HHS Public Access
Author manuscript
Ann Epidemiol. Author manuscript; available in PMC 2022 March 01.

Published in final edited form as:
Ann Epidemiol. 2021 March ; 55: 34–40. doi:10.1016/j.annepidem.2020.12.001.

Using Factor Analyses to Estimate the Number of Female Sex Workers across Malawi from Multiple Regional Sources

Xiaoyue Maggie Niu[1,*], **Amrita Rao**[2,*], **David Chen**[1], **Ben Sheng**[1], **Sharon Weir**[3], **Eric Umar**[4], **Gift Trapence**[5], **Vincent Jumbe**[4], **Dunker Kamba**[5], **Katherine Rucinski**[2], **Nikita Viswasam**[2], **Stefan Baral**[2], **Le Bao**[1]

[1]Department of Statistics, Eberly College of Science, Pennsylvania State University

[2]Center for Public Health and Human Rights, Department of Epidemiology, Johns Hopkins Bloomberg School of Public Health

[3]Department of Epidemiology, University of North Carolina, Chapel Hill

[4]Department of Health Systems and Policy, School of Public Health and Family Medicine, University of Malawi College of Medicine, Blantyre, Malawi

[5]Center for Development of People, Blantyre, Malawi

Abstract

Background—HIV risks are heterogeneous in nature even in generalized epidemics. However, data are often missing for those at highest risk of HIV, including female sex workers. Statistical models may be used to address data gaps where direct, empiric estimates do not exist.

Methods—We proposed a new size estimation method that combines multiple data sources (the Malawi Biological and Behavioral Surveillance Survey, the Priorities for Local AIDS Control Efforts study, and Malawi Demographic Household Survey). We employed factor analysis to extract information from auxiliary variables, and constructed a linear mixed effects model for predicting population size for all districts of Malawi.

Results—On average, the predicted proportion of FSW among women of reproductive age across all districts was about 0.58%. The estimated proportions seemed reasonable in comparing with a recent study PLACE II. Compared to using a single data source, we observed increased precision and better geographic coverage.

Conclusions—We illustrate how size estimates from different data sources may be combined for prediction. Applying this approach to other sub-populations in Malawi and to countries where size estimate data are lacking can ultimately inform national modeling processes and estimate the distribution of risks and priorities for HIV prevention and treatment programs.

E-mail address of the corresponding author: Le Bao, PhD, Pennsylvania State University, Eberly College of Science, 514C Wartick Lab, University Park, PA 16802, lebao@psu.edu, Stefan Baral, MD, Johns Hopkins Bloomberg School of Public Health, 615 N. Wolfe Street, Room E7146, Baltimore, Maryland 21205, sbaral@jhu.edu.
*Contributed equally

Keywords

HIV; Female Sex Workers; Malawi; Prediction; Size Estimation

INTRODUCTION

Despite advancements in treatment coverage and striking declines in mortality among people living with HIV, the HIV pandemic is far from over[1]. Annual incidence rates in many settings remain unchanged[1]. While treatment and prevention strategies have been improving, ongoing implementation challenges related to service delivery and fidelity of interventions, combined with intersecting stigmas, collectively contribute to approximately two million people acquiring HIV each year[1–5]. Close to half of these new infections are among people living across Sub-Saharan Africa[1]. Historically, HIV prevention and treatment programming in Sub-Saharan Africa has been designed and delivered based on reaching the “general population,” which tacitly assumes a relatively equal distribution of risk across all individuals within particular demographic categories[6]. However, data consistently suggest that even in settings with high overall prevalence, heterogeneities in HIV acquisition and transmission remain[7,8]. Providing better services to those living with HIV requires understanding of the actual, rather than assumed, determinants of HIV transmission to inform evidence-based responses.

In Malawi, both sex work and same-sex practices are stigmatized and criminalized, resulting in knowledge gaps around the specific HIV prevention and treatment needs for key populations including sex workers of all genders, gay, bisexual, and other cisgender men who have sex with men, and transgender persons[9]. The national HIV prevalence among female sex workers (FSW) is estimated to be greater than 60%, with one study conducted in Lilongwe in 2014 suggesting a 69% HIV prevalence in a sample of 200 FSW[10,11]. The same study in Lilongwe found that among those living with HIV, just over half were on treatment (72/138) and about 45% (62/138) were virally suppressed[10]. Despite the high burden and gaps in treatment coverage, less is known about the population-level scale of the need – in terms of how many FSW there are in Malawi, and in which districts.

Population size estimates play a critical role in both the relative allocation of funding and in HIV program planning and execution[12–15]. Despite its importance, empiric size estimation can prove especially challenging for populations who, because of continued intersecting stigmas, are fearful of engagement and disclosure in both data collection and in service delivery programs[14–16]. Membership in key population groups is difficult to measure in traditional census and household surveys, and substantial time and resources are required to conduct empiric data collection activities in every setting where size estimates are needed[14,16]. Although there have been several efforts to estimate the number of female sex workers in Malawi, the estimates vary by method and area and there is not yet a consensus on the size estimates at the district level[11,17].

Recently, there has been a series of applications of advanced statistical methods for the extrapolation of size estimation data, generating population size estimates in areas where there is either an absence or a limited number of empiric data[18–20]. Traditional extrapolation approaches that produce estimates where there are no empiric data for FSW and other key populations often rely on simple or stratified imputation using existing estimates from different geographic areas. These approaches assume a degree of homogeneity that is unlikely to hold true given the high levels of migration, tendency to congregate around urban centers, and geospatial diversity[16,19,21]. In many cases, extrapolation has focused on consensus approaches in the absence of empiric data[22]. More analytically rigorous approaches are available and can be used to better fill gaps where direct size estimates do not exist[19,21].

The goal of this study is to contribute to the range of methods being used to estimate the size of key populations where data do not exist, and to generate district-level estimates of FSW population size in Malawi in order to improve efficient and efficacious program delivery. In what follows, we first introduce the existing data sources. Then, we describe the modeling procedure. Next, we present the results and validation. Finally, we discuss the impact of the study and future directions.

MATERIALS AND METHODS

1. Data description

We conducted a comprehensive search for available size estimates in Malawi as part of a broader systematic review[23]. This search process has been previously described[23]. We identified two available data sources for district-level FSW size estimates: 1) the 2014 Malawi Biological and Behavioral Surveillance Survey (BBSS) Report from the National Statistical Office of Malawi[11]; and 2)The Priorities for Local AIDS Control Efforts (PLACE) report from the University of North Carolina[17].

1.1 Malawi Biological and Behavioral Surveillance Survey (BBSS) 2013–2014—The BBSS was conducted between 2013 and 2014 as a collaborative effort by the National AIDS Commission and the National Statistics Office[11]. The BBSS was implemented in order to monitor trends in key indicators among population groups at high risk for HIV acquisition and transmission[11]. One of the primary objectives of the BBSS was to estimate the population size of FSW and provide information that could help guide both program planning and interventions among high-risk groups[11]. To inform the size estimation activities, a “mapping exercise” was conducted to identify possible locations and times where FSW could be found[11]. A total of 626 sites, including bars and clubs, in 14 districts (of 28 total districts) were mapped and 180 (30%) of them were randomly chosen for the size estimation exercise using probability-proportional-to-size (PPS) sampling[11]. This randomly selected list of 180 hotspots constituted a proxy sampling frame[11].

Two different size estimation methods were used in the BBSS[11]. The first was the “Enumeration method[11,24].” All eligible FSW were counted at the selected sites, and this number of enumerated FSW was scaled up by multiplying the average number of people associated with a site by the total number of sites in the sampling frame[11,24].

The second method used was the “Capture-recapture method”[24] (CRM). Two rounds of data collection were conducted, approximately two weeks apart[11,24]. During the first sampling round, sites were visited at the peak of client-seeking activity[11]. Consenting FSW were given a bangle to denote that they had been visited ("captured") in the first round of surveys[11]. Two weeks later, interviewers returned to the sites at the same time and day of the week that they visited previously[11]. FSW present this second time were shown the bangle and asked whether they received one during the previous visit[11].

The following information was documented:

- The number of first capture C_1: all FSW who received a bangle during the first round of data collection
- The number of second capture C_2: all FSW seen during the second round of data collection
- The number of overlaps m: FSW who received a bangle during the first sample and were contacted again during the second sample
- The number of FSW was estimated using the following equation[11,24]:

$$N = \frac{(C_1 + 1)(C_2 + 1)}{m + I}$$

1.2 PLACE I—The second identified source of size estimate data for FSW was the PLACE study[17]. In Malawi, the study had two phases (termed PLACE I and PLACE II). PLACE I was conducted in 2016 and covered six districts: Lilongwe, Blantyre, Zomba, Mangochi, Mzuzu, and Machinga[17]. One of the primary objectives of the PLACE study was to conduct programmatic mapping in selected districts to identify venues where key populations can be reached and estimate population size in each district[17].

For PLACE I, estimation began with identifying public locations where FSW might congregate[17]. By interviewing community informants (e.g. business owners and taxi drivers) and asking where people go to meet new sexual partners, a list of public venues was compiled[17]. According to the protocol, all identified venues were to be visited, characterized and mapped. After venue visits began, however, the target number of venues was decreased due to more sites being listed than expected and difficulty with transportation. Consequently, in addition to the venues visited prior to the protocol change, a random sample of unvisited locations was selected for a visit[17]. The selection process oversampled for venues reported to have FSW or sex work activities on site, and were limited to those within five kilometers of a main road[17]. At the selected sites, a general venue informant (e.g. bar manager and venue staff) and one to three FSW were asked questions about how many FSW regularly come to the site. This information was used to estimate a venue-informant and a FSW-informant estimate of the numbers of FSW at venues at busy times. To develop district level estimates, within each district, these estimates from visited venues were scaled up by the total estimated number of venues. Table 1 summarizes the main features of the two size estimation data sources and compares them side by side. The PLACE study visited more venues while the BBSS covered more districts.

Figure 1 compares the proportion of FSW among women of reproductive age (15–49 years old), estimated by the PLACE report (blue) and BBSS report (enumeration results in green and capture-recapture in red). PLACE size estimates are consistently higher than the BBSS numbers. We will discuss more about this difference in the Discussion section.

2. Statistical Modeling and Prediction of District-level Sizes

2.1 Auxiliary data and selection of covariates—Given these existing data, the goal of the current analysis was to produce size estimates of FSW in Malawi for all **28 districts**. For the districts without any size estimates, additional data was reviewed to understand if there exist other variables that may aid in discerning differences between districts, that is, in serving as covariates in a prediction model of FSW population size. This additional data is often referred to as “auxiliary data[18–20,25].” Therefore, we developed a systematic approach for considering auxiliary variables to improve extrapolation. This idea is similar in concept to the model-based method in Small Area Estimation[25].

We utilized the Malawi Demographic and Health Survey (MDHS) to identify auxiliary variables. MDHS is a nationally representative household survey in Malawi that has been conducted since 1992. We used the 2015–2016 survey, which has over 900 variables summarized at the district-level[26].

First, we screened the candidate auxiliary variables for inclusion in our prediction model. For each source of size estimates (PLACE, CRM, and Enumeration), we calculated pairwise correlations between each predictor and the size estimates, and selected the top 10 predictors based on the smallest ten p-values. These yielded three sets of ten predictors, one for each source of size estimate data. We then combined these three sets of predictors and removed those that had high correlations with the others (correlation coefficient>0.9 or <−0.9).

After this initial screening, we performed factor analysis on those predictors identified from the screening process. Factor analysis is a dimension reduction tool that has the advantage of using a relatively low number of variables while maintaining a majority of the information in the original data. In order to determine the number of factors needed in total, we created a scree plot. A scree plot demonstrates the marginal improvement in terms of proportion of variance explained, by increasing number of factors. It was determined that five factors explained over 70% of the total variance.

Each of the approximately 900 variables in the DHS was assessed for how closely it could be linked to one of the five factors. Factor 1 focuses on education and urbanization (measured as household possession of TV and internet). Factor 2 is a mix of family related variables (e.g. family has electricity), specifically focusing on the female head of the household, such as female earning and how they raised children. Factor 3 is mainly about household male’s education, occupation, and attitudes. Factor 4 is about family living conditions, and factor 5 is about exposure to family planning messaging. Based on factor loadings, these five factors were all relatively interpretable and represent different aspects of a district’s urbanization and socio-economic level. We used those five factors as predictors in the final model.

In addition to the factor analysis, after initial screening, we also performed variable selection within each data source. Then we pooled the three sets of selected variables together (direct pooling). Details of the factor loadings are available in the Table 1 of the Appendix. The variable names that are highlighted in red are the predictors in the final model using direct pooling.

2.2 Fitted models—We compared two approaches to build a joint model combining the estimates from the three data sources. The first approach was via standard variable selection for each data source and directly combining the three sets of predictors as the final predictors. The second approach utilized a factor analysis in the variable selection[27].

Finally, we built a mixed effects model to estimate the FSW population size for each district, with fixed effects being either the directly combined predictors from the first approach, or the factors from the second approach, and the random effects being the district and data source. As shown in Table 1 and Figure 1, the three data sources yielded very different size estimates. In this way, we can use all three data sources, while allowing them to have systematic differences in the model.

We fit the following linear mixed effects model for the FSW prevalence on the log scale (to correct the skewness in the original count scale):

$$\text{Log}(\#\text{ FSW}/\#15\text{–}49\text{ female}) \sim X\beta + b_{\text{district}} + b_{\text{source}} + \text{error},$$

where X are either the factors or the predictors, as discussed above, β are the fixed effects coefficients, $b_{district}$ and b_{source} are the random effects for district and source, respectively. By including the random effects, observations within a district and from the same source are correlated and share some similarity. We estimated the model by Bayesian estimation using R package INLA[28].

3. Validation

The second phase of the PLACE project (PLACE II) was implemented in 2017 in an additional 15 districts of Malawi. For PLACE II, venue identification followed similarly as in PLACE I with community informants, but focused on places designated as Priority Prevention Areas, or predetermined areas of higher risk (fishing villages, mining camps, tourist areas). The venues were then divided into either high priority (contains MSM, sex activities on site, or resident FSW), low priority, or inaccessible (too far, lack of information, outside of time frame). In each district, 90 high priority and 30 low priority venues were randomly selected for interviews[17].

Due to the slightly different sampling framework in PLACE II, rather than combining PLACE II data with PLACE I, in this study, we treated PLACE II data as the validation set to assess the performance of our model. We fit the model with BBSS and PLACE I data and compared the predicted FSW sizes with the PLACE II estimates.

RESULTS

Predicted size estimates

The posterior median and 95% credible intervals of the coefficients from the direct pooling and factor models can be found in Table 2 and Table 3 in the Appendix. Size estimates from the empirically collected data (PLACE, Capture-Recapture, Enumeration), together with the model predicted size estimates at the district level and their credible intervals are presented in Table 2. The two largest predicted estimates were seen in the two major urban centers of Malawi: Lilongwe (Size Estimate: 2838, 95% CI [1248, 6468]) and Blantyre (Size Estimate: 3399, 95% CI [1483, 7787], corresponding to 0.97% and 1.1% of all women of reproductive age, respectively. On average, the predicted proportion of FSW among all women of reproductive age across all districts was about 0.58%. The largest predicted proportions were seen in Mzimba and Rumphi, 1.12% and 1.28%, respectively. The lowest predicted proportions were seen in Mangochi, Dedza, and Machinga, 0.17%, 0.26%, and 0.27% respectively.

We also mapped the geographic distribution of FSW population size. The heat map presented in Figure 2 displays the district-level estimates of predicted FSW size. This heat map shows that the highest predicted proportion of FSW are in districts in the Northern Region, with a high concentration also observed in one district in the Southern Region: Blantyre. The lowest proportions are seen in the districts in the southeast that border Mozambique: Balaka, Mangochi, Dedza, and Salima.

Validation

In Figure 3, we compare the prediction intervals of all districts in Malawi produced by the direct pooling model (red) and the factor model (blue). We see that the factor model produces much more stable and shorter prediction intervals in districts without any observed data, which indicates that the factor model utilizes more of the auxiliary information with fewer parameters. For most of the new districts in PLACE II (plus sign), the prediction intervals covered the PLACE II numbers. In three districts, Mwanza, Salima, and Nkhotakota, the PLACE II estimates of population size were higher than the predicted ones. In two out of the three districts where the PLACE II estimates were outside of the predicted interval, the existing data sources were all highly consistent with each other, but dramatically different from PLACE II.

CONCLUSION

In this study, existing size estimation data from two different sources with three different size estimation activities and auxiliary information were utilized in a factor model to estimate the number of FSW in all districts of Malawi. The final model developed in this analysis produced an average predicted proportion of FSW among all women of reproductive age of 0.58%. The model's prediction is consistent with empiric data collected subsequently. The proposed approach can be applied to other countries or regions where multiple sources of key population size estimates exist. The factor analysis can be especially helpful if the number of observations is small relative to the number of auxiliary variables.

Applying similar approaches where data allow can ultimately inform national modeling processes to estimate the distribution of risks and priorities for HIV prevention and treatment programs.

In our review of existing empiric data in Malawi, we found that available size estimation data covered only a limited number of districts. This is consistent with what is reported elsewhere: in many countries with available key populations size estimation data, data are often inadequate in terms of their geographic coverage, that is estimates are only available for some of the administrative units and are missing for others[14]. On average, estimates from the BBSS were lower than those from PLACE I for the same districts, suggesting some systematic differences between the two. The observed differences may arise from the sampling scheme, the places within each district visited, the estimation methods used, and different definitions of FSW [22]. This demonstrates the potential value of using multiple sources of size estimate data to inform both extrapolation and decision-making in policies and programs, as compared to a single data source. The approach described here allows for the use of multiple data sources and the incorporation of information from over 900 auxiliary variables in the estimation of population size where data do not exist.

Validation of prediction models can bolster confidence in their utility and facilitate decision-making[29]. While validation methods have been used previously as part of size estimation approaches, they have been used primarily during empiric data collection activities[30]. For example, when results were found to be inconsistent between two field teams mapping hotspots in India, remapping and validation was conducted[30]. There are few instances in which extrapolated size estimates have been validated. The natural split between data collection activities of PLACE I and II data allowed us to validate our model and results, by comparing the overlap in predicted estimates with empiric estimates from PLACE II. Our validation results demonstrate the efficiency of the model with prediction intervals covering most of the PLACE II empiric estimates. Compared with other approaches that primarily allow for internal validation and model accuracy, the structure of our empiric estimates allowed us to quantitatively evaluate model predictive accuracy. This validation process may be especially valuable in engaging key stakeholders and policy makers in the modeling approaches and increasing the likelihood of consensus on the final estimates provided[30].

The current study has several limitations. First, the empiric size estimates used in the prediction models are from slightly different years and use different methods. Differences in the timing of size estimation activities could have resulted in the capture of a different underlying population, due to factors such as in- and out-migration, seasonality, or changing population densities. Differences in the methods used could have captured individuals who are differentially engaged in a community, for example venue-based estimates likely capture those who are more visible as key population members and therefore potentially more likely to access services[31]. The method used in the current study did make use of estimates from multiple sources, which is an improvement over existing methods which are often reliant on a single data source[30]. The model assumes that biases of different sources balance out. If all of the biases are in the same direction, without a gold standard to correct it, the estimates would carry biases with the same direction. Second, the predicted proportions are reliant on selection of a total population denominator. For cities, this population denominator can be challenging to define (e.g. main urban center or urban center and all surrounding areas). Interpretation of these results should focus both on the prevalence of FSW (reliant on total population denominator), as well the the estimate and confidence interval (not reliant on total population denominator). Third, the resulting confidence intervals using this approach are wide and may have implications for programs given that the lower and upper bounds differ substantially. Despite this limitation, many existing size estimates have no measure of confidence reported with them, and this approach gives the user of the estimates a quantifiable measure of uncertainty. Fourth, we did not take into account the number of sites visited or the potential reliability of the empiric size estimates. As described in Table 1, the PLACE project visited many more sites than the BBSS report. As a result, we could expect that the PLACE numbers might be more reliable. In this article, we did not make this distinction, but potential weighting to reflect the reliability of different data sources could be implemented. Another limitation of PLACE I data is that the method is based on asking FSW or Venue Informant about how many FSW come to the place, thus relying on perception of respondents. FSW may have an incentive to overestimate. Venue informants may over or underestimate. The definition of FSW also may vary by person.

The methodologic approach developed and tested here not only generates estimates for FSW population size in Malawi, but also represents an example of how empiric size estimates from different data sources may be used for extrapolation and prediction. The use of DHS data, which is available in many resource-constrained settings, as auxiliary data will facilitate the replication of a similar approach in multiple settings. Achieving greater HIV prevention impacts with sustained levels of resources likely necessitates further specificity to optimize HIV incidence reductions with both prevention and treatment approaches[1,32,33]. Applying the approach developed in this study to data for other populations can ultimately provide denominators for assessment of program coverage, facilitate prioritization of HIV programs, inform resource allocation, and generate data inputs for national modeling processes to estimate the distribution of risks and priorities[15].

Supplementary Material

Refer to Web version on PubMed Central for supplementary material.

Acknowledgement

This research was supported by NIAID/NIH R01-AI136664 and NIH F31MH124458.

References

1. Collaborators GH. Estimates of global, regional, and national incidence, prevalence, and mortality of HIV, 1980–2015: the Global Burden of Disease Study 2015. Lancet HIV. 2016;3(8):e361–e387. [PubMed: 27470028]
2. Friedland BA, Sprague L, Nyblade L, et al. Measuring intersecting stigma among key populations living with HIV: implementing the people living with HIV Stigma Index 2.0. J Int AIDS Soc. 2018;21 Suppl 5:e25131. [PubMed: 30033563]
3. Baral S, Phaswana-Mafuya N. Rewriting the narrative of the epidemiology of HIV in sub-Saharan Africa. SAHARA J. 2012;9(3):127–130. [PubMed: 23237066]

4. Kim HY, Grosso A, Ky-Zerbo O, et al. Stigma as a barrier to health care utilization among female sex workers and men who have sex with men in Burkina Faso. Ann Epidemiol. 2018;28(1):13–19. [PubMed: 29425532]
5. Rodriguez-Hart C, Bradley C, German D, et al. The Synergistic Impact of Sexual Stigma and Psychosocial Well-Being on HIV Testing: A Mixed-Methods Study Among Nigerian Men who have Sex with Men. AIDS Behav. 2018.
6. Boily MC, Pickles M, Alary M, et al. What really is a concentrated HIV epidemic and what does it mean for West and Central Africa? Insights from mathematical modeling. J Acquir Immune Defic Syndr. 2015;68 Suppl 2:S74–82. [PubMed: 25723994]
7. Pettifor A, Rosenberg N, Behets F. The need to focus on sex workers in generalized HIV epidemic settings. Sex Transm Dis. 2011;38(4):324–325. [PubMed: 23330153]
8. Mishra S, Boily MC, Schwartz S, et al. Data and methods to characterize the role of sex work and to inform sex work programs in generalized HIV epidemics: evidence to challenge assumptions. Ann Epidemiol. 2016;26(8):557–569. [PubMed: 27421700]
9. Isaack W "Let Posterity Judge": Violence and discrimination against LGBT people in Malawi. 2018; https://www.hrw.org/report/2018/10/26/let-posterity-judge/violence-and-discrimination-against-lgbt-people-malawi.
10. Lancaster KE, Powers KA, Lungu T, et al. The HIV Care Continuum among Female Sex Workers: A Key Population in Lilongwe, Malawi. PLoS One. 2016;11(1):e0147662. [PubMed: 26808043]
11. National Statistical Office of Malawi: Malawi Biological and Behavioural Surveillance Survey Report (BBSS), 2013–2014. 2014.
12. Baral SD, Lyons C, Sullivan E, et al. The uptake of population size estimation studies for key populations in guiding HIV responses across sub-Saharan Africa: a systematic review. Paper presented at: AIDS 20162016; Durban, South Africa.
13. Holland CE, Kouanda S, Lougue M, et al. Using Population-Size Estimation and Cross-sectional Survey Methods to Evaluate HIV Service Coverage Among Key Populations in Burkina Faso and Togo. Public Health Rep. 2016;131(6):773–782. [PubMed: 28123223]
14. Sabin K, Zhao J, Garcia Calleja JM, et al. Availability and Quality of Size Estimations of Female Sex Workers, Men Who Have Sex with Men, People Who Inject Drugs and Transgender Women in Low- and Middle-Income Countries. PLoS One. 2016;11(5):e0155150. [PubMed: 27163256]
15. Viswasam N, Lyons CE, MacAllister J, et al. The uptake of population size estimation studies for key populations in guiding HIV responses on the African continent. PLoS One. 2020;15(2):e0228634. [PubMed: 32101551]
16. Abdul-Quader AS, Baughman AL, Hladik W. Estimating the size of key populations: current status and future possibilities. Curr Opin HIV AIDS. 2014;9(2):107–114. [PubMed: 24393694]
17. UNC, CEDEP, NAC, FHI360. PLACE Report Malawi: September 2018. 2018; https://www.fhi360.org/sites/default/files/media/documents/resource-linkages-malawi-place-report.pdf.
18. Datta A, Lin W, Rao A, et al. Bayesian Estimation of MSM Population Size in Cote d'Ivoire. Stat Public Policy (Phila). 2019;6(1):1–13. [PubMed: 31341935]
19. Edwards JK, Hileman S, Donastorg Y, et al. Estimating Sizes of Key Populations at the National Level: Considerations for Study Design and Analysis. Epidemiology. 2018;29(6):795–803. [PubMed: 30119057]
20. Rao A, Loo V, Saidel T, Datta A, Baral S. Extrapolation of population size estimates for key populations: Impact of method of extrapolation on estimates produced using regional estimates of female sex workers from Tanzania as an illustrative example. Journal of the International Aids Society. 2019;22:81–82.
21. Datta A, Lin W, Rao A, et al. Bayesian estimation of MSM population in Cote d'Ivoire. (Submitted).
22. WHO/UNAIDS. Estimating sizes of key populations: guide for HIV programming in countries of the Middle East and North Africa. Geneva, Swaitzerland: World Health Organization. Regional Office for the Eastern Mediterranean;2016.
23. Rao A, Schwartz S, Sabin K, et al. HIV-related data among key populations to inform evidence-based responses: protocol of a systematic review. Syst Rev. 2018;7(1):220. [PubMed: 30509317]
24. UNAIDS/WHO. Guidelines on Estimating the Size of Populations Most at Risk to HIV. 2010.

25. Rao JNK. Small Area Estimation. Hoboken, New Jersey: Wiley; 2003.
26. National Statistical Office (NSO) and ICF. Malawi Demographic and Health Survey 2015–2016. Zomba, Malawi, and Rockville, Maryland, USA2017.
27. Gorsuch RL. Factor Analysis (2nd Ed.). Hillsdale, New Jersey: Lawrence Erlbaum Associates; 1983.
28. Rue H, Martino S, Chopin N. Approximate Bayesian inference for latent Gaussian models by using integrated nested Laplace approximations. Royal Statistical Society. 2009;71(2):319–392.
29. Moons KG, Kengne AP, Grobbee DE, et al. Risk prediction models: II. External validation, model updating, and impact assessment. Heart. 2012;98(9):691–698. [PubMed: 22397946]
30. Yu D, Calleja JM, Zhao J, Reddy A, Seguy N, Technical Consultation on Lessons Learnt from Size Estimation among Key Populations in Asian C. Estimating the size of key populations at higher risk of HIV infection: a summary of experiences and lessons presented during a technical meeting on size estimation among key populations in Asian countries. Western Pac Surveill Response J. 2014;5(3):43–49. [PubMed: 25320676]
31. Rao A, Stahlman S, Hargreaves J, et al. Sampling Key Populations for HIV Surveillance: Results From Eight Cross-Sectional Studies Using Respondent-Driven Sampling and Venue-Based Snowball Sampling. JMIR Public Health Surveill. 2017;3(4):e72. [PubMed: 29054832]
32. UNAIDS. Ending AIDS: progress towards the 90–90–90 targets. 2017; https://www.unaids.org/en/resources/documents/2017/20170720_Global_AIDS_update_2017.
33. Baral S, Rao A, Sullivan P, et al. The disconnect between individual-level and population-level HIV prevention benefits of antiretroviral treatment. Lancet HIV. 2019;6(9):e632–e638. [PubMed: 31331822]

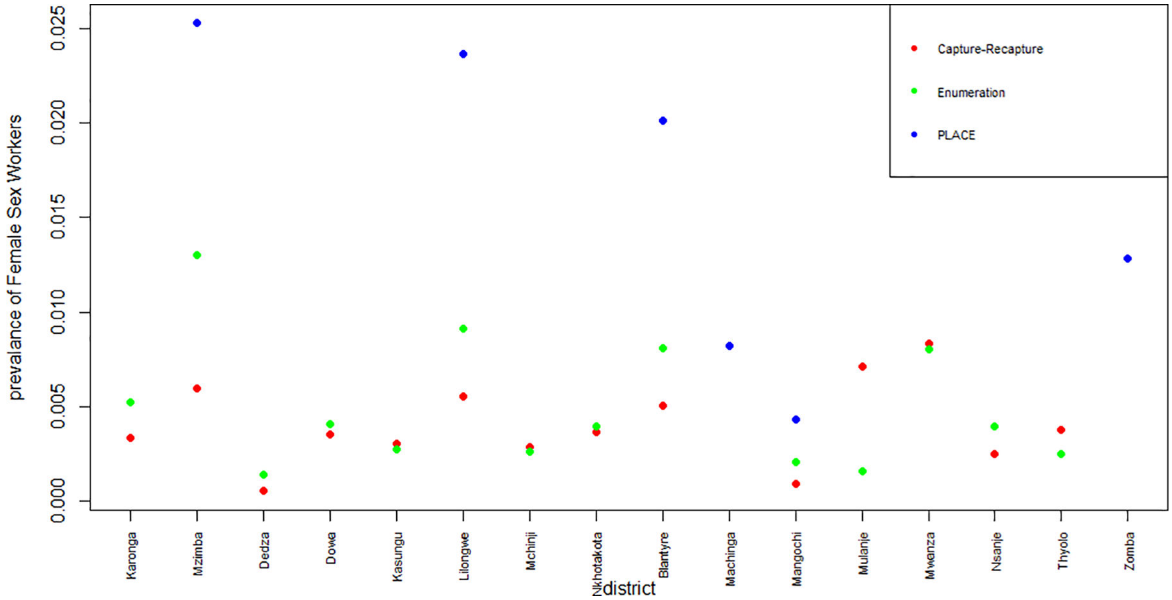

Figure 1:

Comparison of the estimated proportion of Female Sex Workers in the general female population 15–49 years old for districts with available size estimates from the BBSS and PLACE I

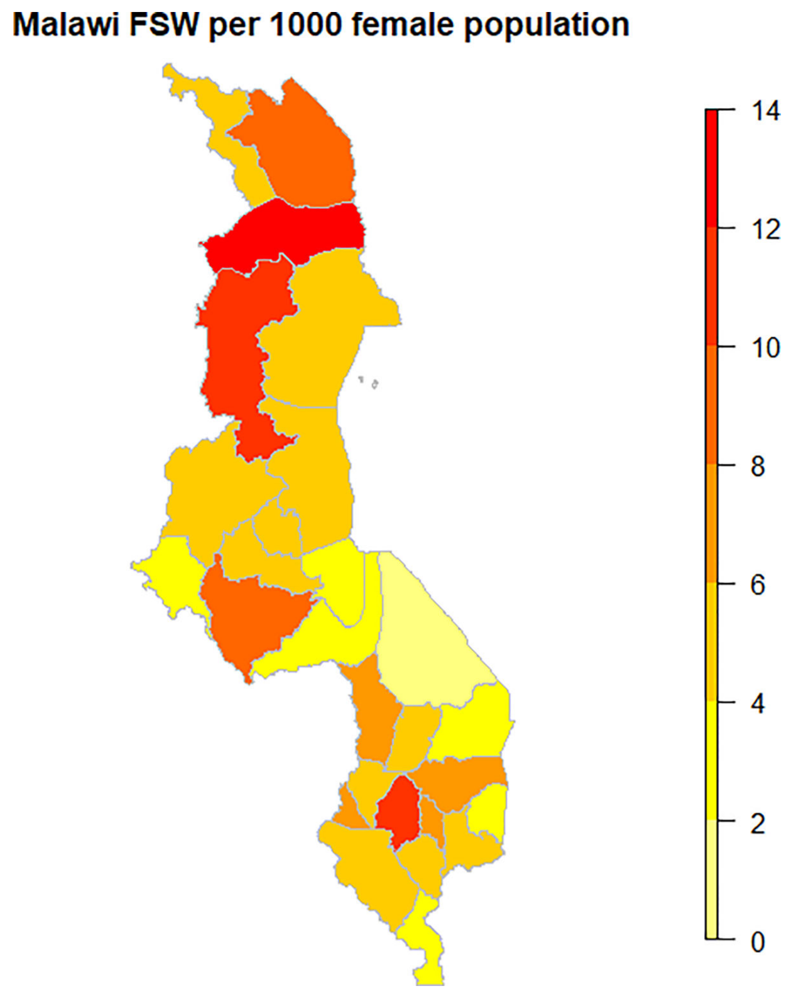

Figure 2:

Predicted Malawi district level Female Sex Workers (FSW) sizes based on factor analysis modeling approach

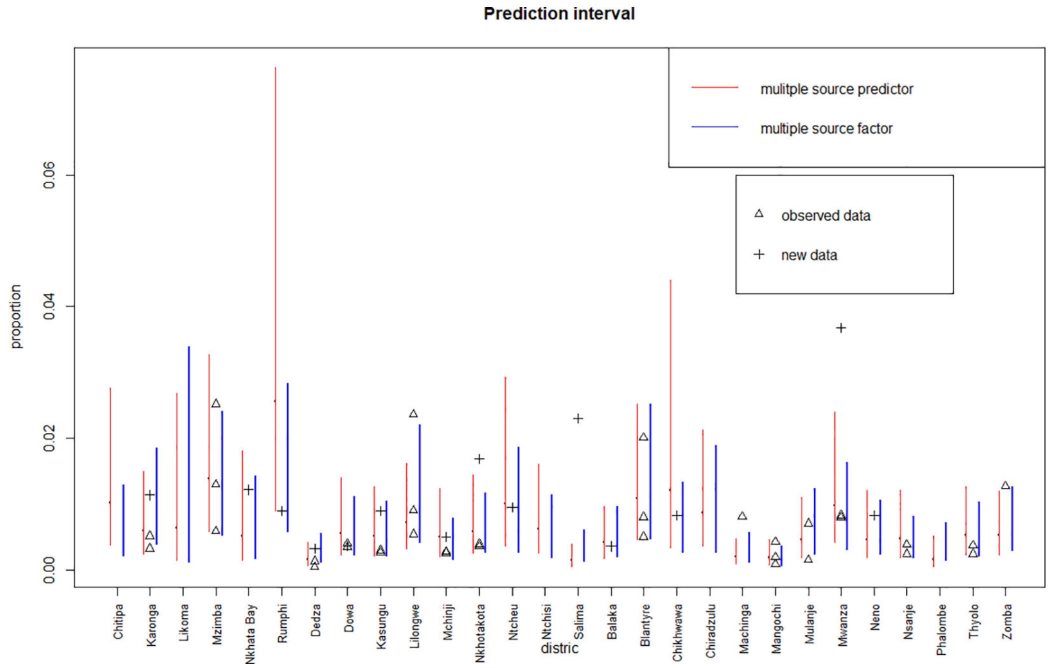

Figure 3:
Prediction interval comparison of direct combined predictors (red), and factor model (blue). The triangles indicate the observed data, and plus signs indicate the PLACE II data.

Table 1:

Comparison of the data collection methods used for the Malawi Biological and Behavioral Surveillance Survey (BBSS) and The Priorities for Local AIDS Control Efforts (PLACE) Study

	PLACE I	Malawi BBSS 2013–2014
Years	2016	2013–2014
Site Identification	General Informants	Field team travels around the district locating sites
How FSW were defined	No formal definition; based on venue-informant and FSW-informant reporting	“Women or girls aged 18 years or more who have exchanged sex for money to earn a living at least once in the last 30 days”[11]
Mapping (site locations)	Cities, rural areas (only visited rural areas not more than 5km off a main road.) Attempted to visit all, realized there were too many, randomly sampled remaining sites within 5km of major road	District towns and all major trading centers in the rural areas. 30% of sites randomly selected for visits
Number of Districts covered	6	14
Number of sites used to estimate FSW size	69–413	5–43
Average number of sites per district used to estimate FSW size	166	13
Size Estimation Calculation	Crude Estimate: (Avg # of FSW at visited sites) × (total # of sites) FSW and General Informant interview estimates. Accounts for FSW visiting multiple sites.	Enumeration: (Avg # of FSW present at visited sites) × (total # of sites) CRM – 2 weeks apart between first and second visit. N = (C1+1) (C2+1)/(m+1)
Size Estimate Constraint	District level, only estimates FSW who go to sites	District level, only estimates FSW who go to sites
Limitations	Assumes all sites to be similar, lacks sites in rural areas off road, sites all weighted the same	Assumes all sites to be similar, missing various rural areas (non-major trading center areas), no weighting, less types of venues compared to PLACE I

Table 2:

District Level Size Estimates: comparing estimates from PLACE, BBSS, and the model predictions

District	PLACE	BBSS_CRM	BBSS_ENUM	size estimate	2.5% lower bound	97.5% upper bound	prevalence%
Chitipa	NA	NA	NA	235	99	563	0.54
Karonga	800	234	366	593	282	1298	0.85
Mzimba	1400	336	732	629	297	1355	1.12
Nkhata Bay	700	NA	NA	298	104	821	0.52
Rumphi	400	NA	NA	563	257	1248	1.28
Dedza	500	81	215	384	185	850	0.26
Dowa	600	576	666	809	391	1830	0.50
Kasungu	1500	511	455	785	377	1755	0.47
Lilongwe	7000	1622	2676	2838	1248	6468	0.97
Mchinji	600	337	310	422	200	930	0.36
Nkhotakota	1300	279	305	432	215	903	0.56
Ntcheu	1100	NA	NA	801	318	2150	0.70
Ntchisi	NA	NA	NA	256	111	640	0.46
Salima	2000	NA	NA	254	120	537	0.29
Balaka	300	NA	NA	374	175	800	0.46
Blantyre	6200	1568	2491	3399	1483	7787	1.10
Chikhwawa	900	NA	NA	628	296	1454	0.58
Chiradzulu	NA	NA	NA	463	180	1256	0.70
Machinga	1000	NA	NA	336	160	714	0.27
Mangochi	900	204	444	363	164	798	0.17
Mulanje	NA	864	193	651	296	1505	0.54
Mwanza	800	182	174	151	69	357	0.69
Neno	300	NA	NA	175	87	382	0.49
Nsanje	NA	140	224	221	111	464	0.39
Phalombe	NA	NA	NA	250	122	553	0.32
Thyolo	NA	500	330	627	300	1358	0.48
Zomba	1800	NA	NA	826	406	1724	0.61