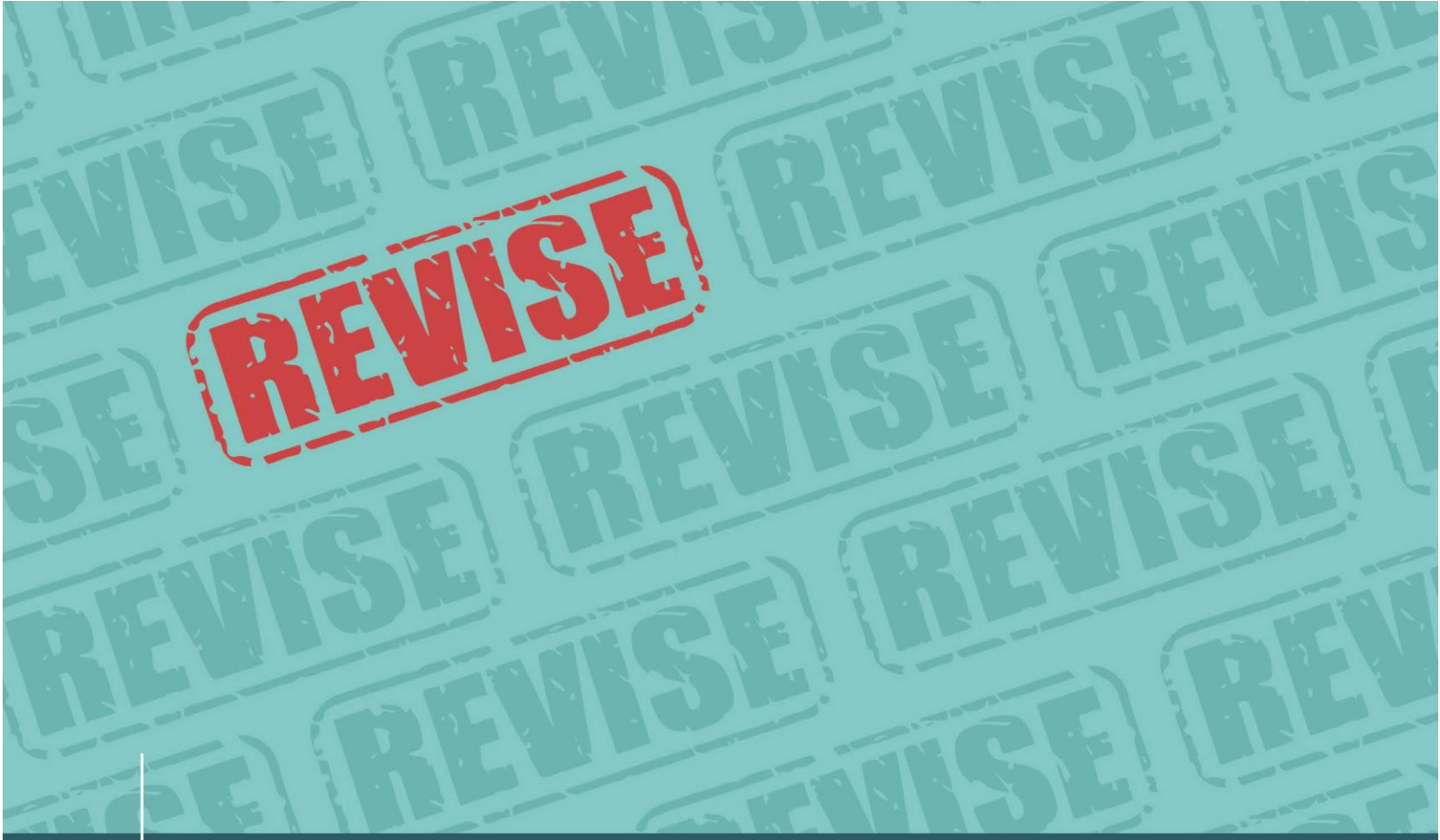


Review and Revise

A Tool for Reviewing and Revising Key Population Size Estimates

July 2018

USAID
FROM THE AMERICAN PEOPLE
PEPFAR
MeSH Consortium
Measurement & Surveillance of HIV Epidemics
MEASURE Evaluation

Review and Revise

A Tool for Reviewing and Revising Key Population Size Estimates

July 2018

MEASURE Evaluation
University of North Carolina at Chapel Hill
123 West Franklin Street, Suite 330
Chapel Hill, North Carolina 27516
Phone: +1 919-445-9350
measure@unc.edu
www.measureevaluation.org

This publication was produced with the support of the United States Agency for International Development (USAID) under the terms of the MEASURE Evaluation cooperative agreement AID-OAA-L-14-00004. MEASURE Evaluation is implemented by the Carolina Population Center, University of North Carolina at Chapel Hill in partnership with ICF International; John Snow, Inc.; Management Sciences for Health; Palladium; and Tulane University. Views expressed are not necessarily those of USAID or the United States government. TR-18-274

USAID FROM THE AMERICAN PEOPLE
PEPFAR
MeSH Consortium Measurement & Surveillance of HIV Epidemics
MEASURE Evaluation

ACKNOWLEDGMENTS

We thank the United States Agency for International Development (USAID) and the United States President's Emergency Plan for AIDS Relief (PEPFAR) for their support of this work.

We thank Lauren Zalla, Zahra Reynolds, Sharon Weir, and Jess Edwards of MEASURE Evaluation at the University of North Carolina at Chapel Hill (NC), and MEASURE Evaluation consultants Sarah Hileman and Tobi Saidel for their contributions to this paper.

We appreciate the members of the MeSH Consortium Key Populations Working Group for their input and review.

We also thank the MEASURE Evaluation knowledge management team for editorial, design, and production services.

Finally, we extend special thanks to our colleagues in Guyana, Haiti, and the Dominican Republic for their time and the valuable information they provided and the opportunity to work with them on size estimates for their country.

CONTENTS

TABLES

ABBREVIATIONS

CEPROSH	Centro de Promoción y Solidaridad Humana
FSW	female sex worker
GFATM	Global Fund to Fight AIDS, Tuberculosis and Malaria
UNAIDS	Joint United Nations Programme on HIV/AIDS
MSM	men who have sex with men
NAPS	National AIDS Programme Secretariat
PEPFAR	United States President's Emergency Plan for AIDS Relief
PLACE	Priorities for Local AIDS Control Efforts
PNLS	Programme National de Lutte contre les IST/Sida
USAID	United States Agency for International Development

WHAT IS THE PROBLEM BEING ADDRESSED?

Key populations—a term used to reference sex workers, men who have sex with men, transgender people, people who inject drugs, and prisoners—are at greater risk for acquiring HIV and are least likely to access services, including treatment. Understanding the size of key populations is important to ensure that resources are allocated to reach these populations with the services and treatment they require. Unfortunately, estimating the numbers of these often hidden and mobile populations who engage in illegal or stigmatized sexual behavior is difficult, with no gold standard method for providing consistently valid and precise estimates. Countries have invested resources in multiple methods to estimate the size of key populations in subnational areas and extrapolate findings to the country as a whole. Consequently, countries frequently have multiple and conflicting size estimates of key populations for subnational areas and for the entire country. Multiple estimates often cause confusion when program targets are set and resources are allocated based on projected need.

Size estimates for the same population may differ for four broad reasons. The first is that the estimates may use differing definitions of the key populations. For example, one definition of a female sex worker could require that an exchange of sex for cash in the past three months. A broader definition would include exchange of sex for cash, gifts, or services in the past year.

The second reason estimates may differ is because they may reflect that the geographic area or the time frame for the two estimates is not aligned. One estimate may be for female sex workers in a particular city. Another may include all female sex workers in the district where the city is located. Most size estimates reflect an estimate at a particular point in time rather than the number in an area over the course of a year. Unless the estimates are very clearly described, we may overlook the fact that they actually capture different populations—so we're comparing apples and oranges.

The United States President's Emergency Plan for AIDS Relief (PEPFAR) supports efforts to reduce HIV transmission. Addressing transmission of HIV among key populations is a major objective for PEPFAR, focusing investments in the geographic areas and populations with the greatest HIV/AIDS burden and on data-driven decision making, accountability, transparency, and using data to do the right things, in the right places, right now, and in the right way.

MEASURE Evaluation, funded by the United States Agency for International Development (USAID) and PEPFAR, works globally to improve health information systems that supply crucial data for making sound decisions on health policy and programs and also conducts research and evaluations on systems and programs that address health. The Measurement and Surveillance of HIV epidemics (MeSH) Consortium develops, tests and implements innovative and efficient methods for routine HIV measurement and surveillance among both adults and children.

Part of MEASURE Evaluation's work is providing technical assistance to countries to estimate the size of key populations

Before comparing two or more estimates, consider the definition of the population being estimated, the time frame, and the geographic scope for the size estimate:

- Who is included in or excluded from the population of interest?
- To what geographic area and time does the estimate refer?

The third reason estimates vary is that methods of data collection differ and size estimates will reflect the bias inherent in the method. Historically, size estimates have been calculated based on a variety of methods ranging from surveillance methods featuring population-based probability samples of the key population to population-based household surveys to expert opinion based on consensus. Increasingly, however, program data offer an additional opportunity for estimating the size of key populations that may have some unique advantages. Although program data are not population-based, program data may include more people than are interviewed in surveillance surveys and may include longitudinal data, offering additional insights. We frequently assume that most methods used to estimate the size of key population underestimate the size of key populations because they are often hidden and stigmatized. However, organizations that provide services to key populations may have an incentive to inflate size estimates for the sake of program funding. Comparing the validity and precision of various estimates requires careful study of potential sources of error and bias in the data collection methods. A review of all the size estimation methods used is beyond the scope of this tool. However, the appendixes give some information about how estimates were calculated in three specific cases.

The fourth reason size estimates can vary is the purpose of the size estimate. Size estimates used for program targets may be lower than size estimates used for program advocacy. Size estimates for program advocacy may be higher than estimates used for program targets. Program targets may choose to focus on the most high-risk subgroups or those most likely to be needing services. Advocacy may include the broadest possible scope for the size estimate in order to focus attention on the scale of the HIV epidemic among key populations.

The four questions listed in Box 1 may help interpretation of size estimates.

This tool was designed to help those engaged in program planning and target setting to understand why size estimates may differ and to make an informed judgment about which ones are most appropriate for use by their programs. The specific objectives and steps of the tool are outlined in the following section.

Box 1. Clarifying Questions to Aid in the Interpretation of Existing Population Size Estimates

1. **Who**. Who is included in the population? Each survey or program may use a slightly different definition of a key population, yielding different size estimates. A program's definition of a key population may be looser, embracing a wider variety of clients, or may include subgroups of the population. Programs may tailor service delivery differently to subgroups of a key population and want to monitor coverage in each subgroup, whereas a survey may not include questions to distinguish who belongs to each particular subgroup.

2. **Where**. What are the geographic boundaries of the population? Size estimates from surveys may not align with the catchment area of a program, which often doesn't cover an entire region or country. Thus they are often not "granular" enough for programs to use as denominators for coverage estimates.

3. **When**. To what point in time does the estimate refer? The method for calculating size estimates often fails to take mobility or turnover of key populations into account. Programs frequently need to target the number of people who can be reached over the period of a year, but cross-sectional survey estimates are taken at a single point in time.

4. **What**. What is the goal of the estimate? It may be intended to capture all members of the key population living in a geographic area at a given time, or only those who are considered reachable by programs. It may characterize the total size of the population or the numbers of people at each step in the prevention and treatment cascade.

WHAT ARE THE OBJECTIVES OF THE REVIEW AND REVISE TOOL?

This tool is meant to help those engaged in program planning and target setting (1) understand what population size estimates are needed, (2) identify currently available estimates, (3) understand the strengths and limitations of those estimates, and (4) reach a consensus on the most appropriate ones for the program to use. If currently available estimates are insufficient, the tool will help stakeholders understand how to revise them using existing data or (5) collect additional data to produce new estimates.

REVIEW

Step 1. Determine what population size estimates are needed.

Step 2. Identify available estimates for a given country or geographic area.

Step 3. Review the available estimates and determine whether they are sufficient, need improvement, or have important gaps.

REVISE

Step 4. Improve currently available estimates through adjustment methods without additional data collection.

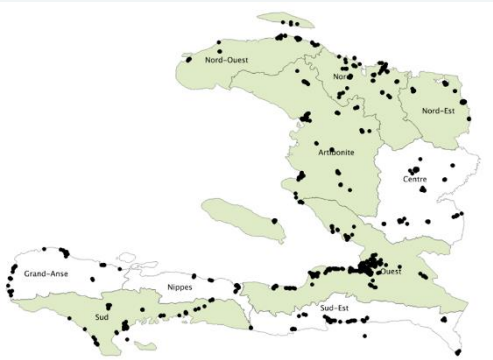
Step 5. Reach consensus on whether additional data collection is needed to improve the estimates.

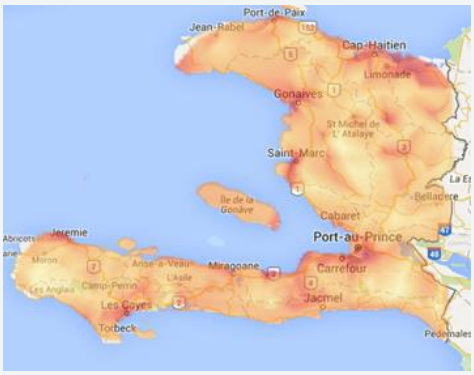
This tool was designed to facilitate consensus on size estimates for female sex workers (FSWs), men who have sex with men (MSM), and transgender people, although it could be adapted for other populations.

It has been tested in Guyana, Haiti, and the Dominican Republic. In each country, it was applied to assist with national target setting for HIV programs by the health ministry. Input was given by the National AIDS Programme Secretariat (NAPS); the Global Fund to Fight AIDS, Tuberculosis and Malaria (GFATM); UNAIDS and PEPFAR, in Guyana; the Programme National de Lutte contre les IST/Sida (PNLS) and PEPFAR, in Haiti; and Centro de Promoción y Solidaridad Humana (CEPROSH) and USAID, in the Dominican Republic.

Box 2. Size Estimate in Haiti

In 2015, USAID asked MEASURE Evaluation, based at the University of North Carolina at Chapel Hill, to conduct a review of all available key population size estimates in Haiti. On the basis of this review, stakeholders determined that the existing estimates were insufficient and commissioned additional data collection to generate new estimates in 2016.





Step 1: Determine What Population Size Estimates Are Needed

All size estimation methods estimate the number of people in a defined population, in a defined geographic area, during a specific time period. These estimates are used to set program targets, assess program coverage, and justify funding requests. The estimates needed for program planning and target setting will depend on the target population, the level of intervention, and any specific subgroups of interest. For example, a program for female sex workers in two urban districts will need estimates of the total population of FSWs in those two districts. Some programs may require estimates of specific subgroups, such as street-based or cross-border sex workers. After considering what estimates are needed, fill out Worksheet 1. Some options are shown in the table. The options shown are not exhaustive. An example for size estimates needed for FSW is shown. Complete one row per program population estimate needed.

Worksheet 1. What Population Size Estimates Are Needed?

	Program Population Needing Size Estimate	Geographic Area(s) for Estimate	Subgroups for Which Estimates are Needed
Options	FSWs; MSM; transgender women	National; capital city; all urban districts; cross-border areas, specific subnational areas	Street-based FSW; specific age groups;
Example	Female sex workers age 18 and older	Estimates needed for each District A, B & C	Separate estimates of FSW who are residents of the country vs. those who are from another country
Estimate 1			
Estimate 2			
Estimate 3			

Step 2: Identify Available Size Estimates

After considering what size estimates are needed, identify those that are currently available. Some places to look for them are listed in Box 3.

To Do:

1. Conduct a review of peer-reviewed and grey literature to identify any published key population size estimates. Also identify any available data sets from which estimates might be derived.
2. Talk with people engaged in service delivery in the country to obtain their estimates of the number of key populations in their catchment areas.
3. Talk with key population organizations to identify subgroups of key populations that may be missed and to obtain size estimates used by these organizations.
4. Using Worksheet 2, record the estimates you find. The worksheet can be expanded or duplicated for use with other key population subgroups or for smaller geographic regions. Describe the source of each estimate you find. Record the confidence bounds and note any reported limitations that may affect the validity or precision of the estimate.
5. For each estimate, consider any additional limitations that may not have been reported. You may note limitations related to the data collection and analysis methods or to the quality and completeness of the data collected. Technical Note 2 lists many common limitations and the situations in which they are likely to apply. Using that list as a guide, record any limitations that apply to the estimates you found in Worksheet 2.

Box 3. Where Can I Find Existing Size Estimates?

1. First, check the Key Populations Atlas, a web database maintained by UNAIDS. (You can also contribute data to the atlas.) http://www.aidsinfoonline.org/kpatlas/

2. Search for recent peer-reviewed and grey literature through library databases and search engines such as PubMed and Google Scholar.

3. Look for regional or global data on the percentage of the total population that belongs to key population groups. For example, female sex workers are thought to make up approximately 0.67% of the adult population ages 15–49 globally, according to estimates pooled by UNAIDS (Table 1). A regional or global estimate of the proportion of the total adult population that belongs to a particular key population can be used for a ballpark estimate that may be adequate for program purposes. This method can quickly estimate population size when estimates derived from surveys are of poor quality or unavailable.

4. In some cases, even if no estimates are available, you may be able to use existing data sets to estimate your population of interest. A data set used for this purpose must contain all the information about behavior and identity that is needed to classify individuals as members of the population of interest. This requirement rules out the many demographic and health surveys that don't collect this information. See Appendix 2 for an example of how key population size estimates were calculated using previously collected program data in Haiti.

Worksheet 2. Available Size Estimates for [Key Population], [Country]

		Available Estimate 1	Available Estimate 2	Available Estimate 3
Step 2	Size estimate			
	Confidence bounds			
	Reference, source and year			
	Inclusion criteria			
	Exclusion criteria			
	Sampling method			
	Sample size			
	Size estimation method			
	Geographic area covered by the estimate			
Step 3	Rate your assessment of the quality / adequacy of each the available size estimates. Use questions in Box 4.	☐ High quality ☐ Good enough ☐ Low quality or unknown	☐ High quality ☐ Good enough ☐ Low quality or unknown	☐ High quality ☐ Good enough ☐ Low quality or unknown
Step 4	Are the available size estimates sufficient for program needs?	☐ Yes ☐ No	☐ Yes ☐ No	☐ Yes ☐ No

Table 1. Proportions of key populations among adult population (15–49)

		UNAIDS regions							Total
		Asia and Pacific	Caribbean	East and Southern Africa	East Europe and Central Asia	Latin America	North Africa and Middle East	West and Central Africa	
Female sex workers	# countries	12	2	7	9	5	3	11	47
	Median	0.35		0.58	0.68	0.49	1.18	1.19	0.67
	range	0.18-2.33	2.40-2.50	0.25-3.00	0.42-1.25	0.12-1.91	0.94-2.78	0.57-3.00	0.12-3.00
	25-75 percentiles	0.26-0.67		0.41-1.66	0.46-0.97	0.16-1.52		0.84-1.88	0.39-1.23
Men who have sex with men	# countries	14	4	9	11	9	3	12	62
	Median	1.69	2.71	1.28	1.39	1.49	0.90	0.72	1.22
	Range	0.09-4.06	0.40-5.00	0.16-3.20	0.18-4.00	0.59-5.47	0.61-2.28	0.05-2.00	0.05-5.47
	25-75 percentiles	0.26-3.0		0.25-1.85	0.62-2.04	0.99-3.30		0.34-1.38	0.50-2.00
People who inject drugs	# countries	10		7	8		1	3	30
	Median	0.06		0.60	2.18		0.06	0.10	0.11
	range	0.001-1.04		0.004-2.72	0.47-3.30			0.05-1.60	0.001-3.30
	25-75 percentiles	0.03-0.16		0.03-1.58	1.11-2.97				0.05-1.41
Transgender people	# countries	3	1	1		5			10
	Median	0.02	0.38	0.31		0.12			0.11
	range	0.02-0.06				0.03-0.42			0.02-0.42
	25-75 percentiles					0.06-0.18			0.04-0.25

Unpublished data: UNAIDS, GFATM, WHO. Source: Keith Sabin, presentation at United States Centers for Disease Control and Prevention

Step 3: Review Available Estimates

The next step is to review available size estimates in terms of quality and usefulness and determine, for each estimate that is needed, if the available estimates are:

- Sufficient
- Need improvement or
- Have important gaps that require additional data collection.

Some size estimates, such as those based on conjecture or outdated data, should be treated with a hefty dose of skepticism. Others are reasonable and can be used with confidence. In this step you will review estimates that are available to assess whether they are good enough or whether improvement or additional data collection is required. Use the questions in Box 4 to review each of the size estimates documented in Worksheet 2. The questions are written so that a "Yes" answer reflects a positive attribute of a size estimate. The number of "No" or "I don't know" responses should guide your assessment of whether the size estimate may have problems.

If multiple well-documented estimates are available that pass the bias review (based on the questions in Box 4), there can still be problems in reaching consensus on a size estimate for a particular population in a particular location if there are multiple estimates that substantially differ. If two or more size estimates of similar populations in the same geographic area are within +/- 15% of each other, they are probably not different enough to warrant significant additional resources to improve. However, even if the estimates appear to differ, you should first consider whether they in fact do. Sometimes apparent differences can be attributed to imprecision or to random error in the underlying data. In other cases, they may arise from how the population was defined. Such situations are described in greater detail below.

1. **Lack of precision:** When size estimates from two data sources have overlapping confidence intervals, we cannot say with certainty that they are different. Typically, very wide confidence intervals are the result of small sample sizes or imprecise estimation methods. When two estimates have wide confidence intervals that overlap, they cannot be statistically distinguished. They can be compared only by judging the quality of the underlying data and the methods used to derive them.
2. **Differing definitions:** When two data sources define the population of interest differently, comparing the resulting size estimates is like comparing apples and oranges. For example, one survey may define female sex workers as women who have received money in exchange for sex in the past six months, whereas another may define them as women who have received money, goods, or favors in exchange for sex in the past twelve months. Sometimes, with access to the underlying data, estimates can be appropriately manipulated to make them more comparable.

Box 4. Bias Checklist: Is a Size Estimate Fairly Recent, Well Documented, and Free of Major Bias?

Use the number of "NO" responses to assess the quality of the size estimate—the higher the number of "NO" responses, the more likely it is that additional analysis or additional data collection might be needed to obtain adequate estimates. NA= Not Applicable.

1.	Is the definition of the key population sufficiently detailed? Was the determination of key population membership operationalized using methods that had been pretested and deemed reasonable? (WHO)	YES	NO	NA
2.	Is the geographic scope of the estimate specified? (WHERE)	YES	NO	NA
3.	Is the estimate based on data collected within the past 5 years? (RECENT)	YES	NO	NA
4.	If a survey was used, were members of the key population engaged in its implementation and design? (ENGAGEMENT)	YES	NO	NA
5.	Was the estimate based on an analysis of empiric evidence collected using standard research methods (rather than based on opinion, conjecture, or a convenience sample of participants)? (EMPIRIC)	YES	NO	NA
6.	Is there sufficient documentation of the method used to calculate the estimate? Could someone independently use the methods with the data and obtain the same estimates? Are the data (with personal identifiers removed) and data collection forms available to the public? (DOCUMENTATION)	YES	NO	NA
7.	If the estimate is based on a survey, were at least 200 people in the key population interviewed using a probability sampling method? (PRECISION)	YES	NO	NA

8.	If a size estimate from one geographic area was extrapolated to others, was the method used to extrapolate well described and justified? (EXTRAPOLATION)	YES NO NA
9.	Is the estimate similar to previously accepted estimates? Is the estimate roughly in line with UNAIDS estimates of the percentage of the population who is a member of a key population? (FACE VALIDITY)	YES NO NA
10.	Were the estimates published in a peer review journal? (PUBLICATION)	YES NO NA

Box 5 lists some basic criteria for deciding whether estimates do in fact differ from each other.

To Do:

1. On the basis of the information you collected in Worksheet 2, eliminate any size estimates that are likely to be invalid, using the criteria outlined in Box 3.
2. Evaluate whether the remaining estimates differ, using the criteria in Box 4.

Box 5. Do the Size Estimates Truly Differ?

1. If confidence bounds were reported around the estimates, it is easy to tell how much they differ. If the confidence bounds overlap, the difference between estimates is statistically insignificant.

2. If no confidence bounds were reported, you may be able to derive them from the underlying data. You should consult a statistician to determine whether this is feasible.

3. Check the inclusion and exclusion criteria for each estimate to determine how the population was defined. If you find explicit differences, you may be able to adjust the estimates to reflect the same population of interest. For example, an estimate of female sex workers ages 15–44 may be adjusted using demographic data to give a size estimate for FSWs ages 15–49, assuming that the proportion of FSWs is the same among women ages 15–44 and 45–49.

4. In some cases, you may not be able to tell whether two or more estimates differ.

When estimates differ substantially, you should weigh them according to their strengths and weaknesses. Technical Note 1 lists some of the most common ones. Although outside the scope of this tool, you may also want to consider the assumptions and potential biases of the methods used to calculate the size estimates. Finally, you may want to consider which estimates stakeholders deem most reasonable on the basis of their knowledge and experience.

Even when available estimates are imperfect, they can be useful for program planning. You may decide that despite their limitations, they are sufficient to meet the program's current needs. For example, if the objective is to estimate how many condoms to procure for distribution to all female sex workers in a country, a ballpark estimate of the total number of FSWs may be sufficient. But if the objective is to decide how best to distribute those condoms by region, a more granular estimate is needed. This decision should be made in consultation with local and national stakeholders working with the population of interest.

To Do:

1. Referring to Technical Note 1, consider the strengths and weaknesses of each estimate remaining in Worksheet 1.
2. Indicate your confidence in the quality of each of the estimates in the penultimate row of Worksheet 1.
3. Considering your confidence in their validity, reach consensus on whether the remaining estimates are sufficient for program needs and fill out the last row of Worksheet 1.

Next Steps

- ➔ If any boxes are checked ***Yes*** in the last row of Worksheet 1: Work with stakeholders to reach consensus on which estimate you will use. This process could take the form of a meeting to review the information you have collected thus far and weigh it together with the judgment of experts.
- ➔ If all boxes are checked ***No*** in the last row of Worksheet 1: Fill out Worksheet 2.

In consultation with stakeholders, consider the current needs of the program and fill out the following table. What improvements to available estimates are needed in each of the following categories?

Worksheet 3. Gaps in Available Size Estimates

	Additional Information Needed
Who *e.g., subgroup estimates*	
Where *e.g., regional estimates*	
When *e.g., estimates that account for population turnover*	
What *e.g., size estimates by step of the treatment cascade*	

Next Steps

After completing the above table, consider whether the necessary improvements can be made using data that already exist. Use your best judgment or consult an expert to answer this question.

➔ Sometimes population size estimates can be improved simply by using information from an existing data source, such as demographic data or program data. If you think this may be the case, proceed to Step 5.

➔ In some cases, you simply won't have enough information to calculate valid estimates with existing data. If so, you should proceed to Step 6.

Step 4: Improve Available Size Estimates Through Adjustment

When you are confident that you have assessed the quality of available estimates and judged they are insufficient to meet the program's current needs, you may be able to improve them without collecting additional data. Perhaps, for example, you have estimates for a few geographic regions of the country and you need some for additional regions. Or perhaps you have a national estimate of all female sex workers and you need provincial estimates of street-based FSWs.

Many methods exist for adjusting size estimates, and it is outside the scope of this tool to describe them all in detail. The appendixes give examples of how they were adjusted in three specific cases:

- In Guyana, estimates from program data were adjusted to account for duplication and to estimate the number of people not reached by the program.
- In Haiti, estimates from survey data were adjusted to account for underreporting of hidden populations and for individual site-visiting behavior.
- In the Dominican Republic, additional data were collected to extrapolate subnational size estimates to additional subnational areas. Adjustment through additional data collection is discussed in Step 6.

These are a few of the most common adjustment methods:

1. Predictive modeling can be used to extrapolate subnational size estimates to the national level. Models can also be used to interpolate size estimates for subnational areas where data were not

collected. The variables in the model are characteristics of the subnational areas, such as socio-demographic ones, that are thought to be associated with the size of the population.

2. Inverse probability weighting applies sampling probabilities to individual participants' survey responses to weight responses up to the population level; this method can be used when individuals are asked to estimate the number of people in a given geographic area.
3. Size estimates can also be adjusted with information collected from population members themselves. This information might be about mobility or other behaviors that contribute to the ebb and flow of the population over time and through space. This type of adjustment helps to improve the precision of estimates by triangulating information from multiple sources.
4. When estimates are derived from program data, it is often necessary to account for overlap between multiple data sources, such as several program sites. It is also necessary to adjust for duplication, which occurs when individual beneficiaries are double counted in the program data. If the program has a system for assigning unique identifiers to its beneficiaries, this adjustment is relatively simple. But if it does not, you may have to make some assumptions about the amount of duplication.

See the appendixes for more about how these adjustment methods were put into practice.

Step 5: Reach Consensus on Whether Additional Data Collection Is Needed

Sometimes funding is available for additional data collection to generate new estimates. Two main approaches can be taken.

(a) Limited data collection to adjust available size estimates. Typically, this is done to generalize available estimates to a different geographic area or a population other than that originally covered by data collection. For example, if an estimate of the size of the FSW population is available from a program that serves women in a certain geographic area, a survey could be conducted to determine what proportion of the women in that area are clients of the program. If it can reasonably be assumed that there is no differential selection of FSWs in the program, the original estimate can be multiplied up to the population level. Another example is described in Appendix 3: in the Dominican Republic, a survey was conducted in 2016 to augment data from a much larger survey conducted in 2014.

(b) A new survey to generate new estimates. Any of the survey methods described in Technical Note 1 can be used to collect the information needed. This option requires more time and funding, but it may be preferable if existing estimates cannot be adjusted confidently. That may be the case if there are many methodological issues with how the data were collected for those estimates. An example of this is given in Appendix 1: in Haiti, existing data were insufficient to calculate key population size estimates, so stakeholders decided to implement a new Priorities for Local AIDS Control Efforts (PLACE) study in 2016.

If stakeholders decide that additional data collection is needed, its type, scope, and scale will depend on what data are already available and what gaps were identified in Worksheet 2.

TECHNICAL NOTE 1: BIAS ASSESSMENT TOOL

A given size estimate may inaccurately represent the size of a population for two reasons: Random error, or bias—also known as systematic error. This technical note briefly defines the two and lists several types of bias that are common in studies used to generate population size estimates.[1]

No matter what method we use to estimate population size, we are essentially collecting information about a group of individuals and using it to make an inference about the population they represent. The process is inherently subject to random error. In other words, the estimate will never be accurate to the last individual, because not all the individuals in the population were actually counted.

The amount of random error, or imprecision, that exists in an estimate depends on several factors. Perhaps the most important is sample size. The larger the sample size for a study, the more precise the resulting estimate will be: An estimate based on surveying 1,000 individuals will be more precise than one based on surveying only 100. However, some small amount of random error is inevitable in any estimate.

The amount of random error is captured in the confidence interval around that estimate, which takes into account sample size and other important elements of the study design. Many studies report confidence intervals around population size estimates. Sometimes it is inappropriate to calculate confidence intervals, but sensitivity analysis can be performed to see how much an estimate will change as the parameters used to calculate it vary. This type of analysis can be very useful when the data underlying estimates are sparse.

Even when a study is very large and the amount of random error is small, the estimate is subject to bias. A biased estimate can have a narrow confidence interval, appear to be precise, but be completely inaccurate. The amount of bias in an estimate depends on the methods used to collect and analyze the underlying data, the quality of the data collection, and the appropriateness of the analysis. All study designs and methods of calculating size estimates are prone to various types of bias. Many of the most commonly encountered types are listed in Table 2. Before reviewing the table, you may want to familiarize yourself with the types of study design most commonly used to generate population size estimates.

[1] For further reading on this topic, consult UNAIDS. *Guidelines on Estimating the Size of Populations Most at Risk to HIV; 2010* (http://www.unaids.org/sites/default/files/media_asset/2011_Estimating_Populations_en_0.pdf) and WHO, CDC, UNAIDS, FHI 360. *Biobehavioral Survey Guidelines for Populations at Risk for HIV.* Geneva: World Health Organization; 2017.

Box 5. Population Data Sources

The data used to generate population size estimates typically come from one of two sources: survey data or program data.

SURVEY DATA

Survey data is distinct from program data in that surveys typically employ a sampling strategy that allows results to be generalized to the population level. The following three sampling strategies are most commonly used in surveys that collect information on key populations.

- **Venue-Based Sampling:** This strategy is used in programmatic mapping studies, which estimate the number of people who can be reached in the public places where key populations socialize or congregate. These are typically high-risk places, such as bars where sex workers meet clients or venues where people inject drugs. Size estimates based on programmatic mapping omit people who do not visit these locations and often rely on a bar manager or other venue informant to estimate the number of key populations who visit the venue. Estimates based on programmatic mapping are typically smaller than other size estimates but may be higher if adjustment for double-counting people who visit multiple venues is not taken into account. Strategies to improve the validity of size estimates from venue-based sampling are available.

- **Respondent-Driven Sampling:** This strategy relies on social networks to recruit a sample of study participants. A few initial participants, called seeds, are asked to invite a predetermined number of other people to participate in the study. Researchers use analysis methods such as the multiplier method or capture-recapture to generate size estimates based on the number of respondents referred to participate. This sampling strategy may help reach populations that are hidden and less likely to be included in venue-based or household surveys. However, it relies on assumptions that may be difficult to meet in typical field conditions and requires participants to self-identify as a member of the population. Strategies to improve the validity of size estimates from respondent-driven sampling are available.

- **Population-based Household Sampling:** Stratified sampling of household clusters is the strategy most commonly used in national household surveys. Its advantage is that the resulting sample is representative of the population at both the national and the regional level. Its disadvantage is that obtaining adequate sample sizes of small or hidden population subgroups, such as men who have sex with men, can be difficult. Key populations may also be under-represented in the household population. Consequently, surveys that use this strategy typically do not include questions to identify individuals as members of these subgroups.

PROGRAM DATA

Data on the numbers of key populations reached by programs may be available at the national, regional, or local level. The main advantage of routine program data is that it does not require the cost of a research study or population-based survey. Its main disadvantage is that inferences drawn from program data cannot be generalized to the population. The individuals captured in program data may be at higher risk than, or have other characteristics that distinguish them from, the rest of the population. As a result, size estimates drawn from program data can be generalized only to the population of individuals reached by the program.

Often additional data can be collected to make estimates more representative of the population. This is discussed further in Step 5.

The table below lists several common sources of bias in studies used to generate population size estimates and the relative susceptibility of venue-based sampling, respondent-driven sampling, and routine program data to the bias.

Bias Assessment List

Potential Sources of Bias or Error in Size Estimates	Examples	If present, size estimate is likely to be:
If method is venue-based and population does not attend venues	• FSWs who work out of their homes are not counted at venues • MSM who only socialize at private homes are not counted at venues	Too low
Individuals are double counted	• The same person appears in routine program data for two different programs • The same person is counted at two different venues	Too high
Misclassification or incorrect classification of individuals to the	• A person younger than 18 is classified as a sex worker when the minimum sex worker age is 18	Too high

Potential Sources of Bias or Error in Size Estimates	Examples	If present, size estimate is likely to be:
population of interest when they are not a member of the population	• A person is classified as a sex worker by a program even though she quit sex work • Transgender women are included as men who have sex with men	
Misclassification of individuals as not belonging to the population when they are a member of the population	• Women who work as exotic dancers and exchange sex for money are not included as sex workers	Too low
Failing to account for population turnover and new persons joining the population	• Female sex workers work in a city for 6 months and then move to another city. The number in the city over the course of a year is higher than the number in any given month.	Too low
Response bias when asking questions about stigmatized behaviors	• Respondents deny behaviors such as sex work or having anal sex with a man	Too low
Missing subgroups of the population that are not connected or linked to the main group	• Female sex workers who are in a country illegally may live and work separately from other sex workers.	Too low
Failing to appropriately extrapolate available estimates to geographic areas without estimates	• Extrapolation may be based on outdated census data and not take areas experiencing high population growth into account.	Too low or too high
Failing to appropriately translate sensitive questions into words that are understandable and acceptable to the population	• People may not understand questions about anal sex.	Too low

These sources of bias may apply to a given study in many ways. Consult an epidemiologist if you are unsure which types of bias apply to a specific study or how they may affect the resulting size estimates. An expert may be able to perform sensitivity analysis to determine whether and how much various sources of bias are expected to affect the estimates.

APPENDIX. GUYANA CASE STUDY: COMPARING VENUE-BASED SAMPLING DATA WITH PROGRAM DATA

The National AIDS Programme Secretariat (NAPS) in Guyana has standardized reporting tools for all key population outreach as defined in 2012 in the Most-At-Risk Populations (MARPs) Guidelines and Standards for Non-Governmental Organizations. All nongovernmental organizations working on prevention outreach for key populations use the reporting tools provided by NAPS. The peer navigators and outreach workers handwrite their activities (persons reached, products distributed, messages given, etc.) in the Peer Educator Diary (see below) when conducting outreach at a specific venue or location. The NGOs compile this information in the Prevention Log for Interventions with Targeted Populations to count total numbers of key populations (KPs) reached at each place visited (see below). These completed logs are provided to NAPS and help identify where KPs are being reached and how many at each venue. Additionally, NGOs complete a list of client codes that they have consolidated, de-duplicated, and provided to NAPS. Peer Educator Diary (see below) provides unique identifiers for each person who received outreach. Guyana has implemented the unique identifier system, and it is currently being used both to identify clients during outreach activities and to link them to testing and care.

Prevention Log for Interventions with Targeted Populations

	NAME OF PLACE VISITED	DATE	Number of KPs & PPs Reached																										
			Miners			Loggers			Adjacent Population			MSM			FSWs			Clients of KPs			Other Persons			Summary			Refer		
		dd/mm/yy																Males											
			N	O	Total	N	O	Total	N	O	Total	N	O	Total	N	O	Total	N	O	Total	N	O	Total	N	O	Total	STI	VCT	
1					0			0			0			0			0			0			0	0	0	0			
2					0			0			0			0			0			0									
3					0			0			0			0															

PEER EDUCATOR DIARY - MSM [ ] FSW [ ] OUTREACH

Name of Peer Ed:
Date:
Location:
Client Code(s):

Persons Reached									
MSM		Transgender		FSWs		Clients		Other Persons	
New	Old	New	Old	New	Old	New	Old	New	Old

Condom Distribution									
MSM		Transgender		FSWs		Clients		Other Persons	
M	F	M	F	M	F	M	F	M	F

Outreach Topic:
Life Skills Topic:

Lubricant Distribution				
MSM	Transgender	FSWs	Clients	Other Persons
# of tubes	# of tubes	# of tubes	# of tubes	# of tubes

Risk Reduction Topic:

Number of BCC Materials Distributed:

Referrals:

Services Covered in Outreach: (Please Tick)

HIV Information

Both PEPFAR and GFATM fund outreach activities in Guyana. Although all NGOs follow the national policies outlined in the MARPs guidelines, reports are submitted separately to the various funding agencies and consolidated by NAPS.

Findings

- Guyana has a well-defined reporting system with consistency across the AIDS program.
- The program data capture unique individuals reached by implementing client codes.
- Reporting is separate for PEPFAR and for GFATM.

Review of Most Recent Data

To compare population size estimates with or develop estimates from the program data, the team reviewed the most recent data available and used multiple sources in an effort to make it complete. The team identified and attempted to solve the following issues:

Delays in Reporting/Incomplete Data

The team wanted to refer to the most recent data available. Data available to NAPS differed according to the partner and the funding agency. NAPS had received consolidated reports from GFATM sub-recipients for the previous six months but no consolidated reports from PEPFAR partners. This prevented comparison between partners and a totaling of all partners in the previous six months.

Additionally, GFATM partners submitted complete lists of client codes to NAPS, whereas PEPFAR partners did not. This prevented a comparison of duplicated client codes for every month but one.

Different Reporting Periods

Partners are required to submit reports to the donor, whether GFATM or PEPFAR, in keeping with the donor's reporting period. GFATM partners submit reports semi-annually, and data are compiled for those six-month periods. PEPFAR partners are required to submit reports quarterly, and data are compiled according to the U.S Government fiscal year (October–September). Semi-annual reports are combined to create an annual report for GFATM that follows the calendar year. This made it difficult to compare GFATM and PEPFAR figures across the same time period, particularly when monthly disaggregation was unavailable.

Overlap and Duplication

GFATM and PEPFAR worked in the same regions in early 2015, and sometimes with the same NGOs, to implement outreach. Because key populations are mobile, access outreach and services in multiple regions, and can be found at multiple outreach venues, there is naturally some overlap between clients reported by PEPFAR and clients reported by GFATM. GFATM client lists are de-duplicated at the NGO, regional, and national levels to remove multiple instances of the same client code. Because PEPFAR NGOs did not submit their client lists to NAPS, we were unable to compare the annual overlap between the clients reached by the two donors' sub-recipients.

Changes in a Program

Program data do not show consistent outreach achievement over time. Large changes in donor-funded programs in the past year are reflected in regional and monthly achievements. These changes affect the ability to generalize about the size of populations or reachable populations.

Understanding Data Quality

The team held a discussion about program data strengths and weaknesses with the NGOs, the National AIDS Programme, and the donors. The discussion identified several issues that could potentially lead to misreporting in their outreach figures.

Definitions of Key Populations

NGOs described various methods for identifying key populations during their outreach activities. Some NGOs mentioned using specific behavioral screening questions to identify key populations. Others described a less formal identification process based on conversation and peer educator knowledge and experience. Peer educators may be failing to apply national guidelines' strict definition of key populations in their outreach. Similarly, although client lists should include only those who are 18 years or older, NGOs mentioned that they sometimes work with younger key population members.

Data Entry and Aggregation

Peer educators, NGO M&E officers, and reporting officers are liable to commit clerical and data entry errors.

Double Counting

All NGOs reported that clients sometimes change the personal information they share, thereby creating new client codes with each visit. The client codes of key populations may also change according to which venues or NGOs they visit, leading to double counting of individuals both within NGOs and across them.

Unique Identifiers

The NGOs were pleased with the implementation of client codes and felt that they were well received. They mentioned, however, that similar initial and birthday combinations among their clients, along with individuals' changing their personal information, created the potential for duplicated client codes.

Adjusting Program Data for Size Estimation

With highly complete data and the NGOs' general confidence in it, the team considered how program data could be adjusted to account for data quality issues and be used to work toward reasonable size estimates of key populations. The numbers reported were considered high, representing larger than expected proportions of the general population when compared with census data. It was acknowledged that program data account only for key populations reached, and the NGOs have yet to reach everyone. The data will need to be adjusted upward for those not reached.

Accounting for Duplication

Duplication of clients between PEPFAR and GFATM was assumed, because the two donors' lists had never been merged and de-duplicated; PEPFAR client lists were not shared previously with NAPS. During this review, PEPFAR's client lists for Region 4 were made available for October and November but no other months. These data were compared with GFATM client list data for October, the only month available. The process for comparing was as follows:

- PEPFAR partners reported 308 unique FSWs in November for Region 4.
- GFATM sub-recipients reported 108 unique FSWs in November for Region 4.
- PEPFAR and GFATM figures were combined on one list with 416 client codes.
- The number of people on both lists for Region 4 was determined through a "remove duplicates" command; 13 records were duplicates.
- The percentage of people on the GFATM list in Region 4 that were also on the PEFFAR list was estimated (13 out of 108, or 12.0%).
- The same percentage of overlap in Regions 2, 6, and 10 was assumed: 12.0%.
- For GFATM (January through November reports were available), 4,396 unduplicated FSWs were reported. For the PEPFAR FY15 (October–September), the total was 3,400.
- The total number of duplications between PEPFAR and GFATM was estimated by applying the monthly percentage of duplicates to all the regions where PEPFAR and GFATM overlapped. GFATM totals were 180 in Region 2; 1,190 in Region 4; 1,094 in Region 6; and 826 in Region 10, for a grand total of 3,290. 12.0% of that total is 395.

- The number of estimated duplicates was subtracted from the total of GFATM and PEPFAR clients reported: 4,396 + 3,400 – 395 = **7,401.**

This same process was applied to MSM. The removal of duplications did not bring down the program totals as much as anticipated.

Estimating the Proportion Not Reached

Another adjustment considered was identifying the proportion of key populations not reached and not represented in the reports submitted to NAPS. To reach a proper size estimate, the team needed to calculate the number reached plus the number not reached. To identify the latter number, the team considered using venue-specific outreach data about key populations, which NAPS receives from each venue.

The NGOs were asked to review the site list used for site verification in the Biobehavioral Surveillance Survey (BBSS). They were asked to report whether the site was closed or unvisited and, if visited, what key populations were found there. This data could have been used to account for key populations not reached. But the review was not completed in favor of focusing on the survey data.

Expert Opinion on Overlap and Double Counting

The NGOs were very proud of and confident in their program data, but they also recognized overlap with other programs and double counting when people change their client codes. One option was to reduce each region's program data by a percentage that reflected the NGOs' assessment of the amount of overlap and double counting. The NGOs suggested reducing the figure by as much as 35% in Region 4 and by 10% in other regions, with perhaps a higher rate in the hinterlands, where populations are more mobile. This option was considered but eventually discarded for lack of evidence to justify the reductions.

RESOLVE

Do the Size Estimates from Program and Survey Data Differ?

Program data and survey data were compared to find differences between estimates for all existing reachable key populations (study size estimates) and key population members reached (program data). The program data consisted of total key population members reached in 2015 as reported by the NGOs. An adjustment to the program data to account for duplication between PEPFAR and GFATM figures could be made but has not been done in the tables below.

Table 2. FSW program data compared with survey data

	Global Fund (Jan–Oct) 2015	PEPFAR 2015	Total GF and PEPFAR 2015	BBSS Size Estimates, with Weight Adjustment
Region 1	58	170	228	575
Region 2	180	106	286	47
Region 3	374		374	539
Region 4	1,191	1,542	2,733	950
Region 5	186		186	132
Region 6	1,094	753	1,847	646
Region 7	327	354	681	792
Region 8	122		122	34
Region 9	39		39	0
Region 10	826	374	1,200	96
Total	**4,397**	**3,299**	**7,696**	**3,811**

Note: Empty cells mean data are not available for those regions.

Table 3. MSM program data compared with survey data

	Global Fund (Jan–Oct) 2015	PEPFAR 2015	Total GF and PEPFAR 2015	BBSS Size Estimates, with Weight Adjustment
Region 1		9	9	29
Region 2	79	0	79	90
Region 3	486		486	130
Region 4	496	2,350	2,846	1,361
Region 5	203		203	37
Region 6	1,235	352	1,587	555
Region 7		77	77	164
Region 8			0	48
Region 9			0	0
Region 10	425	247	672	51
Total	**2,924**	**3,035**	**5,959**	**2,464**

Stakeholders and implementers in general perceived that the program data numbers were too high (particularly in comparison with proportions found in census data) and the BBSS size estimates were too low. The fact that program data are compiled over the course of a year and survey data come from point-in-time estimates explained some of the difference.

Three Options for Determining Appropriate Size Estimates

1. **Survey based.** Use size estimates from the BBSS as the foundation and distribute the number across regions using a justifiable and well-documented approach.
2. **Program data based.** Use estimates from the program data and add the estimated proportion not reached by the program.
3. **Proportion of population**. Use the census to estimate what percentage of the general population are members of a key population.

After discussion with stakeholders and NGOs, it was agreed that survey data made the most sense.

Analysis to Improve the Size Estimates from the Survey

Female Sex Worker Estimates

After stakeholder discussions with national experts, donors, and implementers, this was the proposal agreed upon:

- Use survey data as the base estimate
- Trim weights in the top 1 percent (weights > 50=50)
- Ceiling for each region is 5 percent of population
- Floor for each region is 1 percent of population
- Allocate the shortfall that results from trimming the weights and applying the ceiling across the other regions proportionally

The proposal was discussed at length with donors, program representatives, and national program representatives. The table below shows the regional estimates with the weights trimmed and the 5 percent ceiling and 1 percent floor applied.

It was acknowledged that an estimate of the size of each key population over the course of a year—not just at one point in time (the size estimates reported in the BBSS)—would be necessary. This annual estimate would be useful for planning activities and setting targets for the national program. To create an annual size estimate, the figures were adjusted to account for FSWs who are new to venues. Using the SAS code in the box here, it was found that approximately 4.6 percent of FSWs met the criteria for a new person at the venue. The criteria were:

- FSW meets the definition
- FSW reports that this is the first time at the venue
- FSW reports that she has not gone to other venues that night
- FSW reports that she has not gone to other bars and clubs in the past four weeks

SAS Code to Determine Percentage of New FSW

```
if fsw=1 and q535=7 /*first time at site*/ and q533a in (0,1) /* no other sites visited today */  and q418=4  /* no other bars or clubs visited in past 4 weeks */then fswnewtosite=1
```

We assumed that 4.6 percent of the population was new to the site. We multiplied that by three to get the estimate for the year. The total number of FSWs over the course of a year is 5,256.

Table 4. FSW size estimates and adjustments

Region	Regional Estimates with Extreme Weights Trimmed	5% Ceiling and 1% Floor Applied	New FSW Added to the Estimate ANNUAL
1	575	202	278
2	47	65	89
3	539	749	1,033
4	950	1,320	1,821
5	132	183	252
6	646	898	1,239
7	792	179	247
8	34	47	64
9	0	33	45
10	96	133	183
Total	**3,811**	**3,811**	**5,256**

Females	2002 Census Population Ages 18–49	1% of Census Population	5% of Census Population	Which Estimates Subject to Ceiling and Floor
Region 1	4,043	40	202	575 => 202
Region 2	11,012	110	551	47
Region 3	25,219	252	1,261	539
Region 4	84,971	850	4,249	950
Region 5	7,430	74	372	132
Region 6	29,797	298	1,490	646
Region 7	3,574	36	179	792 => 179
Region 8	1,727	17	86	34
Region 9	3,327	33	166	0 => 33
Region 10	9,404	94	470	96
Total	180,504	1,805	9,025	3,811

Men Who Have Sex with Men Population Size Estimates

The survey data was also compared with census data but applying population proportion floor or ceiling levels was deemed unnecessary. No single regional size estimate was higher than 4 percent of the male population. The comparison we reviewed (with original study size estimates) is shown below:

Males	2002 Census Population Ages 18–49	1.0%	1.5%	2.0%	2.5%	3.0%	3.5%	4.0%	4.5%	5.0%	Original BBSS Study Size Estimates
Region 1	4,918	49	74	98	123	148	172	197	221	246	29
Region 2	11,187	112	168	224	280	336	392	447	503	559	90
Region 3	25,645	256	385	513	641	769	898	1,026	1,154	1,282	130
Region 4	74,517	745	1,118	1,490	1,863	2,236	2,608	2,981	3,353	3,726	1,361
Region 5	12,152	122	182	243	304	365	425	486	547	608	37
Region 6	30,182	302	453	604	755	905	1,056	1,207	1,358	1,509	555
Region 7	4,455	45	67	89	111	134	156	178	200	223	164
Region 8	2,788	28	42	56	70	84	98	112	125	139	48
Region 9	3,710	37	56	74	93	111	130	148	167	186	0
Region 10	9,492	95	142	190	237	285	332	380	427	475	51
Total	**179,046**	**1,790**	**2,686**	**3,581**	**4,476**	**5,371**	**6,267**	**7,162**	**8,057**	**8,952**	**2,464**

The recommended one-year estimate for MSM is the last column in the table below. It is similar to the FSW one-year estimate. Like the annual adjustment to FSW population size estimates, it estimates the percentage of MSM reporting that this is their first time at the venue and that they have not been to other bars or clubs in the past four weeks. The 4.6 percent is multiplied by three to take into account that this number may reflect new MSM at the venue over a three-month period.

The second column reflects another scenario using a methodology to estimate MSM who are newly sexually active, but the former is the preferred methodology.

Table 5. MSM size estimates and adjustments

Region	Regional Estimates with Weight Adjustment	Adding Young MSM for a Year Estimate	Adding New MSM to the Estimate ANNUAL
1	29	29	38
2	90	119	122
3	130	130	176
4	1,361	2,244	1,837
5	37	66	50
6	555	585	749
7	164	164	221
8	48	48	64
9	0	0	0
10	51	66	69
Total	**2,464**	**3,450**	**3,327**

Final Population Size Estimates

The final estimates were presented to stakeholders and agreed upon with the stipulation that they will continue to be refined. These figures represent the annual number of people in each key population who can be found at venues.

Region	FSWs	MSM
1	278	38
2	89	122
3	1,033	176
4	1,821	1,837
5	252	50
6	1,239	749
7	247	221
8	64	64
9	45	0
10	183	69
Total	**5,256**	**3,327**

Next Steps

Population size estimation is ongoing as new and better data become available. The estimates agreed upon will continue to be refined. Potential next steps include:

- Conduct a validity study. Re-interview at a sample of venues using unique identifier codes (capture-recapture approach).
- Continue to improve program data and monitor numbers reached in comparison with the size estimates.
- Plan for an additional size estimation study in the next several years.

MEASURE Evaluation
University of North Carolina at Chapel Hill
123 West Franklin Street, Suite 330
Chapel Hill, North Carolina 27516
Phone: +1 919-445-9350
measure@unc.edu
www.measureevaluation.org

This publication was produced with the support of the United States Agency for International Development (USAID) under the terms of the MEASURE Evaluation cooperative agreement AID-OAA-L-14-00004. MEASURE Evaluation is implemented by the Carolina Population Center, University of North Carolina at Chapel Hill in partnership with ICF International; John Snow, Inc.; Management Sciences for Health; Palladium; and Tulane University. Views expressed are not necessarily those of USAID or the United States government. TR-18-274

USAID
FROM THE AMERICAN PEOPLE

PEPFAR

MeSH Consortium
Measurement & Surveillance of HIV Epidemics

MEASURE Evaluation