPop data (Supplementary Information). In the case where the reference population size could not be calculated, the reference population size from the literature was used if available. We did not use the reference population from the literature by default because many studies used populations other than women of reproductive age as the reference population. Information extracted from the literature included country, spatial region, type of spatial region (e.g., district, region, city, etc.), estimation method (which we later categorized into eight categories: capture-recapture, enumeration, expert opinion, mapping, multiple methods, multiplier, miscellaneous, and not reported) PSE, FSW proportion, publication year, and data year.

**Spatial Resolution.**

Model fitting and proportion estimates used GADM 3.6 administrative-one level units (33) for all countries except Ghana, Uganda, and Nigeria. For these three countries, literature estimates were available in almost all administrative-two level units. Both Ghana and Uganda added many new districts from the creation of the GADM 3.6 shapefile, so shapefiles better matching the literature estimates were used from the government of Ghana (34) and the government of Uganda (35) statistical agencies' websites. The GADM 3.6 administrative-two level units in Nigeria matched with the literature estimates, so these were used. For simplicity, we refer to these administrative units as sub-national areas. Because our analysis includes cities and areas of varying size, our covariates are limited to values standardized by area (e.g., population density is acceptable while total population is not). However, error is introduced both at the response and the covariate level. In the absence of the original spatial units used to estimate the population proportions, the alignment to the sub-national areas was necessary and seemed to be reasonable, and the error introduced was likely small relative to the variation from both the literature estimates and the model.

Literature estimates were classified as either city, sub-national, national, or miscellaneous. For modeling purposes, estimates were classified as a city if they represented a geographic area smaller than the sub-national area and miscellaneous if they represented a geographic area bigger than the sub-national area or corresponded to multiple sub-national areas. Thus, estimates were labeled a sub-national area in cases where estimates corresponded to both a city and a sub-national area (e.g., Brazzaville, Republic of Congo).

**Covariates.**

Our primary analysis included auxiliary spatial covariates (Supplementary Table S6). The auxiliary spatial covariates were all available at a grid cell resolution across sub-Saharan Africa, and the average value for each covariate was calculated for each sub-national area, weighted by WorldPop in each area when relevant. When yearly data was available, estimates were matched with the covariate in their respective years or to the nearest year where data was available. In most cases, only one year was available, so the variable was processed identically regardless of the size estimate year. Many of the auxiliary spatial covariates were highly skewed with heavy upper tails. Variables were plotted against the logit of the observed FSW proportions to evaluate whether a log transformation resulted in a better linear relationship. These transforms are specified in Table 1 and Supplementary Table S2. Additionally, the detailed estimation methods from the literature were mapped to eight general methods and used as a random effect group in the statistical model.

In a secondary analysis, we considered additional covariates from Demographic and Health Surveys (DHS) (36). The DHS data contains information at household clusters across a subset of the SSA countries and includes thousands of unique survey questions which were narrowed down to several hundred using pairwise correlations with the logit-transformed proportions. We considered only DHS surveys with GPS coordinates of cluster locations. DHS surveys were not available in all administrative units. We attempted to impute DHS values for missing sub-national areas using both multiple imputation chained equations (MICE) and via a conditionally autoregressive spatial model (CAR). In both cases, the imputed values were not consistent with observed values, and thus we chose to omit DHS variables from our model. We performed a study to compare estimates with and without the DHS variables. Using only data with available DHS observations, we fit and compared three models: 1. the “combined model” which included both DHS and auxiliary spatial covariates, 2. the “auxiliary model” which included only the auxiliary spatial covariates, and 3. the “DHS model” with included only the DHS covariates. These models were compared using leave-one-out cross-validation. The auxiliary model performed marginally better but was similar in performance to the combined model. The DHS model performed the worst. In addition, the fitted and predicted values from the combined model and auxiliary model exhibited no systematic differences and were very similar. The standard errors of the estimates were essentially identical. Based on these results, we concluded that imputing the missing DHS values would introduce more error than the DHS values would explain, and thus we did not lose predictive accuracy by omitting the DHS covariates. Additional discussion and diagnostics regarding this decision are shown in the Supplementary Information.

### Statistical Model.

We modelled the number of FSW in sub-national area $i$ and replicate $j$ ($Y_{i,j}$) using a Bayesian hierarchical model to allow sharing in FSW estimates between literature and within a country. We excluded national estimates and sub-city estimates since many national estimates were simply the summation of the sub-national area estimates and all sub-city estimates had a matching city estimate which was the sum of sub-city FSW sizes. Including the national and sub-city estimates would lead to smaller than expected uncertainty intervals. We assumed the $Y_{i,j}$ comes from the following hierarchical model:

$$\text{logit}\left(\frac{Y_{i,j}}{N_{i,j}}\right) = \beta_0 + \boldsymbol{\beta_1}\boldsymbol{X}_{i,j} + \alpha_{\text{method}[i,j]} + \gamma_{c[i]} + \epsilon_{\text{i,j}}$$

$$\epsilon_{\text{i,j}} \sim \text{St}(0, \sigma, \nu),$$

where ~ denotes “distributed as”, St denotes the Student’s t distribution, and all $\epsilon_{\text{i,j}}$ are independent. In the above, the logit-transformed FSW proportion in sub-national area $i$ and replicate $j$ came from a Student’s t random variable with location (mean) $\beta_0 + \boldsymbol{\beta_1}\boldsymbol{X}_{i,j} + \alpha_{\text{method}[i,j]} + \gamma_{c[i]}$, scale $\sigma$, and degrees of freedom $\nu$. The number of women of child-bearing age in area $i$ corresponding to replicate $j$ is denoted by $N_{i,j}$. We assumed a linear predictor on the mean, where $\beta_0$ is an overall intercept, $\boldsymbol{\beta_1}\boldsymbol{X}_{i,t}$ is the covariate effect corresponding to the set of predictors $\boldsymbol{X}_{i,t}$ in sub-national area $i$ and replicate $j$, and $\alpha_{\text{method}[i,j]}$ was a gaussian random effect to account for the estimation method associated with the literature PSE. Country-level gaussian random effects centered around zero were included in $\gamma_{c[i]}$, where $c[i]$ represents the country for area $i$. While area $i$ is nested inside $c[i]$, the areas were uniquely coded. The mean includes the index $j$ to note that logit-transformed FSW proportion in sub-national area $i$ was measured repeatedly across years, by different estimation methods, and possibly from a city estimate.

FSW estimates from the fitted model for sub-national area $i$ in year 2015 ($\hat{Y}_i$) were obtained using standard Bayesian methodology with posterior samples, excluding $\alpha_{\text{method}[i,j]}$ to ignore any method effect, using estimated $\gamma_{c[i]}$ for countries with available FSW size estimates, and simulating new $\gamma_{c[i]}$ from a normal distribution with estimated country-level variance. Specifically, for sub-national area $i$, we made inference about $\hat{\beta}_0 + \hat{\boldsymbol{\beta}}_{\mathbf{1}}\boldsymbol{X}_{i,j} + \hat{\gamma}_{c[i]}$, where $\boldsymbol{X}_{i,j}$ is the covariate matrix corresponding to year 2015.

The model was fitted using the RStan (37) and brms (38, 39) packages. Diagnostic plots are found in the Supplementary Information.

### Data availability

The authors declare that the literature female sex worker meta-analysis data supporting the findings of this study are available in the supplementary information and at https://github.com/ilaga/Mapping-FSW-SSA. Auxiliary data are subject to third party restrictions, but links to all data used for the current study are provided in the supplementary information and can be accessed after successful registration.

**Code availability**
All code (in addition to generated figures and results) is available at https://github.com/ilaga/Mapping-FSW-SSA. The nighttime light processing code is available at https://github.com/TheDavidChen/NL_Africa.

**Acknowledgments**
This research was supported by NIH/NIAID R01AI136664. The funders of the study had no role in the data collection, data analysis, data interpretation, or writing of the report.

**Additional information**
Supplementary information is available for this paper.

## References

1. C. E. Lyons, *et al.*, The role of sex work laws and stigmas in increasing HIV risks among sex workers. *Nat Commun* **11**, 1–10 (2020).

2. M. Moazzami, *et al.*, Characterizing multi-level determinants of HIV prevalence among female sex workers in Maseru and Maputsoe, Lesotho. *AIDS Behav* **24**, 714–723 (2020).

3. Joint United Nations Programme on HIV/AIDS, The Gap Report (2014).

4. Joint United Nations Programme on HIV/AIDS, AIDS by the numbers. *Geneva: Joint United Nations Programme on HIV/AIDS (UNAIDS)* (2016).

5. L. Vu, K. Misra, High burden of HIV, syphilis and HSV-2 and factors associated with HIV infection among female sex Workers in Tanzania: implications for early treatment of HIV and pre-exposure prophylaxis (PrEP). *AIDS Behav* **22**, 1113–1121 (2018).

6. S. A. Strathdee, A.-L. Crago, J. Butler, L.-G. Bekker, C. Beyrer, Dispelling myths about sex workers and HIV. *The Lancet* **385**, 4–7 (2015).

7. P. Duff, *et al.*, Pregnancy intentions among female sex workers: recognising their rights and wants as mothers. *Journal of Family Planning and Reproductive Health Care* **41**, 102–108 (2015).

8. S. W. Beckham, *et al.*, Female sex workers' experiences with intended pregnancy and antenatal care services in southern Tanzania. *Stud Fam Plann* **46**, 55–71 (2015).

9. K. K. Hoang, “Dealing in desire” in *Dealing in Desire*, (University of California Press, 2015).

10. M. R. Decker, *et al.*, Human rights violations against sex workers: burden and effect on HIV. *The Lancet* **385**, 186–199 (2015).

11. Joint United Nations Programme on HIV/AIDS, End Inequalities. End Aids. Global Aids Strategy 2021–2026 (2021).

12. R. Steen, S. Jana, S. Reza-Paul, M. Richter, Trafficking, sex work, and HIV: efforts to resolve conflicts. *The Lancet* **385**, 94–96 (2015).

13. E. J. Schauer, E. M. Wheaton, Sex trafficking into the United States: A literature review. *Crim Justice Rev* **31**, 146–169 (2006).

14. C. McClure, C. Chandler, S. Bissell, Responses to HIV in sexually exploited children or adolescents who sell sex. *The Lancet* **385**, 97–99 (2015).

15. R. Curtis, *et al.*, The commercial sexual exploitation of children in New York City. *New York: Center for Court Innovation* (2008).

16. A. S. Abdul-Quader, A. L. Baughman, W. Hladik, Estimating the size of key populations: current status and future possibilities. *Curr Opin HIV AIDS* **9**, 107 (2014).

17. S. Baral, *et al.*, Burden of HIV among female sex workers in low-income and middle-income countries: a systematic review and meta-analysis. *Lancet Infect Dis* **12**, 538–549 (2012).

18. J. MacAllister, *et al.*, A comprehensive review of available epidemiologic and HIV service data for female sex workers, men who have sex with men, and people who inject drugs in select West and Central African countries. *JAIDS Journal of Acquired Immune Deficiency Syndromes* **68**, S83–S90 (2015).

19. N. Viswasam, *et al.*, The uptake of population size estimation studies for key populations in guiding HIV responses on the African continent. *PLoS One* **15**, e0228634 (2020).

20. F. Tanser, T. de Oliveira, M. Maheu-Giroux, T. Bärnighausen, Concentrated HIV sub-epidemics in generalized epidemic settings. *Curr Opin HIV AIDS* **9**, 115 (2014).

21. C. Beyrer, *et al.*, A call to action for concentrated HIV epidemics. *Curr Opin HIV AIDS* **9**, 95 (2014).

22. L. Dwyer-Lindgren, *et al.*, Mapping HIV prevalence in sub-Saharan Africa between 2000 and 2017. *Nature* **570**, 189–193 (2019).

23. K. B. Rucinski, *et al.*, High HIV prevalence and low HIV-service engagement among young women who sell sex: a pooled analysis across 9 Sub-Saharan African countries. *JAIDS Journal of Acquired Immune Deficiency Syndromes* **85**, 148–155 (2020).

24. C. Avoka, P. Adzordor, V. Bawontuo, D. A. Akila, D. Kuupiel, Mapping evidence on HIV status awareness among key and vulnerable populations in sub-Sahara Africa: a scoping review protocol. *Syst Rev* **9**, 1–6 (2020).

25. G. Djomand, S. Quaye, P. S. Sullivan, HIV epidemic among key populations in west Africa. *Curr Opin HIV AIDS* **9**, 506 (2014).

26. E. Papworth, *et al.*, Epidemiology of HIV among female sex workers, their clients, men who have sex with men and people who inject drugs in West and Central Africa. *J Int AIDS Soc* **16**, 18751 (2013).

27. Joint United National Programme on HIV/AIDS, Guidance Note on HIV and Sex Work (2009).

28. R Core Team, R: A language and environment for statistical computing (2013).

29. C. M. Carvalho, N. G. Polson, J. G. Scott, Handling sparsity via the horseshoe in *Artificial Intelligence and Statistics*, (PMLR, 2009), pp. 73–80.

30. C. M. Carvalho, N. G. Polson, J. G. Scott, Handling sparsity via the horseshoe in *Artificial Intelligence and Statistics*, (PMLR, 2009), pp. 73–80.

31. A. Rao, *et al.*, HIV-related data among key populations to inform evidence-based responses: protocol of a systematic review. *Syst Rev* **7**, 1–7 (2018).

32. D. Kerrigan, *et al.*, A community empowerment approach to the HIV response among sex workers: effectiveness, challenges, and considerations for implementation and scale-up. *The Lancet* **385**, 172–185 (2015).

33. Global Administrative Areas, GADM database of Global Administrative Areas. *version 3.6*.

34. Ghana Statistical Service, Shapefiles of all Districts in Ghana 2010 (216 districts) (2012).

35. Uganda Bureau of Statistics with support from WHO, Uganda - Subnational Administrative Boundaries (2020).

36. ICF, Demographic and Health Surveys (various) [Datasets]. *Funded by USAIDS*.

37. Stan Development Team, RStan: the R interface to Stan. *R package version* **2**, 522 (2016).

38. P.-C. Bürkner, brms: An R package for Bayesian multilevel models using Stan. *J Stat Softw* **80**, 1–28 (2017).

39. P. C. Bürkner, Advanced Bayesian multilevel modeling with the R package brms. The R Journal, 10 (1), 395-411 (2018).

## Figures and Tables

**Figure 1. Data availability and FSW Percent in sub-Saharan Africa. a-c**, Number of usable city (a) and sub-national area (b) FSW PSE estimates available from the literature in each area across all years. Sub-national areas in sub-Saharan Africa with no estimates are shown in light grey and countries not in sub-Saharan Africa are shown in dark grey with stripes. In (c), the percent of FSW in each country with respect to the number of women aged 15-49 are plotted against the country ISO-3 codes (dots are jittered for readability). Sub-national estimates are shown in orange and the city estimates are shown in light blue.

**Figure 2. Estimated FSW percent of women aged 15-49 in 2015 at the country and sub-national area level and the relative uncertainty of the area estimates. a-d**, Estimated FSW percent at the sub-national area level (a) and at the country level (c). We do not produce estimates for countries not in sub-Saharan Africa (dark grey with stripes). The relative uncertainty of FSW percent in each sub-national area is shown in (b) and each country in (d). Relative uncertainty is defined by the length of the 95% credible interval divided by the posterior mean. Lighter colors correspond to lower FSW percent and uncertainty, and darker colors correspond to higher FSW percent and uncertainty.

**Table 1.** Predictor coefficient summaries for centered and scaled predictors

| Predictor | Estimate (95% credible interval) |
|---:|---:|
| Intercept | -4.402 (-4.784, -3.992) |
| City | 0.564 (0.000, 0.947) |
| Reference Population (log) * City | -0.299 (-0.472, -0.057) |
| Reference Population (log) | -0.249 (-0.521, -0.022) |
| Population Density | 0.106 (-0.014, 0.448) |
| Percent Urban | 0.060 (-0.007, 0.188) |
| G6PD Freq | 0.038 (-0.009, 0.152) |
| Walking time to Healthcare PW | -0.035 (-0.263, 0.021) |
| HBS Freq PW | 0.025 (-0.030, 0.267) |
| Isothermality | -0.010 (-0.099, 0.024) |

**Table 2.** Method Coefficient Summaries. The baseline FSW percent estimate corresponding to no method used at the average covariate value is 1.23%

| Method | Estimate (95% credible interval) | FSW percent estimate at methods |
|---:|---:|---:|
| Capture Recapture | 0.041 (-0.180, 0.304) | 1.29 |
| Enumeration | 0.209 (-0.013, 0.495) | 1.52 |
| Expert Opinion | -0.212 (-0.599, 0.067) | 1.01 |
| Mapping | -0.084 (-0.385, 0.188) | 1.14 |
| Misc | -0.113 (-0.406, 0.142) | 1.11 |
| Multiple | 0.032 (-0.182, 0.286) | 1.27 |
| Multiplier | -0.051 (-0.301, 0.201) | 1.18 |
| NR | 0.197 (-0.050, 0.528) | 1.50 |