Quantity Versus Quality: A Survey Experiment to Improve the Network Scale-up Method – Web Material

Web Appendix 1

In this section, we describe our strategy for combining two arms of a survey experiment. Our approach is similar to methods used in meta-analysis (1), and the result is similar to a Bayesian shrinkage estimator.

Result 1.1 *Suppose we have two estimators for N_H, $\widehat{N}_H^{(A)}$ and $\widehat{N}_H^{(B)}$. Suppose that the two estimators have sampling variance σ_A^2 and σ_B^2, and that they are unbiased, so that $\mathbb{E}[N_H^{(A)}] = N_H$ and $\mathbb{E}[N_H^{(B)}] = N_H$. Finally, suppose that the two estimators are independent, so that cov$(\widehat{N}_H^{(A)}, \widehat{N}_H^{(B)}) = 0$. Consider the set of all possible estimates*

$$\widehat{N}_H = w\ \widehat{N}_H^{(A)} + (1-w)\ \widehat{N}_H^{(B)}, \tag{A1}$$

where $w \in [0,1]$. Then the weight $w^\star$ which minimizes the mean squared error $\mathbb{E}[(\widehat{N}_H - N_H)^2]$ is

$$w^\star = \frac{\sigma_B^2}{\sigma_A^2 + \sigma_B^2}. \tag{A2}$$

We call the estimate that uses $w^\star$, $\widehat{N}_H = w^\star \widehat{N}_H^{(A)} + (1-w^\star)\widehat{N}_H^{(B)}$, the linear blending *estimate.*

Proof: The mean squared error (MSE) is the sum of the squared bias and the variance, so when the bias is zero, the mean squared error is the same as the variance. Since we have assumed that $\text{cov}(\widehat{N}_H^{(A)}, \widehat{N}_H^{(B)}) = 0$, we can write the variance of the blended estimator as

$$\text{var}(\widehat{N}_H) = \text{var}(w\widehat{N}_H^{(A)} + (1-w)\widehat{N}_H^{(B)}) \tag{A3}$$
$$= w^2\sigma_A^2 + (1-w)^2\sigma_B^2. \tag{A4}$$

So we conclude that, assuming that both estimators are unbiased and independent,

$$\text{MSE}(\widehat{N}_H) = w^2\sigma_A^2 + (1-w)^2\sigma_B^2. \tag{A5}$$

Now we wish to know which value of w will minimize this error; we call this optimum $w^\star$. Taking derivatives, we see that

$$\frac{\partial \text{MSE}(\widehat{N}_H)}{\partial w} = 2w\sigma_A^2 - 2(1-w)\sigma_B^2 \tag{A6}$$

$$= 2w(\sigma_A^2 + \sigma_B^2) - 2\sigma_B^2. \tag{A7}$$

Since we wish to find a minimum, we set this equal to 0 and solve for w to obtain

$$w^\star = \frac{\sigma_B^2}{\sigma_A^2 + \sigma_B^2}. \tag{A8}$$

We did not formally include the constraint that $w \in [0, 1]$ in our derivation, but since variances are always non-negative, $w^\star$ in equation A8 will always satisfy this condition. We can confirm that $w^\star$ is a minimum by differentiating equation A7 again to obtain

$$\frac{\partial^2 \text{MSE}(\widehat{N}_H)}{\partial w^2} = 2(\sigma_A^2 + \sigma_B^2). \tag{A9}$$

As long as we are not in the degenerate case where both variances are identically zero, equation A9 is always greater than 0, meaning that equation A8 is indeed a minimum. (In the degenerate case, the two estimators would always produce identical, exactly correct estimates, since they would both be unbiased estimators with 0 variance.) ■

Finally, we address the question of when it is advantageous to blend the estimates from both arms of an experiment, instead of just using the estimate from one arm. In general, we prefer the blended estimator when we expect it to produce smaller mean squared error than either of the individual arms. Without loss of generality, suppose that experimental condition A outperformed condition B, meaning that we estimate $\text{MSE}_A \leq \text{MSE}_B$. We want to blend as long as $\text{MSE} \leq \text{MSE}_A$. Plugging $w^\star$ into our expression for the mean squared error, equation A5, we have

$$\text{MSE} = (w^\star)^2\sigma_A^2 + (1-w^\star)^2\sigma_B^2 \tag{A10}$$

$$= \sigma_B^2 \frac{\sigma_A^2\sigma_B^2}{(\sigma_A^2 + \sigma_B^2)^2} + \sigma_A^2 \frac{\sigma_A^2\sigma_B^2}{(\sigma_A^2 + \sigma_B^2)^2} \tag{A11}$$

$$= \frac{\sigma_A^2\sigma_B^2}{\sigma_A^2 + \sigma_B^2}. \tag{A12}$$

Comparing this to $\text{MSE}_A = \sigma_A^2$, and assuming $\sigma_A \neq 0$, we have

$$\frac{\sigma_A^2\sigma_B^2}{\sigma_A^2 + \sigma_B^2} \leq \sigma_A^2 \tag{A13}$$

$$\iff \frac{\sigma_B^2}{\sigma_A^2 + \sigma_B^2} \leq 1. \tag{A14}$$

Since $\sigma_A^2 > 0$ and $\sigma_B^2 > 0$ the inequality in equation A14 will always hold. Our conclusion is that, assuming the estimators from each arm are unbiased and independent, blending always produces lower (or equal) expected mean squared error than just choosing the estimate from one arm.

Web Appendix 2

The scale-up literature has long discussed different types of potential bias in basic scale-up estimates (2–11). In this section, we present a framework for assessing the sensitivity of scale-up estimates to these potential sources of bias.

The basic scale-up estimator was introduced in the main text as

$$\widehat{N}_H = \frac{\sum_{i \in s} y_{i,H}/\pi_i}{\sum_{i \in s} \hat{d}_{i,U}/\pi_i} \times N. \tag{A15}$$

However, an equivalent expression for the basic scale-up estimator is,

$$\widehat{N}_H = \left(\frac{\sum_{i \in s} y_{i,H}/\pi_i}{\sum_{i \in s} (\hat{d}_{i,U}/\pi_i)/N_F} \right) \times N/N_F = \frac{\hat{y}_{F,H}}{\hat{\bar{d}}_{F,U}} \times N/N_F \tag{A16}$$

where F is the survey's frame population (in this case, Rwandans aged 15 and over); $\hat{y}_{F,H}$ is the estimated number of connections between people in the frame population F and the hidden population H; and $\hat{\bar{d}}_{F,U}$ is the estimated average number of connections between the frame population F and the entire population U.

Note that, because all connections under consideration are symmetric, the number of connections from F to U is the same as the number of connections from U to F. Therefore, using basic algebra we can re-express equation A16 as

$$\widehat{N}_H = \frac{\hat{y}_{F,H}}{\hat{\bar{d}}_{U,F}}. \tag{A17}$$

Equation A17 allows us to decompose the difference between the basic scale-up estimand and the true hidden population size into several adjustment factors (2):

$$\underbrace{N_H}_{\substack{\text{true size}\\ \text{of hidden}\\ \text{population}}} = \underbrace{\left(\frac{y_{F,H}}{\bar{d}_{U,F}} \right)}_{\substack{\text{basic}\\ \text{scale-up}\\ \text{estimand}}} \times \underbrace{\underbrace{\frac{1}{\bar{d}_{F,F}/\bar{d}_{U,F}}}_{\substack{\text{frame ratio}\\ \phi_F}} \times \underbrace{\frac{1}{\bar{d}_{H,F}/\bar{d}_{F,F}}}_{\substack{\text{degree ratio}\\ \delta_F}} \times \frac{\overbrace{y^{+}_{F,H}/y_{F,H}}^{\text{precision } \eta_F}}{\underbrace{\bar{v}_{H,F}/\bar{d}_{H,F}}_{\substack{\text{true positive rate}\\ \tau_F}}}}_{\text{overall adjustment factor } \alpha_F}, \tag{A18}$$

We now describe each of the four adjustment factors in turn. The frame ratio, ϕ_F, is

$$\phi_F = \frac{\text{avg \# connections from a member of } F \text{ to the rest of } F}{\text{avg \# connections from a member of } U \text{ to } F} = \frac{\bar{d}_{F,F}}{\bar{d}_{U,F}}. \tag{A19}$$

ϕ_F can range from zero to infinity, and in most practical situations we expect ϕ_F will be greater than one.

The degree ratio, δ_F, is

$$\delta_F = \frac{\text{avg \# connections from a member of } H \text{ to } F}{\text{avg \# connections from a member of } F \text{ to the rest of } F} = \frac{\bar{d}_{H,F}}{\bar{d}_{F,F}}. \tag{A20}$$

δ_F ranges from zero to infinity, and it is less than one when the hidden population members have, on average, fewer connections to the frame population than frame population members.

The true positive rate, τ_F, is

$$\tau_F = \frac{\text{\# reported edges from } F \text{ actually connected to } H}{\text{\# edges connecting } F \text{ and } H} = \frac{y^+_{F,H}}{d_{F,H}}. \tag{A21}$$

Here we have written $y^+_{F,H}$ to mean the true positive reports among the $y_{F,H}$; that is, those reports about alters who truly are in the hidden population. τ_F ranges from 0, if none of the edges are correctly reported, to 1 if all of the edges are reported. Substantively, the more stigmatized the hidden population, the closer we would expect τ_F to be to 0.

Finally, the precision, η_F, is

$$\eta_F = \frac{\text{\# reported edges from } F \text{ actually connected to } H}{\text{\# reported edges from } F \text{ to } H} = \frac{y^+_{F,H}}{y_{F,H}}. \tag{A22}$$

The precision is useful because it relates the observed out-reports, $y_{F,H}$ to the true positive out-reports, $y^+_{F,H}$. It varies from 0, when none of the out-reports are true positives, to 1, when the out-reports are perfect.

The estimates produced by the basic scale-up estimator – which was used in many previous studies, and which we use throughout the main text – make the assumption that the product of these adjustment factors is 1. Equation A18 shows that if this assumption does not hold, the estimates will be biased. For example, the results of the internal consistency checks in Figure 3 from the main article show that our internal consistency check estimate for the number of teachers is too high. This result suggests that the product of the four adjustment factors is less than 1. This could happen if, for example, teachers have more average connections to members of the frame population than members of the frame population have to each other, leading the degree ratio, $\delta_F = \bar{d}_{H,F}/\bar{d}_{F,F}$ to be greater than 1.

Previous research has discussed several potential sources of bias in basic scale-up estimates, including i) structural differences between the networks of hidden population members and the general population, which have been called *barrier effects* (3, 7, 10); ii) respondents who are not perfectly aware of the characteristics of their network alters, which has been called *transmission error* (3, 7–9, 12); and iii) respondents who make errors in their network reports, which is called *recall error* (3, 5–7). All three of these possible sources of bias are accounted for by the framework summarized in equation A18: barrier effects will impact the degree ratio δ_F; and transmission and recall errors will impact the reporting terms, η_F and τ_F.

The results above and the results in the main text all assume that the conditions for the known population estimator are met; see the article by Feehan and Salganik (2). If those conditions are not met, then researchers could use the robustness results in the article by Feehan and Salganik (2) to construct additional adjustment factors.

Researchers can use the theoretical results from the article by Feehan and Salganik (2), which are summarized above in equation A18, and the empirical results from our study (Web Table 1) to produce their own estimates under different assumptions about the magnitude of potential biases. Researcher can also apply the blending weights derived in Section 1 to estimates based on their own assumptions to produce a blended estimate, $\widehat{N}_H^{\dagger}$.

Web Table 1. Quantities needed for sensitivity analysis.

	Meal definition		Acquaintance definition	
	$\widehat{N}_H^{(A)}$	$\widehat{\sigma}_A^2$	$\widehat{N}_H^{(B)}$	$\widehat{\sigma}_B^2$
Male clients of sex workers	33,800	13,753,311	23,632	11,158,493
Female sex workers	34,213	11,894,878	30,466	6,810,654
Men who have sex with men	2,219	1,186,857	1,368	75,333
People who inject drugs	613	74,056	2,141	573,264

Web Table 2. Possible assumptions a researcher might make about ϕ_F, δ_F, τ_F, and η_F for estimating the number of female sex workers from each arm of the study.

Quantity	Meal definition	Acquaintance definition
frame ratio (ϕ_F)	1.30	1.30
degree ratio (δ_F)	0.90	0.50
true positive rate (τ_F)	0.80	0.40
precision (η_F)	0.99	0.95

As a concrete example, suppose a researcher thinks that the most plausible values for ϕ_F, δ_F, τ_F, and η_F, are the ones in Web Table 2. Using these values, the implied value for the overall adjustment factor for the arm with the meal definition, $\alpha_F^{(A)}$ is

$$
\begin{aligned}
\alpha_F^{(A)} &= \frac{\eta_F^{(A)}}{\phi_F^{(A)}\ \delta_F^{(A)}\ \tau_F^{(A)}} && \text{(A23)}\\
&= \frac{0.99}{1.30 \times 0.90 \times 0.80} && \text{(A24)}\\
&\approx 1.06 && \text{(A25)}
\end{aligned}
$$

This means that the unbiased estimator produces

$$
\begin{aligned}
\widehat{N}_H^{(A)\dagger} &= \alpha_F^{(A)} \times \widehat{N}_H^{(A)} && \text{(A26)}\\
&= 1.06 \times 34,213 && \text{(A27)}\\
&\approx 36,187 && \text{(A28)}
\end{aligned}
$$

For the other arm, which used the acquaintance definition, we have

$$
\begin{aligned}
\alpha_F^{(B)} &= \frac{\eta_F^{(B)}}{\phi_F^{(B)}\ \delta_F^{(B)}\ \tau_F^{(B)}} && \text{(A29)}\\
&= \frac{0.95}{1.30 \times 0.50 \times 0.40} && \text{(A30)}\\
&\approx 3.65 && \text{(A31)}
\end{aligned}
$$

In this case, the unbiased estimator produces

$$
\begin{aligned}
\widehat{N}_H^{(B)\dagger} &= \alpha_F^{(B)} \times \widehat{N}_H^{(B)} && \text{(A32)}\\
&= 3.65 \times 30,466 && \text{(A33)}\\
&\approx 111,320 && \text{(A34)}
\end{aligned}
$$

Next, we note that, based on the estimated sampling variances, the optimal blending weight, $w^\star$ (equation A2), is:

$$
\begin{aligned}
w^\star &= \frac{\sigma_B^2}{\sigma_A^2 + \sigma_B^2} && \text{(A35)}\\
&= \frac{6,810,654}{11,894,878 + 6,810,654} && \text{(A36)}\\
&\approx 0.36 && \text{(A37)}
\end{aligned}
$$

Applying the blending weight to the adjusted estimates results in a final estimate of

$$
\begin{aligned}
\widehat{N}_H^{\dagger} &= w\widehat{N}_H^{(A)\dagger} + (1-w)\widehat{N}_H^{(B)\dagger} && \text{(A38)}\\
&= 0.36 \times 36,187 + (1-0.36) \times 111,320 && \text{(A39)}\\
&= 83,964 && \text{(A40)}
\end{aligned}
$$

Under these assumptions, then, the blended point estimate is 83,964 female sex workers.

Web Appendix 3

Here, we give additional details about data collection and evaluation. First, we describe how we obtained information about the groups of known size for Rwanda. Next, we discuss several balance checks we conducted to ensure that randomizing the tie definition followed the study plan. Finally, we describe a manipulation check which demonstrates that respondents pay attention to the definition of a tie they are asked to report about.

We collected information about groups of known size for Rwanda from several different sources (Table 1 in the main article). Although there is currently no general method that researchers can use to compile their own list of groups of known size in a different setting, our experience in Rwanda suggests that this would also be feasible in other developing countries. Our goal was to find about 20 groups of known size whose prevalence in the general population varied from about 0.1% to 3%, consistent with the design of earlier studies. As an illustration, we briefly describe the three main strategies we used to compile groups of known size for this study.

First, we obtained several of the groups of known size by asking local institutions to consult their administrative records. We contacted the Catholic Church to ask for the total number of priests in the country; we contacted the Ministry of Education to obtain the number of teachers; and we contacted the Ministry of Health to obtain the number of male community health workers and the number of nurses or doctors. We also used an official report from the International Committee of the Red Cross to obtain the number of incarcerated people.

Second, several of the groups of known size come from quantities that can be estimated as part of a demographic and health survey (DHS). However, DHS surveys typically interview only women aged 15-49 and men aged 15-59. Therefore, we had to extrapolate estimates from the DHS frame population to all Rwandans. We did so by taking proportions estimated from the DHS and applying them to population projections by age group for Rwanda, provided to us by the National Institute of Statistics Rwanda (NISR).

Third, we obtained information about the prevalence of several names by asking the Rwandan Government for a tally of names from the national identity card database. This database has two types of names: Kinyarwanda names and Christian names. Since not all Rwandans have Christian names, we used Kinyarwanda names. We retained all names whose frequency was at least 1% of the most frequent name, by gender; for males, we retained all names with at least 885 people, and for females, we retained all names with at least 716 people. We also removed all names that occurred for both males and females. Finally, we sorted the names by popularity and then looked for male and female pairs that had similar popularity (over a range of popularity values). For each pair we checked to make sure that it did not have multiple spellings or a nickname.

Each of these known population totals may have error. However, the article by Feehan and Salganik (2) shows that the known population estimator is consistent and unbiased when the *sum* these known population totals is accurate, not the size of each individual one. If there is random error in the known population sizes, then using a large number of groups of known size may lead these errors to cancel out. If researchers think there is non-random error in the known population sizes, then the methods of Feehan and Salganik (2) can be used to assess the impact these errors will have on size estimates.

In order to confirm that households were assigned to the different tie definitions according to the randomization procedure required by our study plan, we checked the balance between the two arms. These checks suggest that the randomization of households to tie definitions was conducted according to the study design.

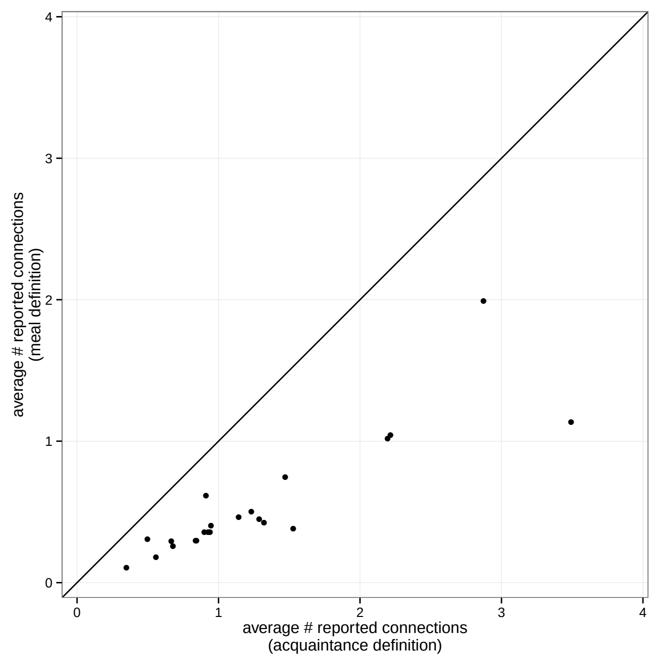
Web Table 3 compares the distribution of five household-level variables between households assigned to the meal definition and households assigned to the acquaintance definition. No substantively meaningful differences are apparent. We used random permutation tests, accounting for our blocked and clustered randomization, to compute p-values for the difference in household covariates between each arm; Web Table 3 has the results. However, these p-values are difficult to interpret because 1) there is a linear dependence among the income quintiles; and 2) they do not account for the fact that we make multiple comparison.

Web Table 3. Comparison between households assigned to the acquaintance tie definition and households assigned to the meal tie definition. All values are fractions, with the exception of the average number of respondents. P-values are from non-independent random permutation tests. Sampling weights were not used in computing these means.

	Acquaintance	Meal	Two-Sided P Value
Num. respondents in household	2.263	2.209	0.335
Lowest wealth quintile	0.088	0.116	0.054
Second wealth quintile	0.151	0.153	0.946
Third wealth quintile	0.203	0.208	0.957
Fourth wealth quintile	0.210	0.210	1
Highest wealth quintile	0.348	0.312	0.086

Therefore, we also conducted an omnibus test for imbalance (13). This test accounts for the blocked and clustered nature of respondents' random assignment to a tie definition; furthermore, the omnibus test avoids the problem of testing multiple hypotheses, which arises when separately testing several individual covariates for balance. Finally, the omnibus test permits us to test both individual and household-level characteristics. The omnibus test included as covariates the number of respondents in the household, the five wealth quintiles, the total age of each household, the total number of females in each household, and the total number of people in each of the three education categories in each household. The (13) omnibus test did not find any significant evidence of imbalance (χ^2 test on 9 degrees of freedom produced a two-sided $p = 0.22$). Therefore, we conclude that the randomization of households to different definitions of a network tie was conducted according to the study design.

We also performed a manipulation check to assess whether or not responses differed between the two experimental arms: Web Figure 1 shows that the average number of reported connections to each of the 22 groups of known size was lower for the meal definition than the acquaintance definition. This test suggests that respondents assigned to the meal definition did indeed respond differently from respondents assigned to the acquaintance definition.



Web Figure 1. Average number of connections by tie definition for 22 groups of known size (Table 1 in the main article). On the x axis is the average number of connections from the experimental arm that used the acquaintance definition of a network tie, and on the y axis is the average number of connections from the arm that used the meal definition. The average number of reported connections for the meal definition is consistently lower than for the acquaintance definition.

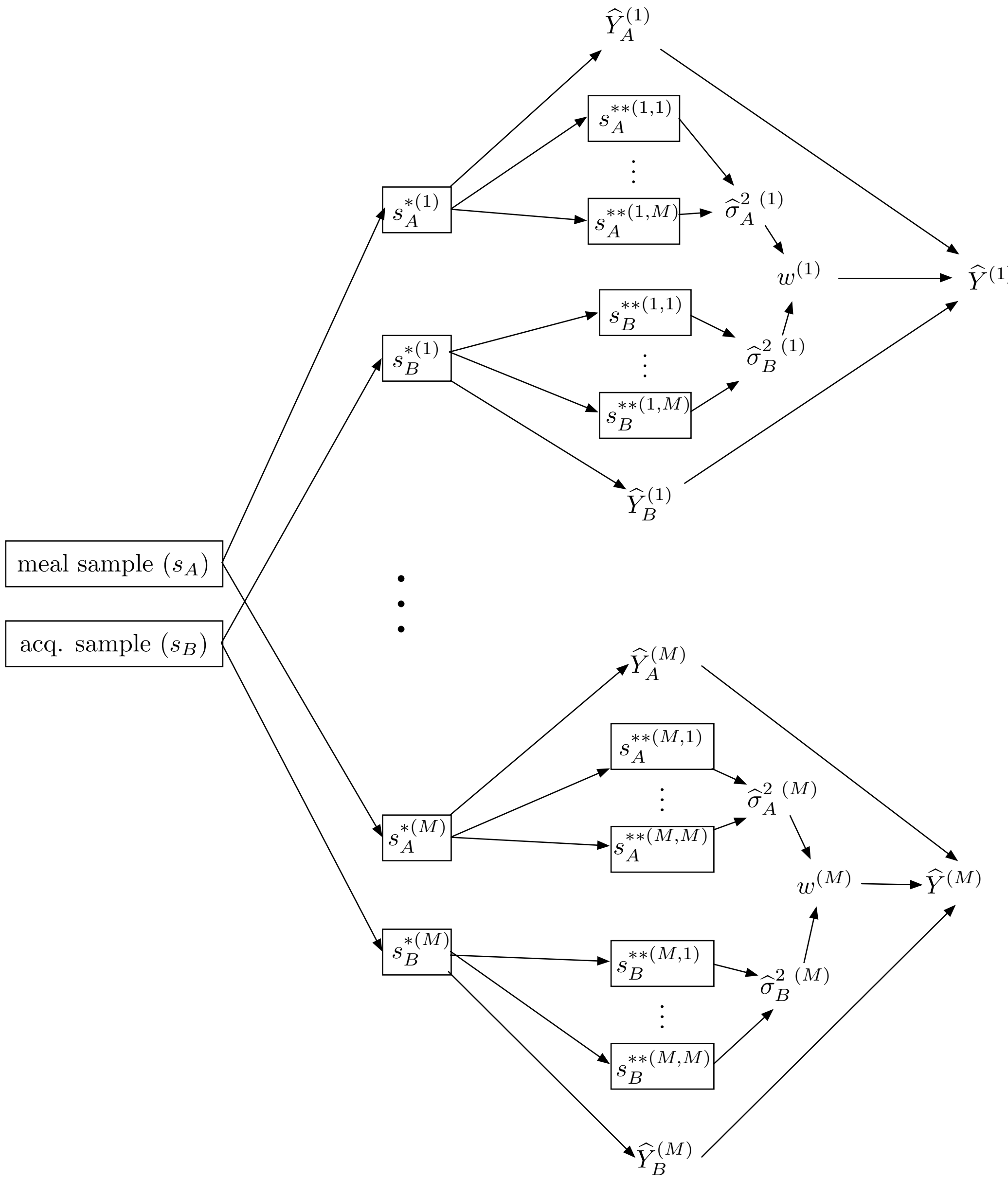
Web Appendix 4

In order to compare the internal consistency results for the two tie definitions, and to put intervals around our non-blended estimates, we obtained 2,000 bootstrap resamples of our data set using the rescaled bootstrap method, which accounts for our study's complex sample design (2, 14, 15). We constructed confidence intervals from the central 95% of the resampled estimates.

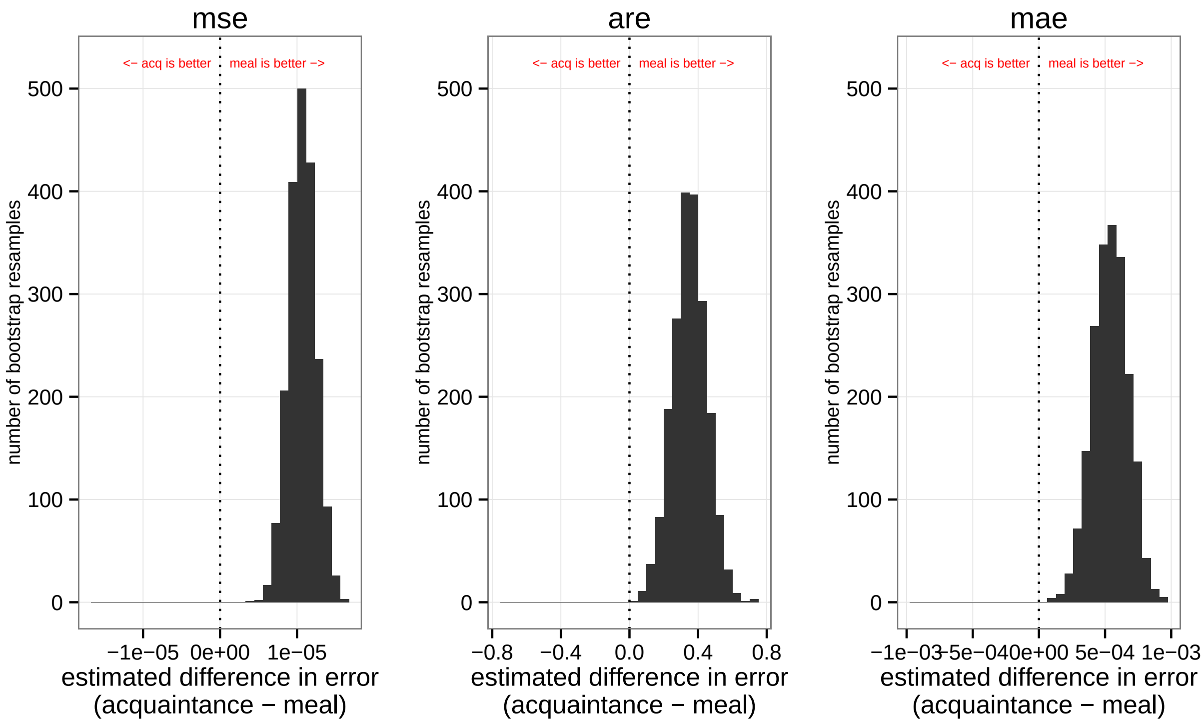
The procedure for producing a measure of uncertainty of the blended estimate was more complex (Web Figure 2). We took $M = 2,000$ bootstrap resamples $s_A^{(1)}, \ldots, s_A^{(M)}$ and $s_B^{(1)}, \ldots, s_B^{(M)}$ from each experimental arm. Within the ith bootstrap resample for each arm, we computed a size estimate $\widehat{N}_H^{(A)(i)}$ and $\widehat{N}_H^{(B)(i)}$. Also within the ith bootstrap resample for each arm, we obtained M bootstrap resamples

$s_B^{\star\star(i,1)}, \ldots, s_B^{\star\star(i,M)}$ and $s_B^{\star\star(i,1)}, \ldots, s_B^{\star\star(i,M)}$, which we use to estimate the sampling variance $\widehat{\sigma}_A^{2\ (i)}$ and $\widehat{\sigma}_B^{2\ (i)}$. We use equation A2 to estimate the blending weight $w^{(i)}$ from these estimated sampling variances. Finally, we produce a blended estimate $\widehat{N}_H^{(i)}$ for the ith bootstrap resample using $w^{(i)}$. We constructed confidence intervals from the central 95% of the resampled blended estimes.

We also performed the internal consistency checks on each set of $M = 2,000$ bootstrap resamples to reach our conclusion that the meal definition produces lower error than the acquaintance definition by three error metrics: the mean squared error, the average relative error, and the mean absolute error. For each bootstrap resample, for each arm, and for each error metric, we produce a summary of the error across the groups of known size. Web Figure 3 show our results. For all three error metrics, and for all bootstrap resamples, the meal definition attains a lower error than the acquaintance definition ($p < 0.001$).



Web Figure 2. Illustration of the procedure used to estimate the sampling uncertainty in the blended estimates. We took $M = 2,000$ bootstrap resamples from each experimental arm. Within each bootstrap resample, a size estimate is computed. Also within each bootstrap resample, an additional M resamples are taken to estimate the sampling variance. The blending weight is computed from the estimated sampling variance for each arm, and a blended estimate is formed. We end up with M blended estimates, and we take the middle 95% for our uncertainty interval.



Web Figure 3. Three summaries of the errors in estimates from the internal consistency checks. Results are shown for, from left to right, mean squared error, average relative error, and mean absolute error. If the true known population sizes are $X_1, \ldots, X_K$ and our estimated sizes are $\widehat{X}_1, \ldots, \widehat{X}_K$, then the mean square error is $\text{MSE} = \frac{1}{K}\sum_{i=1}^{K}(\widehat{X}_i - X_i)^2$; the average relative error is $\text{ARE} = \frac{1}{K}\sum_{i=1}^{K}\frac{\widehat{X}_i}{X_i}$; and the mean absolute error is $\text{MAE} = \frac{1}{K}\sum_{i=1}^{K}|\widehat{X}_i - X_i|$. Each panel shows the distribution, across bootstrap resamples, of the difference in estimated total error between the acquaintance and meal definitions. The dashed vertical line is at 0, which is what we would expect to see if there was no difference in error between the two tie definitions. Positive values, shown to the right of the dashed line, indicate bootstrap resamples for which the meal definition performed better by attaining lower error. The meal definition outperforms the acquaintance definition on all three error criteria.

Web Appendix 5

Here, we provide additional details needed to compare our estimates to earlier estimates from Rwanda and to benchmarks from the Joint United Nations Programme on HIV/AIDS (UNAIDS). First, we present the UNAIDS benchmarks and describe how they are constructed. Next, we give additional detail about the comparison with previous studies that was discussed in the main text. Finally, we compare the estimates from our study with both the UNAIDS benchmarks and the previous studies from Rwanda.

The UNAIDS benchmark estimates are derived from published literature on key population size estimates from around the world. UNAIDS benchmark estimates for the sizes of hidden populations are presented as a prevalence within some portion of the entire population. For example, the benchmark estimated number of sex workers in a country in Sub-Saharan Africa is 0.4% to 4.3% of females age 15-49 (16). Web Table 4 reproduces UNAIDS benchmark ranges suggested for Rwanda for all of the hidden populations except for men who have sex with men, for which no benchmarks are available for African countries. For injecting drug users, the UNAIDS guidelines do not specify the portion of the population to which the estimates apply, so we assume it to be both men and women aged 15-49. In order to estimate quantities such as the number of women between 15-49, we take the age-sex distribution from the United Nations Population Division and apply it to the total population estimate for Rwanda in 2011, which was provided to the study team by the National Institute of Statistics of Rwanda (Web Table 5). The resulting benchmark estimates for Rwanda are shown in Web Table 6.

For example, in order to obtain the benchmark estimate for the number of female sex workers, we take the number of females aged 15-49 – 2,592,039 – and multiply it by 0.4% and 4.3%, to get a benchmark range of 10,368 - 111,458. In order to obtain a point estimate, we take the midpoint of the UNAIDS benchmark range, 2.35%; in this case, that gives us 64,492.

Web Table 4. UNAIDS benchmark ranges for the populations most at-risk for HIV/AIDS from Spectrum Software manual (16, pg. 22).

Quantity	Region	Benchmark	Denominator
injecting drug users	Middle East, North Africa, Sub-Saharan Africa	0.05 - 2.07%	all aged 15-49
female sex workers	Sub-Saharan Africa	0.4-4.3%	females aged 15-49
clients of sex workers	Eastern and southern Africa	10-11%	males aged 15-49
men who have sex with men	Africa	(not available)	

In the main text, we compare our study's estimates to previous studies that were conducted in Rwanda. In this section, we give more detail about those other studies.

Web Table 7 compares the definitions of the hidden population used in our study and in the three other studies to which we compare. For female sex workers, there are three existing estimates. The first is from a capture-recapture study (18). The capture was the 2010 Behavioural Surveillance Survey (BSS) of female sex workers (19). The recapture was part of a mapping of sex workers in which enumerators

Web Table 5. Population estimates used to produce UNAIDS benchmark values, and also to produce hidden population size estimates. UNPD is the United Nations Population Division estimate of Rwanda's age-sex distribution in 2010 (17), and NISR is the National Statistical Institute of Rwanda, who provided the study team with an estimate for Rwanda's total population in 2011.

Quantity	Source	Estimate
Total population in 2011	NISR	10,718,378
% males aged 15-49	UNPD	23%
% females aged 15-49	UNPD	24%
Number of males aged 15-49	our calculations	2,426,821
Number of females aged 15-49	our calculations	2,592,039
Number of either sex aged 15-49	our calculations	5,018,860

Web Table 6. Benchmark estimates for hidden populations in Rwanda computed from the information in Web Table 4 and Web Table 5.

Hidden population	Benchmark for Rwanda
injecting drug users	64,492 (25,094 - 103,890)
female sex workers	60,913 (10,368 - 111,458)
clients of sex workers	254,816 (242,682 - 266,950)
men who have sex with men	(not available)

asked sex workers whether or not they participated in the BSS (20). The second existing estimate for the number of female sex workers is based on enumeration of female sex workers at sites selected through time-location sampling (18). The third existing estimate for the number of female sex workers is based on a participatory mapping of female sex workers in Rwanda in 2011 and 2012 (20).

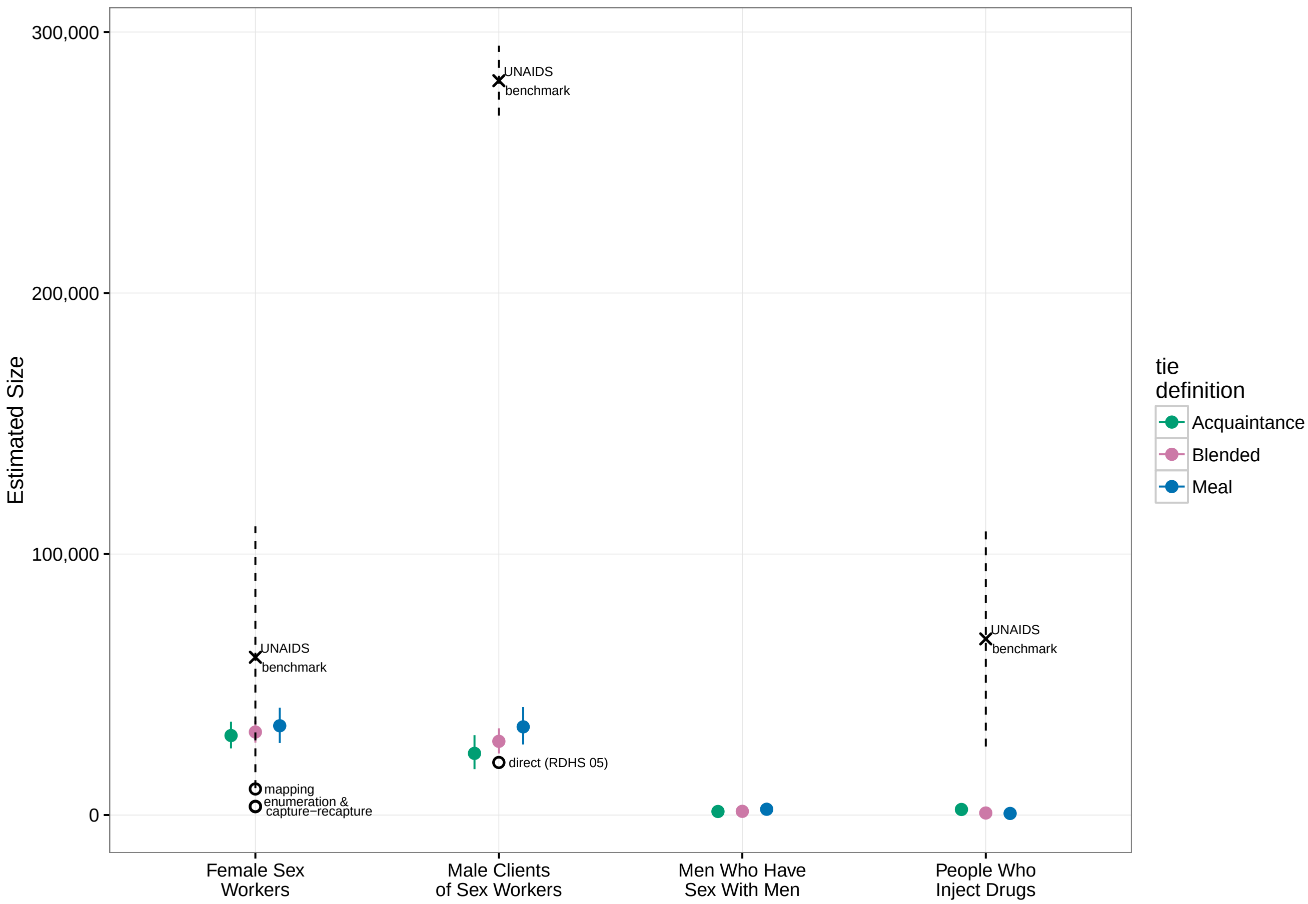
For the number of male clients of sex workers, the existing estimate comes from a direct question posed to a nationally representative sample of men in the 2005 Rwanda DHS (21). The survey asked men whether they had paid for sex in the past 12 months and also whether they had ever paid for sex. The DHS survey requires that these questions be asked when no other person is within hearing distance of the interview. Despite this precaution, men are likely to underreport this behavior while being interviewed in their own homes. The survey found that 0.83 percent of men reported having paid for sex in the 12 months prior to the survey. Assuming that this proportion has not changed between 2005 and 2011, we can multiply 0.83 percent by the male population aged 15-49 (Web Table 5) to estimate that approximately 20,142 men 15-49 paid for sex in the preceding 12 months (20).

In all cases, the definitions are not exactly comparable, which may contribute to the differences between our results and previous estimates.

Finally, we compare our study's estimates to both the UNAIDS benchmarks and to the results of the previous studies (Web Figure 4). Our estimates are generally higher than the ones produced by the three comparison studies, but lower than the UNAIDS benchmarks, with the exception of sex workers, where our estimates are within the UNAIDS benchmark range, and men who have sex with men, where no comparisons and no UNAIDS benchmarks are available.

Web Table 7. Question wording used to estimate the size of four key populations at risk for HIV in Rwanda. Differences in the precise definition of membership in the hidden population may contribute to discrepancies in size estimates. Fsw is female sex workers, csw is male clients of sex workers, msm is men who have sex with men and idu is injecting drug users.

Source	Hidden population	Wording (English)	Wording (Kinyarwanda)
Our study (acq.)	fsw	"How many females do you know who are practicing prostitution?"	"Ni abantu bangahe bigitsina gore uzi bakora umwuga wuburaya?"
Our study (acq.)	csw	"How many men who buy prostitutes do you know?"	"Ni abantu bangahe bigitsina gabo bagura indaya/ umuntu wigitsina gore?"
Our study (acq.)	msm	"How many men do you know who have sex with men?"	"Ni abantu bangahe bigitsina gabo bakora imibonano mpuzabitsina nabandi bagabo?"
Our study (acq.)	idu	"How many people of your acquaintances inject drugs in their veins?"	"Mu gace mutuyemo, ni abantu bangahe bitera ibiyobyabwenge mu mitsi?"
Our study (meal)	fsw	"How many females sex workers have you shared a meal or drink with?"	"Ni abantu bangahe b'igitsina gore bakora umwuga w'uburaya mwasangiye ibiryo cg se icyo kunywa?"
Our study (meal)	csw	"How many clients of female sex workes have you shared a meal or drink with?"	"Ni abagabo bangahe bagura indaya mwasangiye ibiryo cyangwa se icyo kunywa?"
Our study (meal)	msm	"How many men you shared a meal or drink with have sex with other men?"	"Ni abagabo bangahe bakora imibonano mpuzabitsina n'abandi bagabo mwasangiye ibiryo cg se icyo kunywa? "
Our study (meal)	idu	"How many acquaintances you shared meal or drink with use injectable narcotics?"	"Ni abantu bangahe bitera ibiyobyabwenge mu mitsi mwasangiye ibiryo cg se icyo kunywa?"
Rwanda 2005 DHS	csw	"In the last 12 months, did you pay anyone in exchange for sex?"	"Wigeze ugura igitsina mu mezi cumi nabiri ashize?"
Capture-recapture 2010	fsw	Women who are engaged in either street or establishment based sex work	Abagore bakora umwuga wuburaya ku muhanda cyangwa mu kigo gikora umwuga wuburaya.
Enumeration 2010	fsw	Women who are engaged in either street or establishment based sex work	Abagore bakora umwuga wuburaya ku muhanda cyangwa mu kigo gikora umwuga wuburaya.
Participatory mapping 2011-12	fsw	Women who have sex for money, home based, street based and venue based	Abagore bakora imibonano mpuzabitsina ku mafaranga, mu rugo, mu muhanda cyangwa mu kigo gikora umwuga wuburaya



Web Figure 4. Comparison between our estimates, the UNAIDS benchmarks, and the estimates from other studies in Rwanda. Our estimates are higher than all earlier estimates from Rwanda and are either lower than or within the bounds given by the UNAIDS benchmarks.

References

1. Hedges LV. *Statistical Methods for Meta-Analysis*. Orlando: Academic Press, 1 edition ed., 1985.
2. Feehan DM, Salganik MJ. Generalizing the Network Scale-Up Method: A New Estimator for the Size of Hidden Populations. *Sociological Methodology* In Press;.
3. Maltiel R, Raftery AE, McCormick TH, et al. Estimating population size using the network scale up method. *Annals of Applied Statistics* 2015;9(3):1247–1277.
4. Bernard HR, Hallett T, Iovita A, et al. Counting hard-to-count populations: the network scale-up method for public health. *Sexually Transmitted Infections* 2010;86(Suppl 2):ii11–ii15. doi: 10.1136/sti.2010.044446.
5. McCormick TH, Zheng T. Adjusting for recall bias in "How many X's do you know?" surveys. In: *Proceedings of the joint statistical meetings*. Salt Lake City, UT.
6. Killworth PD, McCarty C, Bernard HR, et al. Two interpretations of reports of knowledge of subpopulation sizes. *Social Networks* 2003;25(2):141–160. doi:10.1016/S0378-8733(02)00040-0.
7. Killworth PD, McCarty C, Johnsen EC, et al. Investigating the variation of personal network size under unknown error conditions. *Sociological Methods & Research* 2006;35(1):84–112.
8. Shelley GA, Killworth PD, Bernard HR, et al. Who knows your HIV status II?: Information propagation within social networks of seropositive people. *Human Organization* 2006;65(4):430–444.
9. Shelley GA, Bernard HR, Killworth P, et al. Who knows your HIV status? What HIV+ patients and their network members know about each other. *Social Networks* 1995;17(3-4):189–217. doi: 10.1016/0378-8733(95)00262-M.
10. Zheng T, Salganik MJ, Gelman A. How many people do you know in prison?: Using overdispersion in count data to estimate social structure in networks. *Journal of the American Statistical Association* 2006;101(474):409–423.
11. Cowan SK. Secrets and Misperceptions: The Creation of Self-Fulfilling Illusions. *Sociological Science* 2014;1:466–492.
12. Salganik MJ, Mello MB, Abdo AH, et al. The game of contacts: Estimating the social visibility of groups. *Social Networks* 2011;33(1):70–78. doi:10.1016/j.socnet.2010.10.006.
13. Hansen BB, Bowers J. Covariate balance in simple, stratified and clustered comparative studies. *Statistical Science* 2008;:219–236.
14. Rust K, Rao J. Variance estimation for complex surveys using replication techniques. *Statistical Methods in Medical Research* 1996;5(3):283 –310. doi:10.1177/096228029600500305.
15. Rao JNK, Wu CFJ. Resampling Inference With Complex Survey Data. *Journal of the American Statistical Association* 1988;83(401):231–241. doi:10.2307/2288945.

16. Joint United Nations Programme on HIV/AIDS. *Quick Start Guide for Spectrum.* Geneva, Switzerland: Joint United Nations Programme on HIV/AIDS, 2014.

17. United Nations. *World Population Prospects: The 2012 Revision.* United Nations New York, NY, USA, 2012.

18. Mutagoma M, Kayitesi C, Gwiza A, et al. Estimation of the size of the female sex worker population in Rwanda using three different methods. *International Journal of STD & AIDS* 2015;26(11):810–814. doi:10.1177/0956462414555931.

19. Rwanda Biomedical Center. Behavioral and biological surveillance survey among female sex workers. Tech. rep., Center for Treatment and Research on AIDS, Malaria, Tuberculosis and Other Epidemics, 2010.

20. Rwanda Biomedical Center. Estimating the size of key populations at higher risk of HIV through a household survey (ESPHS) Rwanda 2011. Tech. rep., Calverton, Maryland, USA: RBC/IHDPC, SPF, UNAIDS and ICF International, 2012.

21. Institut National de la Statistique du Rwanda, ORC Macro. Rwanda Demographic and Health Survey 2005. Tech. rep., Institut National de la Statistique du Rwanda (INSR) and ORC Macro, Calverton, Maryland, U.S.A., 2006.