

# UCSF

## UC San Francisco Previously Published Works

### Title

Software Application Profile: The Anchored Multiplier calculator—a Bayesian tool to synthesize population size estimates

### Permalink

<https://escholarship.org/uc/item/7j26p5px>

### Journal

International Journal of Epidemiology, 48(6)

### ISSN

0300-5771

### Authors

Wesson, Paul D  
McFarland, Willi  
Qin, Cong Charlie  
et al.

### Publication Date

2019-12-01

### DOI

10.1093/ije/dyz101

Peer reviewed



---

## Software Application Profile

# Software Application Profile: The Anchored Multiplier calculator—a Bayesian tool to synthesize population size estimates

Paul D Wesson, <sup>1\*</sup> Willi McFarland,<sup>2</sup> Cong Charlie Qin<sup>3</sup> and Ali Mirzazadeh<sup>2</sup>

<sup>1</sup>Center for AIDS Prevention Studies, Traineeship in AIDS Prevention Studies Fellowship Program, University of California San Francisco, San Francisco, CA, USA, <sup>2</sup>Department of Epidemiology and Biostatistics, University of California San Francisco, San Francisco, CA, USA and <sup>3</sup>Mission One, Fremont, CA, USA

\*Corresponding author. Center for AIDS Prevention Studies/Prevention Research Center, University of California San Francisco, 550 16th St, 3rd Floor, San Francisco, CA 94158, USA. E-mail: paul.wesson@ucsf.edu

Editorial decision 15 April 2019; Accepted 27 April 2019

## Abstract

Estimating the number of people in hidden populations is needed for public health research, yet available methods produce highly variable and uncertain results. The Anchored Multiplier calculator uses a Bayesian framework to synthesize multiple population size estimates to generate a consensus estimate. Users submit point estimates and lower/upper bounds which are converted to beta probability distributions and combined to form a single posterior probability distribution. The Anchored Multiplier calculator is available as a web browser-based application. The software allows for unlimited empirical population size estimates to be submitted and combined according to Bayes Theorem to form a single estimate. The software returns output as a forest plot (to visually compare data inputs and the final Anchored Multiplier estimate) and a table that displays results as population percentages and counts. The web application ‘Anchored Multiplier Calculator’ is free software and is available at [<http://globalhealthsciences.ucsf.edu/resources/tools>] or directly at [<http://anchoredmultiplier.ucsf.edu/>].

**Key words:** Bayesian modelling, population size estimation, key populations, data synthesis

---

## Introduction

Estimating the size of the population at risk has become an explicit aim of studies investigating HIV infection among hidden and hard-to-reach ‘key populations’ (e.g. men who have sex with men, female sex workers, people who inject drugs, transgender individuals etc.). International aid organizations (including CDC, WHO and UNAIDS) have

made the enumeration of key populations a priority in the strategic effort to pivot resources to populations in which the epidemic is concentrated.<sup>1</sup> This enumeration, also referred to as population size estimation (PSE), serves to provide the denominator to calculate the burden of disease, to advocate for resources for marginalized populations and to help set priorities among populations at risk for HIV.

Numerous methods are available to epidemiologists and public health practitioners to directly count or indirectly estimate the size of the population at risk.<sup>2,3</sup> These methods vary in how data are leveraged to estimate the population size: from mapping known venues where risk behaviour occurs (mapping and enumeration, Reverse Tracking Method),<sup>4</sup> to assessing the overlap of selected lists of the target population and noting the frequency with which unique individuals appear on multiple lists (capture-recapture)<sup>5-7</sup> and to crowd-sourcing opinions of the population size from a sample of the target population itself (wisdom of the crowd).<sup>8-10</sup> A frequently implemented method is the multiplier method,<sup>11</sup> a variant of capture-recapture, where the number of people who appear on a programmatic list that serves the population is divided by the proportion of a representative sample who report participating in that programme. Surprisingly, multiple population size estimates calculated from the multiplier method (i.e. different service provider lists are used as counts in the numerator) are often inconsistent with each other; their point estimates differ and corresponding confidence intervals do not overlap.<sup>2</sup>

Calculating multiple size estimates using different methods may suggest a range of possible size estimates or may bolster confidence in an estimate when there is agreement across the different approaches. Ultimately, however, a single most likely or 'best estimate' is needed for programme planning and quantifying the burden of disease. Investigators often combine population size estimates by taking the median of point estimates, along with the median of upper and lower bounds.<sup>8-12</sup> This approach is believed to balance biases from individual PSE methods that may be operating in different directions. However, this approach assumes that all methods, and the resulting estimates, are equally biased (or equally valid). Combining multiple and often disparate estimates is also done in stakeholder meetings, where persons familiar with the population arrive at a consensus guided by prior knowledge and examination of locally generated estimates.<sup>8-10</sup> The question arises as to how to arrive at a best estimate that makes use of all sources without being dependent on subjective decisions about some estimates or giving unrealistic weight to others.

We recently developed a new method, the Anchored Multiplier, which synthesizes PSEs using a Bayesian model, which we now offer as a free web browser-based calculator. Details of the model methodology are described elsewhere.<sup>13</sup> In brief, prior belief of the population size is taken from previous studies, estimates from similar populations or stakeholder input and converted into a beta probability distribution. Empirical results from PSE methods (the data) are also converted into beta probability

distributions. Using Bayes Theorem, the prior distribution is iteratively combined with the probability distribution for the data to generate a single posterior probability distribution of the population size. This posterior distribution is generated using Markov Chain Monte Carlo (MCMC) simulation. For the MCMC simulations we specified three chains, each with 5000 iterations and burn-in period of 2500. From the posterior distribution, we take the mean as the best estimate and the 2.5th percentile and 97.5th percentile as the 95% credible interval. Because the entire probability distribution of population size estimates has been generated, the 95% credible interval has a simple and intuitive interpretation; that is, there is 95% probability that the population size lies between the 2.5th percentile and the 97.5th percentile. In this software application profile, we present the Anchored Multiplier calculator as an application where investigators enter results from their PSE studies and generate a final consensus estimate.

## Implementation

The Anchored Multiplier calculator is implemented as a free web browser-based application, accessible at [<http://globalhealthsciences.ucsf.edu/resources/tools>] and directly at [<http://anchoredmultiplier.ucsf.edu/>]. For optimal performance, we recommend that users access the calculator with Google Chrome. The software is written in Python, adapted from previous R programming that incorporated the *rjags*, *R2jags* and *forestplot* packages. A user-friendly feature of the Anchored Multiplier calculator is the ability of the user to enter size estimates (Figure 1, Panel A and Panel B) (point estimate, lower bound estimate and upper bound estimate), as either a count or a percentage of the general population. Using a loop, the calculator will fit the data inputs to a beta distribution, finding the best beta distribution shape parameters that reflect the PSE values and the certainty associated with the estimate (indicated by the spread of the distribution). The final beta distribution that is fitted to the data input is guaranteed to have at least 70% of the probability mass of the distribution contained within the lower and upper bounds. The beta distribution that is fitted to the data is then combined with the prior distribution in order to generate the posterior probability distribution. We decided to fit PSE values, as population proportions, to beta distributions because these probability distributions are flexible and mathematically convenient for quantities constrained to lie between 0 and 1.<sup>14</sup>

Example: calculating the number of female sex workers in Tehran, Iran, from multiple empirical point estimates from different methods

To demonstrate the features of the Anchored Multiplier calculator, we draw from the example of female sex



**Figure 1.** Panel A. Data input page for Anchored Multiplier calculator, entering the general population size and the prior information (example of female sex worker population size estimation in Tehran, Iran). Panel B. Data input page for Anchored Multiplier calculator, entering the empirical population size estimation fields (example of female sex worker population size estimation in Tehran, Iran) Panel C. Data output page for Anchored Multiplier calculator, forest plot of data input and results (example of female sex worker population size estimation in Tehran, Iran). Panel D. Data output page for Anchored Multiplier calculator, table of results and data input (example of female sex worker population size estimation in Tehran, Iran).

workers (FSW) in Tehran, Iran. In 2015, in collaboration with Kerman Knowledge Hub and CDC-Iran, we implemented a survey in 13 cities in Iran to assess the prevalence of HIV infection and associated risk factors among FSW. Two PSE methods, the wisdom of the crowd (WOTC) and the multiplier method, were integrated into this survey. In a previous general population survey, the network scale-up method (NSUM) was implemented.<sup>15,16</sup> The NSUM measures the characteristics of the social networks of a representative sample of the general population and extrapolates these characteristics to calculate population prevalence. For example, if on average 10% of the sample's social network are teachers, then it would be estimated that 10% of the population are teachers. The median of the estimates from all three PSE methods was taken as the final size estimate for each survey site. Details of the study design and main outcomes are reported elsewhere.<sup>17</sup> The present example uses size estimates of FSW in Tehran, the largest city and capital of Iran.

To implement the Anchored Multiplier calculator, users must first input the size of the general population, from which the target population is a subset. For FSW, the general population size will be the number of adult women (15–49 years old) in the study site. The general population size is required so that size estimates from PSE methods can be converted to population proportions, and so that final estimates from the Anchored Multiplier can be presented as both population percentages and as population counts. We used 2 581 019 as the size of the adult female population in Tehran (Figure 1, Panel A).<sup>18</sup>

Next, the user must specify a 'prior' estimate for the population size of the target population. Priors are central to the Bayesian paradigm. Priors quantify knowledge or belief in the value of the parameter of interest (in this case, the population size) and the degree of certainty of that prior knowledge, operationalized by the width of the interval between the lower and upper bounds (the narrower the interval, the stronger the belief in the estimate, and vice versa).<sup>19,20</sup> Selecting an appropriate prior is a frequently debated topic for Bayesian analyses.<sup>19,21</sup> Generally, priors reflect information from previous studies or reviews. For population size estimation, the prior will often come from previous PSE studies, where size estimates (and a plausible range) are reported for similar populations from neighbouring geographical settings. Therefore, researchers should look to the published scientific literature as well as the grey literature as sources to define the prior. Priors may also be estimated from consultation with stakeholders—community members and service providers who work with and have knowledge of the target population. If the user wishes to rely heavily on the empirical PSE results, or has very little prior knowledge to specify, a uniform prior may

be specified. When specifying a uniform prior, the user simply checks the box next to 'Uniform' and inputs the plausible range of population size estimates, providing a value for the lower bound and for the upper bound. The uniform prior will then assign equal weight to the range of values spanning the user-defined lower and upper bounds. Whether using an informative prior (with a point estimate and bounds) or a less informative uniform prior, the prior should represent information apart from the data that are the source of the empirical size estimates.<sup>19</sup> It is common to vary the prior in order to test the sensitivity of the posterior distribution to the choice of priors.<sup>13,19</sup> This practice, along with the general use of priors, adds to the transparency in Bayesian analysis, as a priori belief must be quantified and justified. For FSW in Tehran we used 1.54% (1.36–1.71%) as the prior, taken from the NSUM estimate for Tehran which was calculated before the 2015 FSW survey (Figure 1, Panel A). Both the prior and the data (PSE) fields can be entered as either counts or percentages of the population.

After specifying the general population size and the prior, the user then enters empirical results from PSE methods. At a minimum, the prior and at least one PSE needs to be entered to calculate the Anchored Multiplier. Additional PSE inputs may be entered by clicking the 'ADD SIZE ESTIMATE' button (shown at the bottom of Figure 1, Panel A and Panel B). There is no limit to the number of empirical size estimates that the user can provide.

Drawing upon the random effects meta-regression literature to incorporate additional uncertainty into the calculation, the software also calculates a Variance Adjusted Anchored Multiplier estimate.<sup>14</sup> The variance of true effects/size estimates (the parameter, or estimand,  $\tau^2$ ) is estimated and added to the variance of each individual size estimate, effectively increasing the width of the final 95% credible interval. The variance adjusted estimate will only be calculated when at least two empirical size estimates are provided and when the estimate for  $\tau^2$  is positive, indicating heterogeneity in size estimates that must be accounted for by the calculator. If the estimate for  $\tau^2$  is negative, the value will be set to zero, consistent with the meta-regression literature, and only the Anchored Multiplier estimate will be calculated. Of note, whereas the estimand  $\tau^2$ , cannot be negative because it is a variance component, the estimate of  $\tau^2$  can be negative.<sup>14</sup>

After the user completes all the input fields and submits the data, the program will calculate the Anchored Multiplier estimate and (if there is evidence of heterogeneity in size estimates as indicated by a positive value for  $\tau^2$ ) the Variance Adjusted Anchored Multiplier estimate (Figure 1, Panel C and Panel D). The program generates a forest plot that summarizes the range of estimates provided

by the prior and empirical PSE results, along with estimates from the Anchored Multiplier and Variance Adjusted Anchored Multiplier. Additionally a table is generated, summarizing the data inputs, the beta distribution parameters calculated for each data input and the Anchored Multiplier estimates (as both population percentages and population counts). Both the forest plot and the tables can be exported as .PNG files and .CSV files, respectively. The input data for the Anchored Multiplier calculator is not stored on the website.

After entering the prior and all size estimates from the survey of FSW in Tehran, we estimated by the Variance Adjusted Anchored Multiplier method that FSW make up 1.54% of the adult female population in Tehran (95% credible interval: 1.37–1.72). This estimate closely matches the NSUM estimate, due to the narrowness of the prior's 95% confidence interval relative to the width of the remaining data estimates. As a comparison, in a separate analysis we replaced the prior with a less informative uniform prior covering the range of NSUM estimates for the 13 cities in Iran (0.14–2.44%; lower and upper bounds of the uniform prior). With all other data input the same, we calculated the Variance Adjusted Anchored Multiplier estimate to be 1.71% (95% CI: 1.0–2.59%).

## Discussion

The Anchored Multiplier calculator implements a Bayesian model to synthesize population size information from multiple data sources, including prior knowledge, thus reconciling potentially discrepant estimates to arrive at a single consensus estimate. Although the Bayesian model and the software were originally designed to reconcile discrepant size estimates from multiple implementations of the multiplier method (hence the name, Anchored Multiplier), this approach can also be used to synthesize size estimates from diverse methods.

In practice, researchers and public health programme implementers often aggregate multiple population size methods, using the median of all point estimates and their corresponding lower and upper bounds.<sup>10,12</sup> The rationale behind this approach is that taking the median of all estimates will wash out biases influencing any particular method. However, the implicit assumption is also that all estimates are equally biased (or equally valid), and that they should all be weighted equally when contributing to the final estimate. Our model, in contrast, acknowledges the variation in the degree of certainty around estimates, giving more weight to estimates with narrower intervals. This approach is also distinct from taking a weighted mean where estimates that have wide intervals, and are therefore down-weighted, result in 'pulling' the final estimate

towards zero. Estimates that are down-weighted by the Anchored Multiplier calculator do not pull the final estimate closer to zero. Instead, the posterior distribution is influenced more by the prior distribution. Therefore, when estimates with wide intervals are incorporated into the model, the posterior distribution remains anchored by the prior distribution (as illustrated by the example given in this paper).

As previously discussed, a notable feature of the Anchored Multiplier calculator is that the beta distributions estimated for each size estimate are guaranteed to have at least 70% of the probability mass contained within the lower and upper bounds of the size estimate. If the calculator is unable to fit a beta distribution with this property after 25 iterations of the loop, an error message will indicate which size estimate cannot be fitted to a beta distribution. Depending on the number of size estimates entered, this process should not take longer than 1 min. In the event that the calculator cannot fit a beta distribution to the data, users should re-assess the size estimate in question and how the bounds were originally estimated. Some PSE methods do not calculate lower and upper bounds using traditional formulas for confidence intervals or using percentiles of a probability distribution, possibly resulting in an error when fitting to a beta distribution.<sup>2</sup> If the error persists, users may opt to delete the problematic PSE input and continue the Bayesian synthesis without that size estimate.

Size estimates for the population at risk are fundamental to epidemiology and public health. Common practice in the field recommends implementing multiple PSE methods, as any one particular method may be biased in unpredictable ways.<sup>2</sup> The Anchored Multiplier is a novel approach to synthesize and reconcile multiple estimates, while also allowing for the incorporation of local knowledge. We developed the Anchored Multiplier calculator as a freely available, user-friendly software tool that may be easily adopted by ministries of health, public health departments and public health researchers.

## Funding

This work was supported by the National Institutes of Health [T32 MH19105]. We also received support from the University of California, San Francisco's International Traineeships in AIDS Prevention Studies (ITAPS), U.S. NIMH, R25MH064712.

**Conflict of interest:** None declared.

## References

1. Office of the U.S. Global AIDS Coordinator. *PEPFAR 3.0—Controlling the Epidemic: Delivering on the Promise of an AIDS-free Generation*. Washington, DC: U.S. Global AIDS Coordinator, 2014.

2. Wesson P, Reingold A, McFarland W. Theoretical and empirical comparisons of methods to estimate the size of hard-to-reach populations: a systematic review. *AIDS Behav* 2017;**21**:2188–206.
3. Abdul-Quader AS, Baughman AL, Hladik W. Estimating the size of key populations: current status and future possibilities. *Curr Opin HIV AIDS* 2014;**9**:107–14.
4. Vadivoo S, Gupte MD, Adhikary R *et al*. Appropriateness and execution challenges of three formal size estimation methods for high-risk populations in India. *AIDS* 2008;**22**(Suppl 5):S137–48.
5. International Working Group for Disease Monitoring and Forecasting. Capture-recapture and multiple-record systems estimation. I: history and theoretical development. *Am J Epidemiol* 1995;**142**:1047–58.
6. International Working Group for Disease Monitoring and Forecasting. Capture-recapture and multiple-record systems estimation II: applications in human diseases. *Am J Epidemiol* 1995;**142**:1059–68.
7. Wesson P, Murgai N. Evaluating the completeness of HIV surveillance using capture—recapture models, Alameda County, California. *AIDS Behav* 2018;**22**:2248–57.
8. Khalid FJ, Hamad FM, Othman AA *et al*. Estimating the number of people who inject drugs, female sex workers, and men who have sex with men, Unguja Island, Zanzibar: results and synthesis of multiple methods. *AIDS Behav* 2014;**18**(Suppl 1):S25–31.
9. Quaye S, Raymond H, Atuahene K *et al*. Critique and lessons learned from using multiple methods to estimate population size of men who have sex with men in Ghana. *AIDS Behav* 2015;**19**:16–23.
10. Okal J, Geibel S, Muraguri N *et al*. Estimates of the size of key populations at risk for HIV infection: men who have sex with men, female sex workers and injecting drug users in Nairobi, Kenya. *Sex Transm Infect* 2013;**89**:366–71.
11. Sabin K, Zhao J, Maria J *et al*. Availability and quality of size estimations of female sex workers, men who have sex with men, people who inject drugs and transgender women in low- and middle- income countries. *PLoS One* 2016;**11**:1–11.
12. Raymond HF, Berekyei S, Berglas N, Hunter J, Ojeda N, McFarland W. Estimating population size, HIV prevalence and HIV incidence among men who have sex with men: a case example of synthesising multiple empirical data sources and methods in San Francisco. *Sex Transm Infect* 2013;**89**:383–87.
13. Wesson PD, Mirzazadeh A, Mcfarland W. A Bayesian approach to synthesize estimates of the size of hidden populations: the Anchored Multiplier. *Int J Epidemiol* 2018;**47**:1636–44.
14. Borenstein M, Hedges LV, Higgins JPT, Rothstein HR. *Introduction to Meta-Analysis*. Pondicherry, India: Wiley, 2009.
15. Salganik MJ, Fazito D, Bertoni N, Abdo AH, Mello MB, Bastos FI. Assessing network scale-up estimates for groups most at risk of HIV/AIDS: evidence from a multiple-method study of heavy drug users in Curitiba, Brazil. *Am J Epidemiol* 2011;**174**:1190–96.
16. Guo W, Bao S, Lin W *et al*. Estimating the size of HIV key affected populations in Chongqing, China, using the network scale-up method. *PLoS One* 2013;**8**:e71796.
17. Sharifi H, Karamouzian M, Baneshi MR *et al*. Population size estimation of female sex workers in Iran: synthesis of methods and results. *PLoS One* 2017;**12**:e0182755.
18. Statistical Center of Iran. Selected Findings of the 2016 National Population and Housing Census, 2018. <https://www.amar.org.ir/english/Population-and-Housing-Censuses> (3 May 2019, date last accessed).
19. Greenland S. Bayesian perspectives for epidemiological research: I. Foundations and basic methods. *Int J Epidemiol* 2006;**35**:765–75.
20. Maclehoose RF, Hamra GB. Applications of Bayesian methods to epidemiologic research. *Curr Epidemiol Rep* 2014;**1**:103–09.
21. Greenland S. Bayesian perspectives for epidemiological research. II. Regression analysis. *Int J Epidemiol* 2007;**36**:195–202.