Title
If you are not counted, you don't count: Estimating the size of hidden populations

Author
Wesson, Paul Douglas

Publication Date
2016

Peer reviewed|Thesis/dissertation

If You Are Not Counted, You Don't Count: Estimating The Size of Hidden Populations

by

Paul Douglas Wesson

A dissertation submitted in partial satisfaction of the

requirements for the degree of

Doctor of Philosophy

in

Epidemiology

in the

Graduate Division

of the

University of California, Berkeley

Committee in charge:

Professor Arthur Reingold, Co-chair
Professor William McFarland, Co-chair
Professor Nicholas Jewell
Professor Mark Wilson

Spring 2016

If You Are Not Counted, You Don't Count: Estimating The Size of Hidden Populations

Abstract

If You Are Not Counted, You Don't Count: Estimating The Size of Hidden Populations

by

Paul Douglas Wesson

Doctor of Philosophy in Epidemiology

University of California, Berkeley

Professor Arthur Reingold, Co-chair

Professor William McFarland, Co-chair

Background: Despite advances in treatment and prevention services, HIV infection remains a leading cause of morbidity and mortality worldwide, identified by the 2010 Global Burden of Disease report as the fifth leading cause of global disability adjusted live years. While the epidemiologic features of HIV infection vary globally, marginalized populations, such as men who have sex with men (MSM), female sex workers (FSW), and injection drug users (IDUs) are consistently at increased risk for HIV infection relative to the general population. Targeting such marginalized, or hidden, populations has become a global priority to maximize the effectiveness of the public health response to the HIV pandemic. Members of these populations are often difficult to find, and the size of these populations is largely unknown, making it difficult to calculate epidemiologic measures of disease and to evaluate the reach and coverage of public health programs.

Methods: Through three separate analyses, this dissertation will investigate the reliability and the plausibility of population size estimation methods when applied to hidden populations. Chapter 1 systematically reviews the literature on population size estimation methods and assesses the degree to which different methods, applied to the same population, calculate similar estimates of the target population. Chapter 2 evaluates a novel size estimation method, the SS-PSE, by applying it to a Respondent-Driven Sampling study of African-American MSM in San Francisco, and comparing results to other methods. Chapter 3 applies capture-recapture models to evaluate the completeness of the Alameda County HIV surveillance system, and examines the role of sampling bias in this application.

Discussion and Significance: The results of the analyses featured in this dissertation demonstrate that variability in population size estimates from different size estimation methods is common, though often unaddressed. Population size estimation is fundamental to public health surveillance, serving as the basis for policy decisions and quantifying the magnitude of disease. To produce reliable population size estimates, which have implications for the allocation of limited public health resources to marginalized populations, investigators should consistently apply multiple size estimation methods and carefully consider the influence of sampling bias.

To my family

Thank you for your unending support through the good times and the difficult times, and for all the laughs in between. You keep me grounded.

Contents

List of Figures

List of Tables

Acknowledgments

I would first like to thank my committee chair and academic advisor at UC Berkeley, Dr. Arthur Reingold, who always believed in my potential and treated my success as a foregone conclusion. I am extremely grateful for all of his efforts to make my research pursuits possible. As his advisee, I never heard "no", but rather, "no one has ever done that before, but I dont see any reason why you can't." I also heard "stay focused."

I am also extremely grateful to my co-advisor and mentor, Dr. Willi McFarland, who first introduced me to the topic of my dissertation, stressed its global importance, and made me excited to be a part of this area of research. As my mentor, Willi has reminded me to look for the silver lining in every situation. The balance he strikes between family-life, a career with global reach, and surfing continues to inspire me.

I would also like to thank my other committee members, Drs. Nicholas Jewell and Mark Wilson, who both emphasized a step-wise, methodical, approach to addressing research questions. I admire (and strive to emulate in my own career) how they are both able to communicate complicated scientific methods in a way that is easy for many to understand, regardless of background.

In addition to what I gained from my dissertation committee, my tenure at UC Berkeley was greatly enriched by my relationship with my unofficial mentors: Dr. Barbara Abrams, Dr. Jennifer Ahern, Dr. Michael Bates, the late Dr. Patricia Buffler, Dr. Ray Catalano, and Dr. Mahasin Mujahid. Our formal and informal interactions have taught me so much about scientific inquiry, scientific rigor, and prioritizing self-care. I am fortunate to have had my scientific and professional perspective shaped by their own. I will always remember Dr. Catalano distilling the importance of our discipline: "It is our job [as epidemiologists] to take that which offends us and to shine a light on it."

Thank you, also, to the staff in the Epidemiology office (especially Janene Martinez), whom I have always known to be exceptionally friendly and helpful.

Chapter 4 of this dissertation would not have been possible without the enthusiastic support of the Alameda County Public Health Department HIV Epidemiology and Surveillance Unit; specifically: Dr. Neena Murgai, Richard Lechtenberg, Oliver Heitkamp, George Banks, Jesus Altamirano, and Dr. Janet Tang.

I am especially grateful to my classmates, most notably: Rose Kagawa, Irene Headen, Alison Cohen, Caitlin Gerdts, Divya Vohra, Karl Pope, Ann Webber, Benjamin Chaffee, Anna Decker, and Deb Karasek. I have learned so much from being taught by you, teaching with you, and debating with you. Most importantly, you have all made this journey fun. I look forward to collaborating with you in the future.

Finally, I am deeply grateful to my parents, Oliver and Barbara Wesson, who sat me down when I was five years old to watch V: The Final Battle. From that moment, I dreamed of becoming a scientist.

Chapter 1

Introduction

1.1 Motivation

Despite advances in treatment regimens and prevention strategies, HIV remains a leading cause of morbidity and mortality worldwide. The 2010 Global Burden of Disease (GBD) report identified HIV/AIDS as the fifth leading cause of global disability adjusted life years (DALYs), third among communicable disease disorders.[42] This increase in ranking from the 1990 GBD ranking for causes of DALYs highlights the shortcomings of public health action to effectively control this disease.

Globally, populations engaging in certain risk behaviors are especially vulnerable to HIV infection. These key populations include female sex workers (FSW), men who have sex with men (MSM), and injection drug users (IDUs), among others. Due to biological, behavioral, and structural vulnerabilities, the prevalence of HIV infection is typically higher in these groups than in the general population. Stigma-related discrimination, imprisonment, and violence contribute to reasons why members of these groups may not generally disclose their identity, which poses a challenge to public health efforts targeting these groups for health interventions. Thus, these groups are often referred to as hidden populations.[37]

Reaching these hidden populations has been identified as a key strategy for achieving maximum effectiveness in the public health response to HIV infection.[45, 52] In December 2014, phase III of the Presidents Emergency Plan For AIDS Relief (PEPFAR), an international collaboration to control the HIV pandemic, initiated a pivot strategy that directs resources to where HIV infection is concentrated in order to achieve the greatest impact of the treatment and prevention programs funded by PEPFAR.[8] Hidden populations were identified as being among the key populations where HIV infection is concentrated, and thus where resources must be invested.

Estimating the size of hidden populations is an emerging priority for global HIV prevention. Enumerating key populations allows epidemiologists to quantify the burden of disease in these populations through estimation of the population prevalence of HIV infection and calculating an attributable fraction. Furthermore, population size estimates are useful parameter inputs for modeling population dynamics contributing to disease transmission through sexual and social networks. Finally, population size estimates contribute to the evaluation of programs targeting at-risk populations with respect to their reach and coverage.

A complete census of hidden populations is logistically implausible due to the very nature of a population being hidden; therefore, the size of these populations must be estimated from samples of the hidden population.[1, 61] The breadth of population size estimation methods available to public health researchers often depends on representative samples in order to draw unbiased inferences of the target population. In many surveillance studies of hidden populations, the available data are neither complete nor a random (representative) sample.[33, 41] Bias in the sampling process may occur if certain members of the target population preferentially select into the sampling process, while other segments of the target population avoid selection. The characteristics of the study population then do not reflect the characteristics of the target population. While public health researchers benefit from a wide range of options to estimate the size of a population, the current public health literature on population size estimation methods mainly distinguishes between methods with respect to their implementation; that is, the data structure required to apply a particular method.[61] Less focus is given to the agreement in population size estimates calculated from these different methods, and their performance, relative to each other, under biased sampling conditions. Differences in estimates calculated from different population size estimation methods applied to the same population compromise the reliability of any one method to produce estimates from which public health policy is based. Reliable estimates of the sizes of hidden populations are needed in order to responsibly invest in programs targeting key populations and allocating limited public health resources to curb the HIV epidemic.

The purpose of this dissertation is to describe and evaluate the range of methods applied in public health to estimate the size of hidden populations. I begin by assessing the degree to which different methods produce similar estimates of the same population; noting how the methods account for (or leverage) the bias that may exist in the sampling process. I then apply a selection of relevant population size estimation methods and modeling techniques to hidden populations of public health importance to the HIV epidemic in the San Francisco Bay Area. From this direct application I provide insight into approaches to produce plausible estimates of the sizes of the target populations, while accounting for biases in the sampling

process.

Population size estimation methods are a useful set of tools for public health researchers to monitor and quantify disease in a population. Despite the breadth of options available, the degree to which different population size estimation methods produce estimates that are in agreement with each other, is unknown. Current guidelines for population size estimation do not address how methods may be differentially impacted by biased sampling, thus producing estimates of the same population size that are not in agreement with each other. Addressing this gap in the literature, while also evaluating methods that can recover plausible size estimates despite biased sampling, is necessary to improve public health surveillance of hidden and hard-to-reach populations.

Chapter 2

Theoretical and empirical comparisons of methods to estimate the size of hard-to-reach populations: A systematic review

2.1 Introduction

Obtaining an accurate count of human populations is fundamental to the distribution of resources. A census of the general population is a difficult task, complicated by the likelihood of undercounting segments of the populations that are considered hard-to-reach, possibly leading to a selection bias in the final count. This problem is particularly salient in public health, where hard-to-reach populations, sometimes termed hidden populations, are defined by behaviors, identities, or characteristics that lead to stigmatization and discrimination. These hidden populations (e.g., female sex workers [FSW], men who have sex with men [MSM], and people who inject drugs [PWID]) often face a disproportionate risk of HIV infection and of other sexually transmitted infections compared to the general population. Estimating the size of these populations is often necessary for allocation of scarce public health resources, for evaluating the reach and coverage of social services targeting these populations, and for using the appropriate denominator in epidemiologic measures of disease burden and incidence. Population size estimation (PSE) of key populations at risk for HIV infection is a priority for international organizations, such as the World Health Organization (WHO), and organizations that finance global health programs, such as the Presidents Emergency Plan For AIDS Relief (PEPFAR). In December 2014,

phase III of PEPFAR initiated a pivot strategy that directs resources to where the HIV epidemic is concentrated in order to achieve the greatest impact of the treatment and prevention programs it funds.[8] As a result, reliable estimates of the sizes of key populations are necessary to set program targets, assess their coverage, and evaluate their impact.

Many methods are available for counting the hidden, hard-to-reach, and unobserved with origins from diverse disciplines, including sociology, population biology, and natural resource management. In public health, population size estimation methods are often imported from these other fields and adapted for human populations. In 2010, the UNAIDS/WHO updated their guidelines concerning PSE methods appropriate for key populations. The document identified five PSE methods (i.e., census and enumeration, capture-recapture, multiplier, population surveys, and network scale-up) and discussed their data requirements, implementation procedures, and limitations.[60] Shortly after, Abdul-Quader et al. published a literature review of PSE methods commonly used to estimate the size of key populations in studies published in the peer-reviewed literature between 2011 and 2013.[1] The authors summarized six PSE methods (those listed above and single survey based methods, such as the successive sampling population size estimation method [18]), highlighting their theoretical and practical strengths and limitations, and making recommendations for their implementation based on lessons learned from individual articles that served as case studies. The guidelines and the literature review provide a menu of options from which public health researchers can choose when attempting to estimate the size of a hidden population based on the availability of the appropriate data structure and resources.

However, such a menu implies that any option will produce a consistent estimate of the true population size. Salganik et al. highlighted the questionability of this assumption when they applied the network scale-up method to estimate the size of heavy drug users in Curitiba, Brazil. They noted that estimates from this method, with and without statistically correcting for assumption violations, were significantly different from estimates derived using alternative PSE methods.[50] The discrepancy presents a dilemma in choosing which result is correct in the absence of a gold standard. Other researchers apparently embrace variability in estimates from multiple PSE methods, arguing that different methods are likely prone to different biases and therefore using a central estimate (usually a median) can reduce the potential for severe bias when relying solely on a single method.[44, 47] No study has explicitly explored the variability in size estimates from multiple PSE methods, formally assessing the reliability of these methods, and determining if any methods appear to consistently under- or over- estimate compared to others in different populations.

Our objective was to assess the extent to which different PSE methods provide

the same estimate of a target population by systematically reviewing peer-reviewed studies that used at least two PSE methods to estimate the size of the same target population. In doing so, we build upon previous work by identifying additional methods not included in past reviews, make side-by-side empirical comparisons of results, and summarize the consistency of estimates for the size of the same target populations. For a more detailed description of each PSE method we refer the reader to the UNAIDS 2010 guidelines, the review by Abdul-Quader et al, or citations provided for additional methods.

2.2 Methods

Between February 6, 2015 and March 6, 2015 we conducted a systematic review of the peer-reviewed literature on population size estimation methods. Articles published by March 6, 2015 were eligible for inclusion in the review. As a primary inclusion criterion, studies were required to include at least two PSE methods that allow for a comparison of the resulting estimates of the same population. Alternative modeling approaches for the same PSE method were treated as distinct methods; for example, the network scale-up method and the generalized network scale-up method were treated as distinct methods. Of the studies that provided the results obtained using multiple methods, studies were included only if the research team designed and implemented the methods that were used. This eligibility criterion addressed the concerns that different research teams may use different definitions of the target population when estimating its size, and that temporal differences may affect the population being enumerated (i.e., if different research teams estimated the same population, but at different points in time). Furthermore, articles were eligible for inclusion only if they included empirical, as opposed to simulated, data (as a result, mathematical models were excluded); estimated the size of a human population; and estimated the size of a current population (rather than make a projection into the future or the past). Review papers were excluded.

In order to obtain a comprehensive sample of studies we used several search strategies. Two of the three search strategies used the PubMed database. The first strategy (PubMed1) searched using the term population size estimation. Studies were filtered to include only human populations. No language filter was used. The second strategy (PubMed2) used the search terms ((estimat* prevalence) OR (estimat* incidence) OR (estimat* size population)). Due to the large number of results matching this query, even after filtering for studies that included human populations (n=184,530) and using the PubMed function for ordering results by relevance, we reviewed studies by title until we encountered an interval of 100 titles that were not

relevant to the review. Reviewing titles for PubMed2 ended after this interval was reached. The third search strategy used Google Scholar, employing the following search terms: men OR women OR children population size estimation. Results were filtered for Articles and sorted by relevance. No temporal restrictions were specified.

After results from the three search strategies were aggregated, duplicates were identified and removed. The remaining de-duplicated results were assessed for relevance by reviewing the abstract. Based on the review of abstracts, articles considered not relevant were removed from the pool of studies for potential review. For remaining studies, the full text was read, and based on that reading additional studies were determined to be ineligible. For the final set of studies judged to be eligible for inclusion, the following data were abstracted: target population, setting (geographic/location), time period (year(s)), study design/sampling method, study population, sample size, population size estimation method, point estimates and confidence intervals, discussion of bias, and statistical corrections.

Studies were not given a rating according to quality because the field of population size estimation in epidemiology is still relatively new and recommendations for standardized reporting have not been agreed upon or published. Additionally, the guidelines for documenting the size estimation process put forward by UNAIDS/WHO in their 2010 guidelines were published after a subset of the eligible studies included in this review were published.[59] Because heterogenous populations were estimated across studies, we determined it to be inappropriate to conduct a meta-analysis to statistically synthesize results across studies. Instead, we chose to visualize the data, using the reported population size estimates and upper and lower bounds in order to present the level of agreement with which standard and novel methods in this field estimate the size of the same target population. Dot plots were generated using R.[46]

2.3 Results

The search strategies generated 341 unduplicated, potentially eligible publications (Figure 2.1). Of the 341 abstracts screened for possible inclusion in the review, 65 were deemed relevant to the objective of this study and underwent full text review. Based on the full text review, 25 of these 65 were considered eligible for inclusion. The remaining articles were ineligible primarily due to not including a second size estimation method implemented by the authors for comparison.

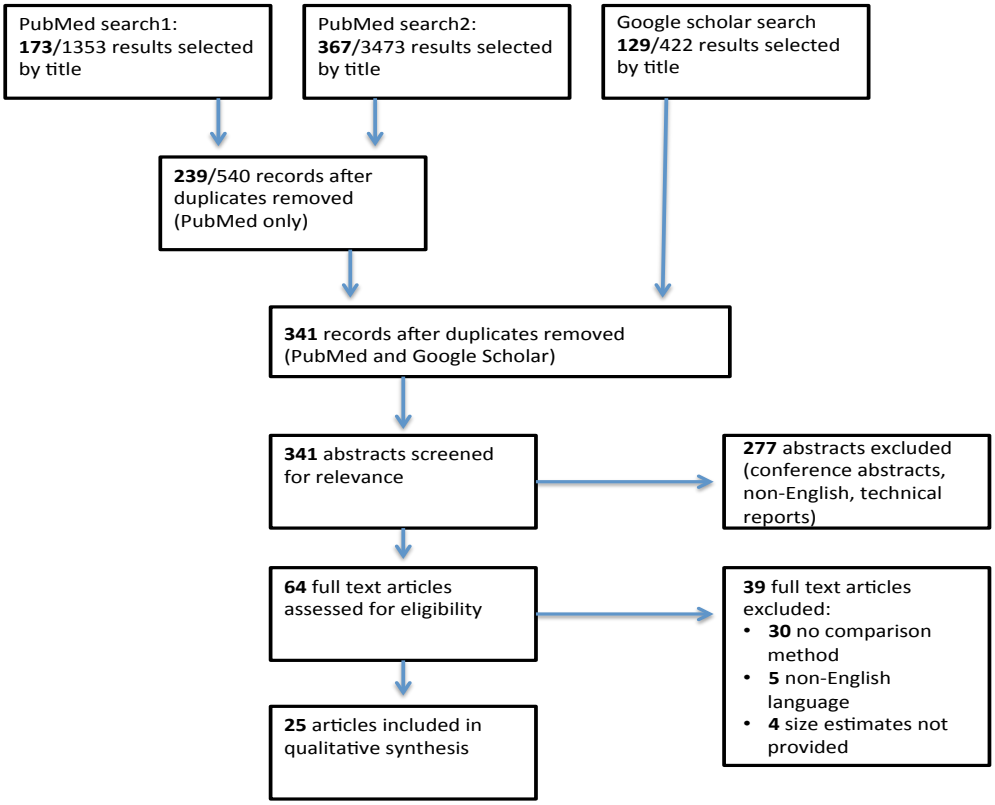


Figure 2.1: Flow diagram of study selection

Across the 25 studies included in the review, covering 16 countries, 21 unique PSE methods were featured. Of these 21 PSE methods, 15 were considered generalizable methods; that is, estimation approaches that could be applied in settings other than that in which the study took place, given the availability of appropriate data (Table 2.1). The most common method included in the review was the service multiplier method (n=21 studies), followed by capture-recapture (i.e., multiple systems estimation) (n=11 studies), and the unique object multiplier method (n=8 studies). Although no restrictions were placed on the type of population examined during the initial literature search, all populations in the final sample of eligible studies were key populations at risk for HIV; people who inject drugs (PWID) and problem drug users were the most commonly estimated population among studies included in the review.

Table 2.1: Generalizable population size estimation (PSE) methods and sources of bias

	Target population				Methodological considerations		
PSE method	Female sex workers	Men who have sex with men	People who inject drugs/problem drug users	Other[1]	Data source	Common sources of bias	Statistical adjustment for bias
Service multiplier	(Khalid et al., 2014) (Mutagoma, Kayitesi, Gwiza, & Ruton, 2014) (Okal et al., 2013) (Vadivoo et al., 2008) (L. G. Johnston et al., 2013)	(Archibald et al., 2001) (Livak et al., 2013) (Khalid et al., 2014) (Jing, Qu, Yu, Wang, & Cui, 2014) (Quaye et al., 2015) (Okal et al., 2013) (Luan et al., 2005) (Vadivoo et al., 2008) (Raymond et al., 2013)	(L. Johnston et al., 2011) (Archibald et al., 2001) (Salganik et al., 2011) (Sawitri, Blogg, & Angela, 2012) (Okal et al., 2013) (Kimber, Hickman, Degenhardt, Coulson, & van Beek, 2008) (L. G. Johnston et al., 2013)		-Count of attendees at service during time interval (benchmark) - Representative population survey (multiplier)	-Non-representative population survey -Duplicate counts from service (counting visits not people) -Inaccurate recall among survey participants -Inconsistent eligibility criteria between service (benchmark) and survey (multiplier) -Source dependency between benchmark and multiplier	

	Target population				Methodological considerations		
PSE method	Female sex workers	Men who have sex with men	People who inject drugs/problem drug users	Other	Data source	Common sources of bias	Statistical adjustment for bias
Unique object multiplier	(Khalid, Hamad, & Othman, 2014) (Vadivoo et al., 2008)	(Khalid et al., 2014) (Quaye, Raymond, & Atuahene, 2015) (Vadivoo et al., 2008)	(L. Johnston, Saumtally, Corceal, Mahadoo, & Oodally, 2011) (Khalid et al., 2014) (Vadivoo et al., 2008)		-Count of unique objects distributed (benchmark) - Representative population survey (multiplier)	-Non-representative population survey -Unable to track number of objects actually distributed to population -Non-random distribution of objects -Inconsistent geographic coverage and eligibility for benchmark and multiplier	

	Target population				Methodological considerations		
PSE method	Female sex workers	Men who have sex with men	People who inject drugs/problem drug users	Other	Data source	Common sources of bias	Statistical adjustment for bias
Capture-recapture (multiple systems estimation)	(Khalid et al., 2014) (Mutagoma et al., 2014)	(Khalid et al., 2014) (Luan et al., 2005) (Chen et al., 2013)	(Archibald et al., 2001) (Khalid et al., 2014) (Xu, Fyfe, Walker, & Cowen, 2014) (Hope, Hickman, & Tilling, 2005) (Kimber et al., 2008) (Kuhnert & Bohning, 2009)		≥ 2 independent lists of target population members with information to identify the same individual across lists	-Source dependency -Open population (especially if notable temporal distance between lists) -Inaccurate matching across lists (misclassification of subjects)	-Huggins or Pledger models to account for source independence assumption in 2 sources (Xu et al. 2014) -Regression model with covariates to control for heterogeneity in capture probabilities (Hope et al., 2005) -Stratify by measured covariates (Kimber at al., 2008)

	Target population				Methodological considerations		
PSE method	Female sex workers	Men who have sex with men	People who inject drugs/problem drug users	Other	Data source	Common sources of bias	Statistical adjustment for bias
Single source capture-recapture			(van der Heijden, Cruts, & Cruyff, 2013), (Cruyff & van der Heijden, 2008)	Smit, Reinking, & Reijerse, 2002)	Repeated visit on a single source	-Underestimation if not true that people with 1 or 2 visits are similar to people with 0 visit -Unobserved heterogentiy in capture probabilities -Change in individual Poisson parameters may be related to prior capture	-Stratify by measured covariates (Smit et al, 2002) -Include covariates in regression model to control for heterogeneity (van der Heijden et al., 2013) (Cruyff & van der Heijden, 2008)

	Target population				Methodological considerations		
PSE method	Female sex workers	Men who have sex with men	People who inject drugs/problem drug users	Other	Data source	Common sources of bias	Statistical adjustment for bias
Network scale-up	(Guo et al., 2013)	(Jing et al., 2014) (Guo et al., 2013)	(Salganik et al., 2011) (Guo et al., 2013)	(Guo et al., 2013)	Representative population survey	-Respondents misunderstand question and include people in network who do not fit eligibility criteria -Respondents do not know that members of their network belong to the target population (transmission error)	-Augment with survey of the target population to account for transmission error and differential network size (Generalized Network Scale-Up) (Jing et al., 2014) (Salganik et al., 2011)

	Target population				Methodological considerations		
PSE method	Female sex workers	Men who have sex with men	People who inject drugs/problem drug users	Other	Data source	Common sources of bias	Statistical adjustment for bias
						-Differential network size for target population members and the general population -Study sampling frame less likely to include people who know certain segments of the target population in their networks.	-Scaled social respect factor to adjust for likelihood that survey participants hesitate to report knowing member of target population (Guo et al., 2013)
Wisdom of the crowds	(Khalid et al., 2014) (Okal et al., 2013)	(Khalid et al., 2014) (Quaye et al., 2015) (Okal et al., 2013)	(Khalid et al., 2014) (Okal et al., 2013)		Target population-based survey	-Subjectivity -Respondents may misunderstand question	

	Target population				Methodological considerations		
PSE method	Female sex workers	Men who have sex with men	People who inject drugs/problem drug users	Other	Data source	Common sources of bias	Statistical adjustment for bias
Delphi	(Khalid et al., 2014)	(Archibald et al., 2001) (Livak et al., 2013) (Khalid et al., 2014)	(Archibald et al., 2001) (Khalid et al., 2014)		Local stakeholders and community experts working with target population [Optional] Delphi members revise initial estimate after seeing data from other population size estimation methods	Subjectivity and panel member biases	
Program estimate	(Vadivoo et al., 2008)	(Vadivoo et al., 2008)	(Vadivoo et al., 2008)		Estimates provided by program personnel working with target population	-Double counting among programs/services (unique individual attends multiple programs) -Misclassification (non-target population members included in count)	

	Target population				Methodological considerations		
PSE method	Female sex workers	Men who have sex with men	People who inject drugs/problem drug users	Other	Data source	Common sources of bias	Statistical adjustment for bias
						-Program funding linked to population size estimates	
General population survey (simple proportion)		(Archibald et al., 2001) (Livak et al., 2013) (Raymond et al., 2013)	(Archibald et al., 2001) (Salganik et al., 2011)		Representative survey of the general population	-Potential underestimation of self-report of sensitive behaviors -Population survey not representative -Non-sampling error -Differential non-response	
Mapping and enumeration		(Mutagoma et al., 2014)	(Quaye et al., 2015)		Mapping hotspots within catchment area	-May miss clandestine venues -Does not include people who do not frequent selected venues -Stigmatized populations may alter frequency of attending known venues to avoid discrimination Differential non-response	

	Target population				Methodological considerations		
PSE method	Female sex workers	Men who have sex with men	People who inject drugs/problem drug users	Other	Data source	Common sources of bias	Statistical adjustment for bias
Reverse tracking method	(Vadivoo et al., 2008)	(Vadivoo et al., 2008)			-Count of target population at venues/clusters during mapping -Count of target population at venues/clusters at second visit (population survey using venue-based sampling)	-Does not account for populations that are hidden, mobile, or do not visit the cluster on the day of the survey -Overestimation of the population size if individuals frequent multiple venues/clusters	
Literature review	(Khalid et al., 2014) (Okal et al., 2013)	(Khalid et al., 2014) (Quaye et al., 2015) (Okal et al., 2013)	(Khalid et al., 2014) (Okal et al., 2013)		Published size estimates of target population from similar countries/ geographic region	Publication bias (estimates perceived by researchers as too low or high)	

	Target population				Methodological considerations		
PSE method	Female sex workers	Men who have sex with men	People who inject drugs/problem drug users	Other	Data source	Common sources of bias	Statistical adjustment for bias
Ratio estimation		(Raymond et al., 2013)	(Hope et al, 2005)		Ratio of absolute count or proportion of a known population to an unknown population	Invalid estimates of the counts or proportions for known populations	
Modified LMS (Laska, Meisner and Siegel)		(Chen et al., 2013)			Survey measuring frequency with which respondent has visited venue (tangible or internet) in past K weeks	-Misidentification of visible target population members at tangible venues -Misclassification of target population members at tangible and internet venues -Inadequate recall	

	Target population				Methodological considerations		
PSE method	Female sex workers	Men who have sex with men	People who inject drugs/problem drug users	Other	Data source	Common sources of bias	Statistical adjustment for bias
Nominator method			(Sawitri et al., 2012)		Individual members of the target population name other target population members also in their social network according to an investigator-specified criteria (i.e., utilized a service, died during a time interval)	Poor recall by study participants	

Table 2.1 also outlines the common sources of bias described by the eligible studies for each of the generalizable PSE methods. Most sources of bias pertain to violations in the theoretical assumptions for valid estimation. In several cases, as noted in the table, investigators attempted to account for and correct the bias through statistical adjustment. Capture-recapture and the network scale-up were the only two PSE methods for which statistical adjustment for potential bias was observed among articles included in this review.

Many of the PSE methods use a single formula to estimate the size of the hidden population. For example, the size of the hidden population according to the multiplier methods (service multiplier, unique object multiplier, and simple capture-recapture) is calculated as the benchmark (an absolute count of the hidden population meeting a specified criteria) divided by the multiplier (a proportion of the hidden population from a second, representative, sample that matches the same criteria as the benchmark). For some PSE methods, such as adjusted capture-recapture and network scale-up, multiple modeling approaches have been developed to estimate the unobserved population and to correct for violations in theoretical assumptions that may otherwise result in biased estimates. These modeling approaches use additional data or specify alternative distributions to adjust for potential biases or account for violations in the underlying assumptions for the PSE method. Six studies in this review included only alternative estimators for the same method for comparison; that is, a single PSE method was implemented, but estimates from multiple models were calculated. For example, Smit et al. compared the Zelterman estimator and the Chao estimator, alternative analysis approaches to the same capture-recapture method, to estimate the number of potential clients using a facility for homeless people. The authors concluded that had the estimators produced statistically different estimates, this would indicate heterogeneity in the capture probabilities and the Zelterman estimator should be preferred. For this study, the confidence interval of the Chao estimator overlapped the point estimate of the Zelterman, suggesting homogeneity in the capture probabilities for the target population.

Across the 25 studies, the sizes of 80 target populations were estimated using various PSE methods. Two of the target populations estimated included methods specific to that context, and thus not generalizable to broader applications of size estimation for any population. The results of the remaining 78 size estimations are depicted graphically in Figure 2.2. Corresponding tables of exact point estimates and confidence intervals are detailed in the table in the Appendix. Overall, dot plots of the estimated sizes of target populations demonstrated substantial variability in the estimates produced by different methods. For 18 of the 80 target populations,

[1] potential clients for homeless facility, clients of female sex workers

population size estimates were consistent across all PSE methods implemented; that is, there was overlap of the upper and lower bounds associated with individual point estimates across all PSE methods. For some methods, these bounds were calculated as 95% confidence intervals; in other methods, such as the wisdom of the crowds and Delphi methods, these upper and lower bounds represented the maximum and minimum guesses of the sample surveyed. The majority (n=15) of these 18 target populations included one or more methods with intervals that were extremely wide, sometimes with a width greater than 20,000 people for a point estimate of 4,300, as was the case for PWID in Montreal(Archibald et al., 2001).

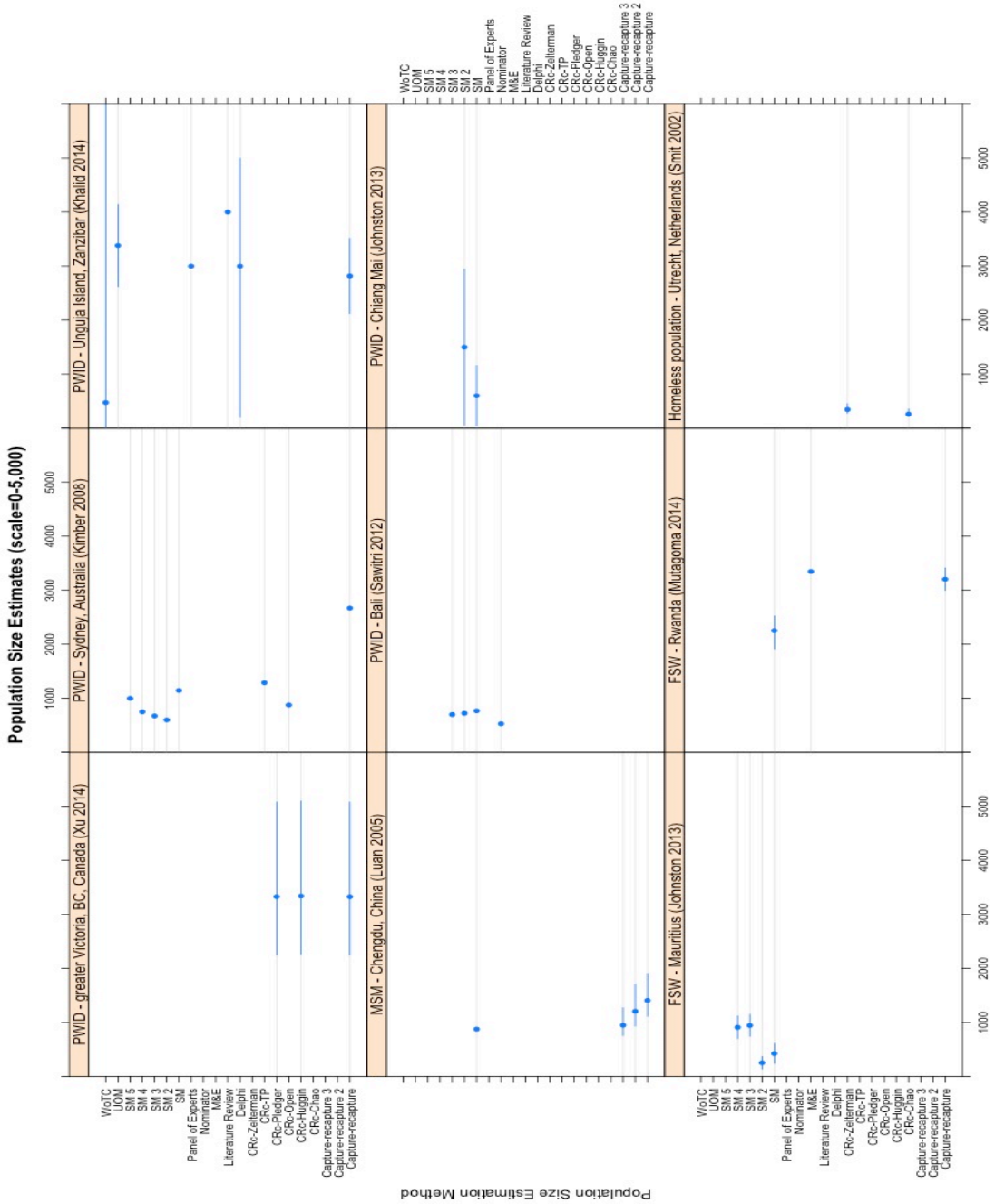
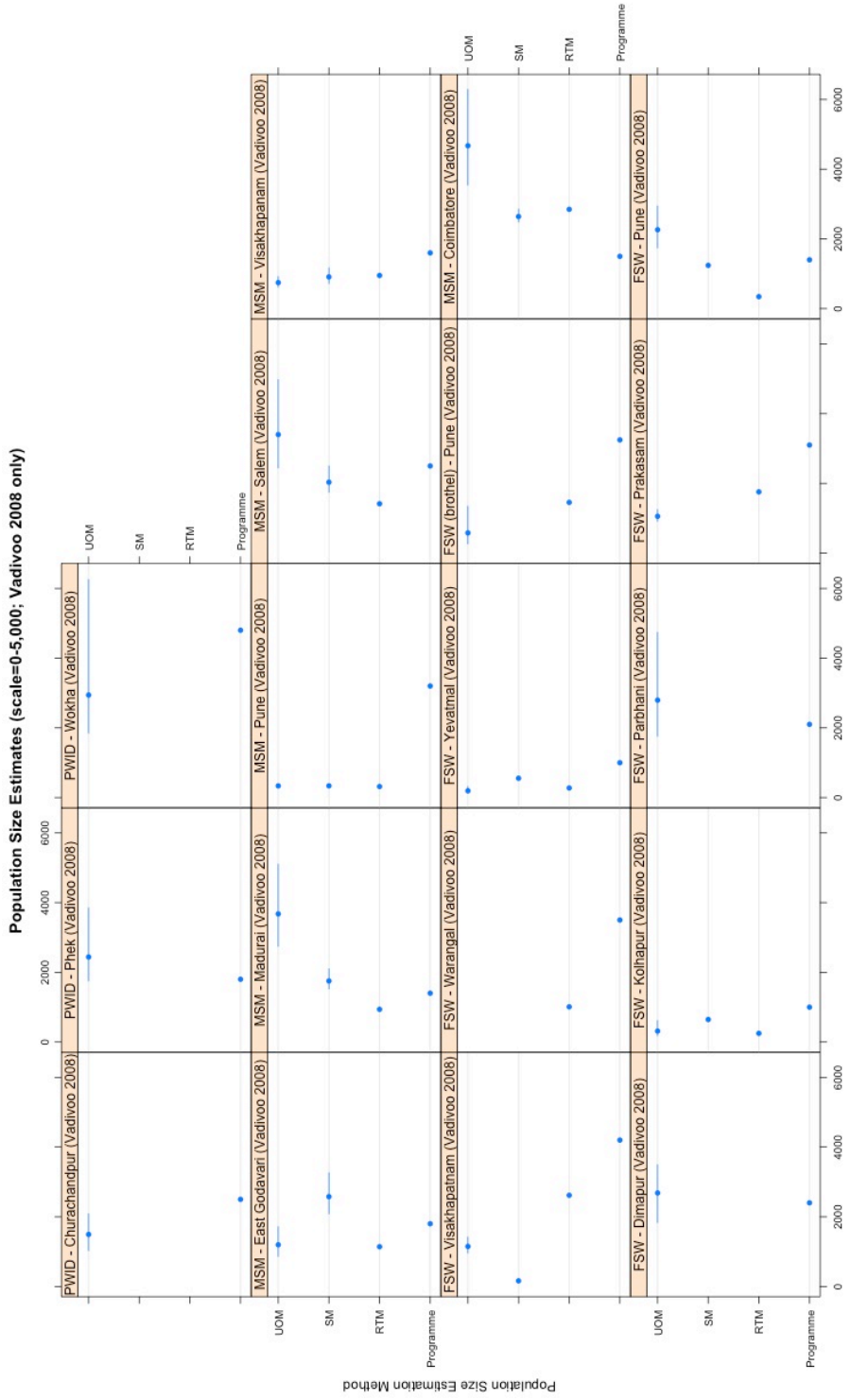
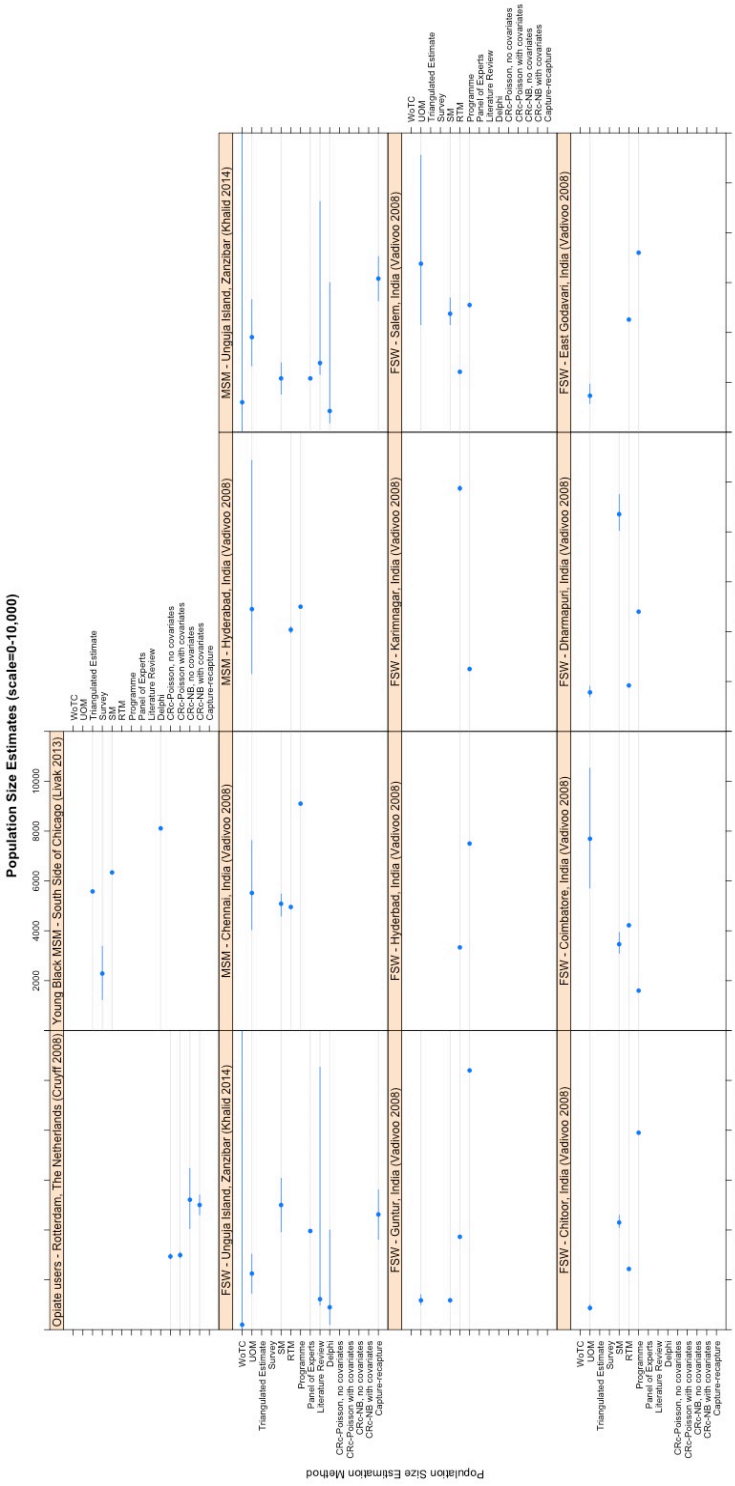


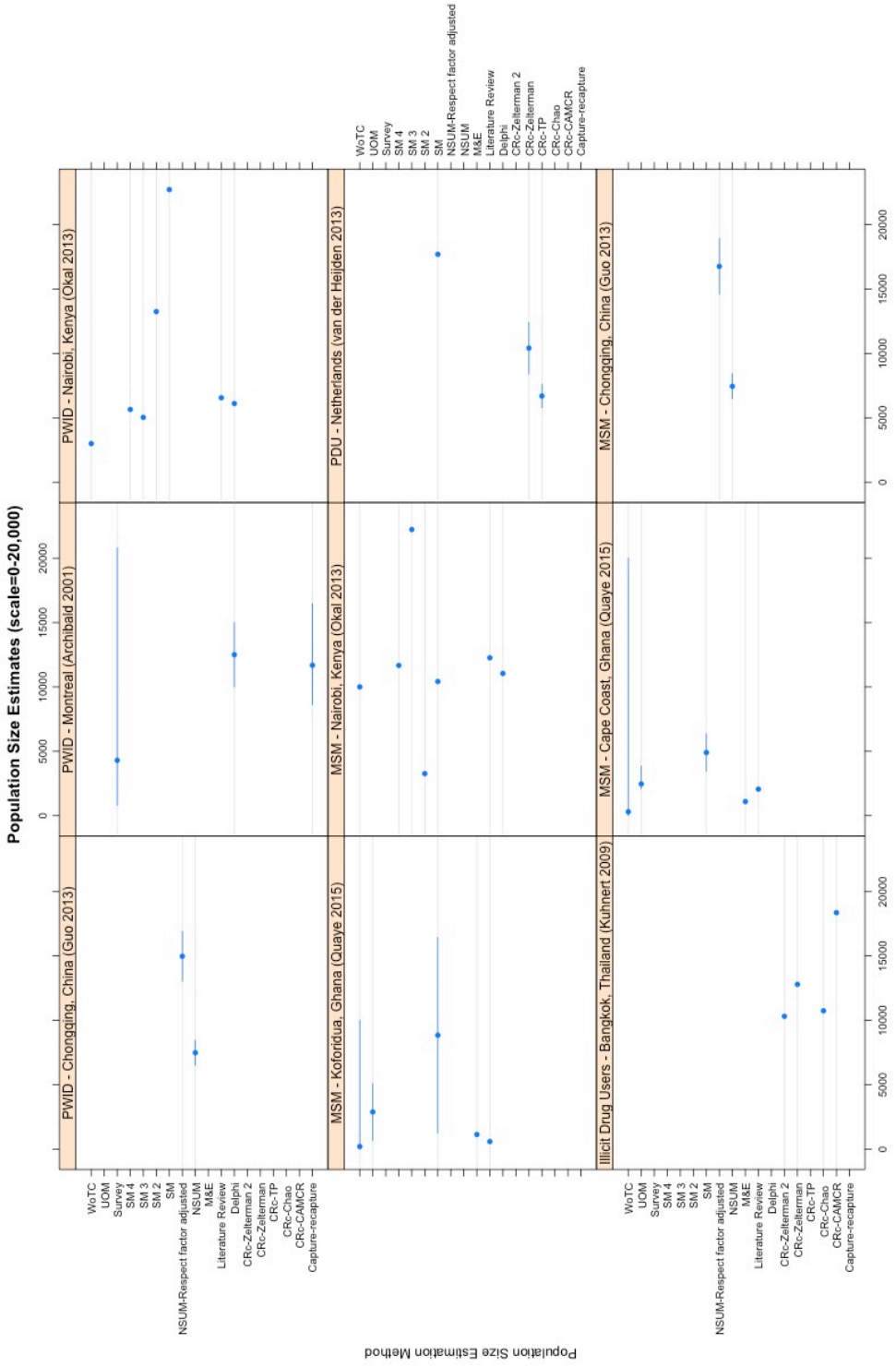
Figure 2.2: Comparison of Population Size Estimation methods grouped by target population and scale of estimated population size.[2,3]

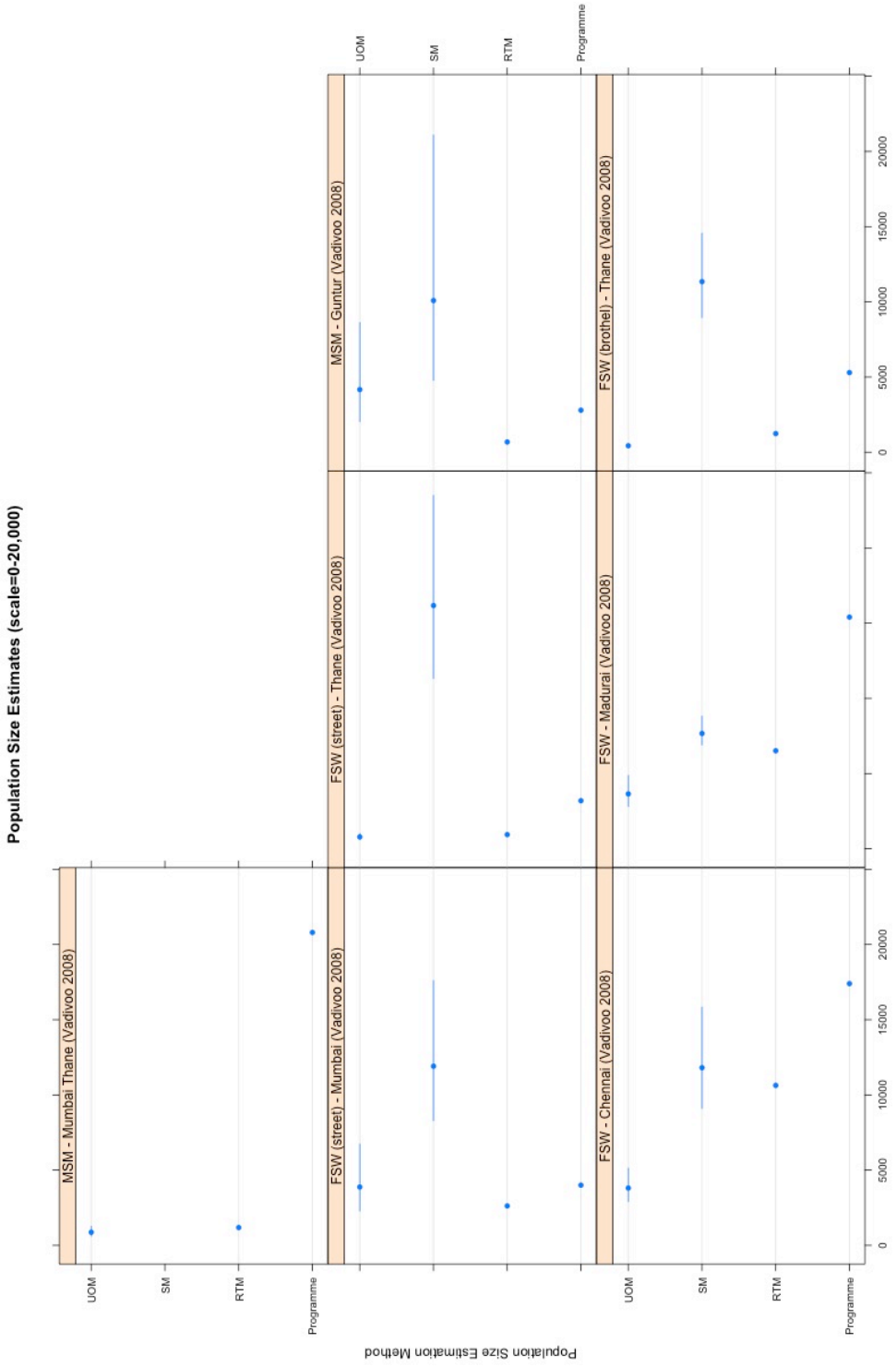
[2]Abbreviations: FSW=female sex workers; MSM = men who have sex with men; PWID = people who inject drugs; PDU = problem drug users; WoTC = wisdom of the crowds; UOM = unique object multiplier; SM = service multiplier; M&E = mapping and enumeration; CRc=Zelterman = Capture-recapture using Zelterman estimator; CRc-TP = Capture-recapture using truncated Poisson; CRc-Pledger = Capture-recapture using Pledger's method; CRc-Huggin = Capture-recapture using Huggin's method; CRc-Open = Open population capture-recapture; RTM = reverse tracking method; Programme = program estimate; Survey = general population survey; CRc-Poisson = capture-recapture with Poisson distribution; CRc-NB = capture-recapture with negative binomial distribution; CRc-Chao = capture-recapture using Chao estimator; CRc-CAMCR= capture-recapture using mixture model; NSUM = network scale-up; LMS* = modified LMS.

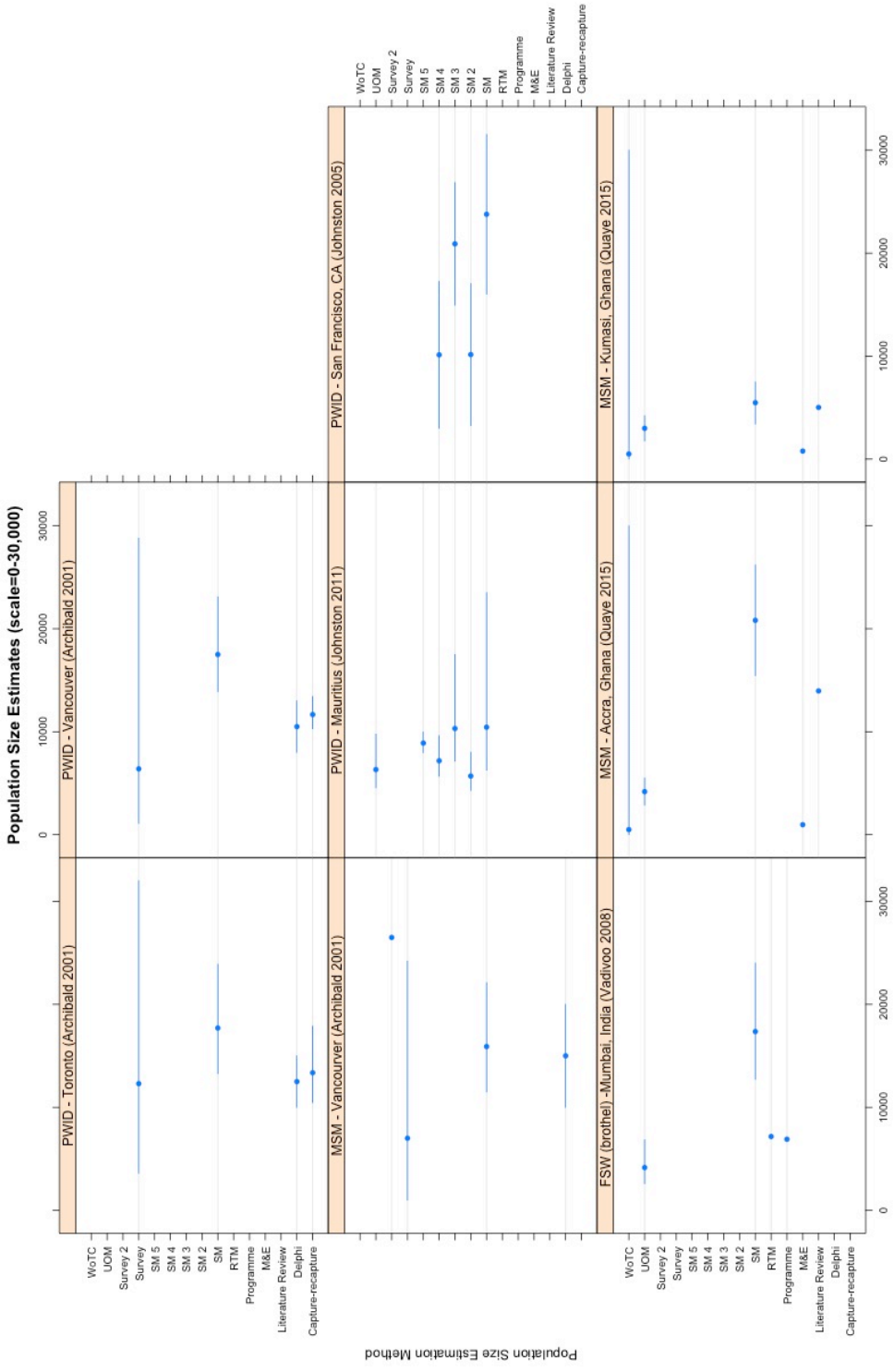
[3]Numerical indexing of method (i.e. SM 2) refers to multiple implementations of the same method, using different data inputs.

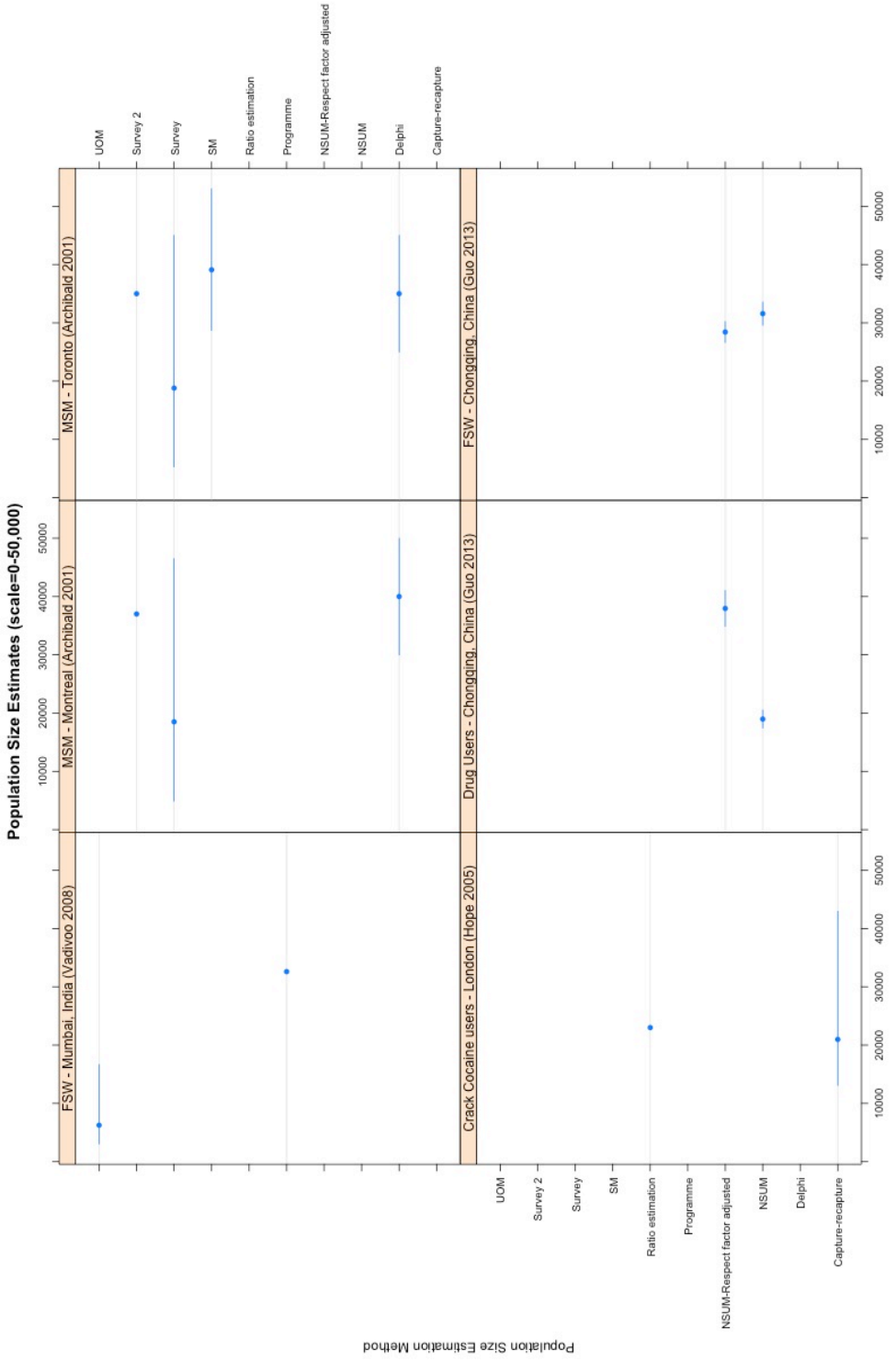

Population Size Estimates (scale=0-5,000; Vadivoo 2008 only)
PWID - Churachandpur (Vadivoo 2008)
PWID - Phek (Vadivoo 2008)
PWID - Wokha (Vadivoo 2008)
MSM - East Godavari (Vadivoo 2008)
MSM - Madurai (Vadivoo 2008)
MSM - Pune (Vadivoo 2008)
MSM - Salem (Vadivoo 2008)
MSM - Visakhapanam (Vadivoo 2008)
FSW - Visakhapatnam (Vadivoo 2008)
FSW - Warangal (Vadivoo 2008)
FSW - Yevatmal (Vadivoo 2008)
FSW (brothel) - Pune (Vadivoo 2008)
MSM - Coimbatore (Vadivoo 2008)
FSW - Dimapur (Vadivoo 2008)
FSW - Kolhapur (Vadivoo 2008)
FSW - Parbhani (Vadivoo 2008)
FSW - Prakasam (Vadivoo 2008)
FSW - Pune (Vadivoo 2008)
UOM
SM
RTM
Programme
0
2000
4000
6000
Population Size Estimation Method

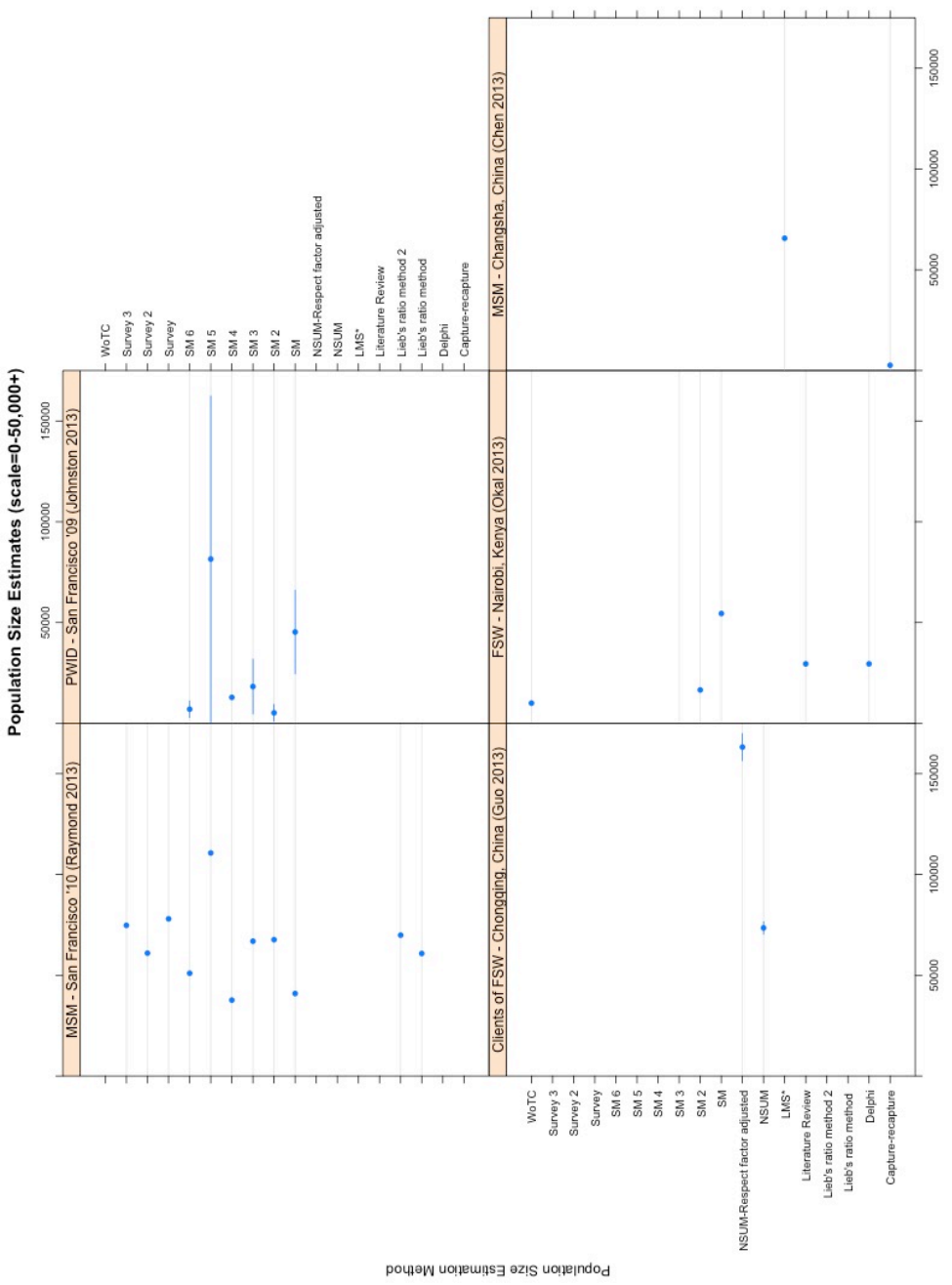

Population Size Estimates (scale=0-10,000)
Opiate users - Rotterdam, The Netherlands (Cruyff 2008)
Young Black MSM - South Side of Chicago (Livak 2013)
FSW - Ungula Island, Zanzibar (Khalid 2014)
MSM - Chennai, India (Vadivoo 2008)
MSM - Hyderabad, India (Vadivoo 2008)
MSM - Ungula Island, Zanzibar (Khalid 2014)
FSW - Guntur, India (Vadivoo 2008)
FSW - Hyderbad, India (Vadivoo 2008)
FSW - Karimnagar, India (Vadivoo 2008)
FSW - Salem, India (Vadivoo 2008)
FSW - Chitoor, India (Vadivoo 2008)
FSW - Coimbatore, India (Vadivoo 2008)
FSW - Dharmapuri, India (Vadivoo 2008)
FSW - East Godavari, India (Vadivoo 2008)
2000
4000
6000
8000
10000
WoTC
UOM
Triangulated Estimate
Survey
SM
RTM
Programme
Panel of Experts
Literature Review
Delphi
CRc-Poisson, no covariates
CRc-Poisson with covariates
CRc-NB, no covariates
CRc-NB with covariates
Capture-recapture
Population Size Estimation Method

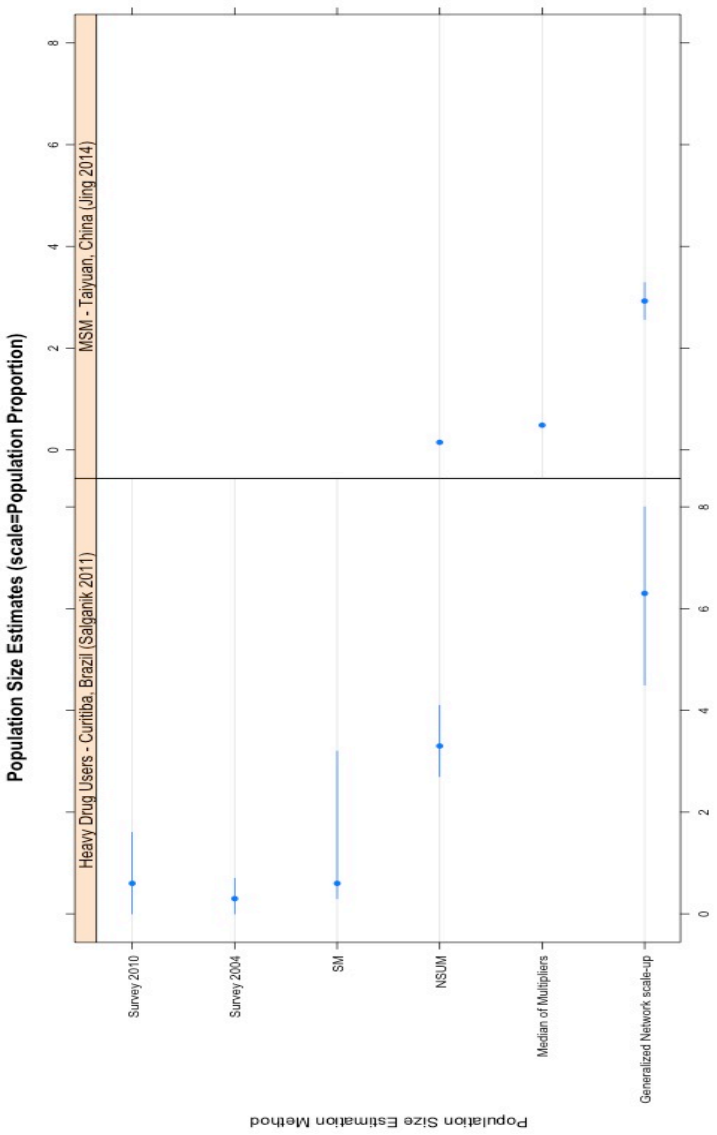

Population Size Estimates (scale=0-20,000)
PWID - Chongqing, China (Guo 2013)
PWID - Montreal (Archibald 2001)
PWID - Nairobi, Kenya (Okal 2013)
MSM - Koforidua, Ghana (Quaye 2015)
MSM - Nairobi, Kenya (Okal 2013)
PDU - Netherlands (van der Heijden 2013)
Illicit Drug Users - Bangkok, Thailand (Kuhnert 2009)
MSM - Cape Coast, Ghana (Quaye 2015)
MSM - Chongqing, China (Guo 2013)
WoTC
UOM
Survey
SM 4
SM 3
SM 2
SM
NSUM-Respect factor adjusted
NSUM
M&E
Literature Review
Delphi
CRc-Zelterman 2
CRc-Zelterman
CRc-TP
CRc-Chao
CRc-CAMCR
Capture-recapture
0
5000
10000
15000
20000
Population Size Estimation Method


Population Size Estimates (scale=0-20,000)
MSM - Mumbai Thane (Vadivoo 2008)
FSW (street) - Mumbai (Vadivoo 2008)
MSM - Guntur (Vadivoo 2008)
FSW (street) - Thane (Vadivoo 2008)
FSW - Chennai (Vadivoo 2008)
FSW - Madurai (Vadivoo 2008)
FSW (brothel) - Thane (Vadivoo 2008)
UOM
SM
RTM
Programme
0
5000
10000
15000
20000
Population Size Estimation Method


Population Size Estimates (scale=0-30,000)
PWID - Toronto (Archibald 2001)
PWID - Vancouver (Archibald 2001)
MSM - Vancouver (Archibald 2001)
PWID - Mauritius (Johnston 2011)
PWID - San Francisco, CA (Johnston 2005)
FSW (brothel) -Mumbai, India (Vadivoo 2008)
MSM - Accra, Ghana (Quaye 2015)
MSM - Kumasi, Ghana (Quaye 2015)
WoTC
UOM
Survey 2
Survey
SM 5
SM 4
SM 3
SM 2
SM
RTM
Programme
M&E
Literature Review
Delphi
Capture-recapture
0
10000
20000
30000
Population Size Estimation Method


Population Size Estimates (scale=0-50,000)
FSW - Mumbai, India (Vadivoo 2008)
MSM - Montreal (Archibald 2001)
MSM - Toronto (Archibald 2001)
Crack Cocaine users - London (Hope 2005)
Drug Users - Chongqing, China (Guo 2013)
FSW - Chongqing, China (Guo 2013)
10000
20000
30000
40000
50000
UOM
Survey 2
Survey
SM
Ratio estimation
Programme
NSUM-Respect factor adjusted
NSUM
Delphi
Capture-recapture
Population Size Estimation Method


Population Size Estimates (scale=0-50,000+)
MSM - San Francisco '10 (Raymond 2013)
PWID - San Francisco '09 (Johnston 2013)
Clients of FSW - Chongqing, China (Guo 2013)
FSW - Nairobi, Kenya (Okal 2013)
MSM - Changsha, China (Chen 2013)
WoTC
Survey 3
Survey 2
Survey
SM 6
SM 5
SM 4
SM 3
SM 2
SM
NSUM-Respect factor adjusted
NSUM
LMS*
Literature Review
Lieb's ratio method 2
Lieb's ratio method
Delphi
Capture-recapture
50000
100000
150000
Population Size Estimation Method



To account for the possibility of one method being poorly implemented when estimating the size of a target population, we chose to look at a subset of studies with at least three PSE methods and examine agreement (i.e., overlap of point estimates within each others confidence intervals) between all but one PSE method implemented for a given population. For 14 out of the 62 target populations that included at least three PSE methods (range 3 to 11 different PSE methods), point population size estimates were consistent across all but one PSE method. Of the

estimates of the sizes of these populations, over half (n=8) included a method that yielded estimates of low precision, producing very wide intervals that overlapped other point estimates. Three-fourths (75%) of the estimated target population sizes included at least one estimate that did not present associated upper or lower bounds.

Comparison of PSE methods across studies indicated little agreement in size estimates between methods, although comparisons did provide insight into the relative performance of specific methods. Estimates from organizations that provide services to the target population tended to be among the higher estimates for a given target population. While this finding is based on 37 of the populations estimated, they all originate from a single study (Vadivoo et al.). The wisdom of the crowds approach and the Delphi approach similarly base their estimation of the population size on the opinion of stakeholders or experts, often coupled with examination of data collected locally or from the published literature. The wisdom of the crowds method, which asks survey respondents how large they believe the target population is and then summarizes all responses into a median or mean, consistently produced the lowest or second lowest estimate of the size of the target population, although the range of responses were extremely wide. As described above, these were not statistically calculated confidence intervals, but rather the range of estimates provided by individuals sampled. The Delphi approach, which solicits population size estimates from a small group of stakeholders, often after showing this group the results of other PSE methods, tended to produce estimates near the medians of the other population size estimates calculated.

Two studies included a nationally representative survey, applying the proportion of self-identified MSM and PWID in the survey to the size of the general population to calculate the sizes of these target populations. In one study (Archibald et al.), six populations were estimated using a representative population survey, among other methods. For all six populations, while the rank order of the point estimates from the population survey varied relative to the other methods, this method consistently produced estimates with the widest confidence intervals. The authors concluded that the small sample size for the general population-based surveys and the low frequency of the MSM and PWID subgroups within this general population contributed to the relatively large width of the associated confidence intervals. In the second study (Livak et al.), the representative population survey was the only method to produce confidence intervals around the point estimate, so agreement between methods, as defined by this review, could not be assessed. While the confidence interval for the estimate from the representative population survey did not span any of the point estimates from the other methods, the difference between the estimate from the population survey (the lowest estimate) and the Delphi approach (the highest estimate) was less than 6,000, a relatively small difference compared to the difference

between other size estimates included in this review of similar magnitude.

PSE methods that rely on enumerating the visible members of the target population (i.e., mapping and enumeration) consistently produced the lowest estimates of the size of the target populations. As an exception to this trend, the enumeration method included in the study by Mutagoma et al. produced the highest estimate of the size of the population of FSW in Rwanda; however, the authors noted limitations in the other two PSE methods (capture-recapture and multiplier method) that may have resulted in underestimates being produced by these two methods. For capture-recapture, the venues sampled on the first capture occasion were largely the same as the venues sampled on the second capture occasion. This positive dependence between capture occasions will result in an underestimation of the target population. For the multiplier method, the researchers noted that the time period specified for the benchmark and the multiplier, while overlapping, was not exactly the same. This inconsistency could result in an underestimation of the target population if the proportion that is used in the denominator of the calculation is inflated because there is a longer eligible time period to recall visiting a service or program. The reverse tracking method featured in only one study (Vadivoo et al.), but was used to estimate 31 populations (FSW, MSM, and PWID in multiple Indian cities and towns). The reverse tracking method augments mapping and enumeration data, with additional enumeration of the target population at a venue on a second visit, to statistically estimate the size of the target population at each venue. Population size estimates from the reverse tracking method tended to fall in the mid to lower range of estimates calculated by various methods.

The literature review method, which applies estimated proportions of the target population in the general population from published studies of comparable populations in similar settings to census figures in the country of interest to estimate the size of the target population, yielded mixed results. Estimates produced using this method consistently fell within the range of estimates produced by other methods. When reported, the upper and lower bounds for the literature review method were wide.

The capture-recapture method generally produced estimates consistent with other methods, as well as reasonably narrow 95% confidence intervals. The multiplier method produced estimates with more variability. No clear trend emerged when comparing the relative performance of this method to other methods included in the review. These methods calculate size estimates from data on people using a program (service multiplier) or people receiving a specific object distributed to the target population (unique object multiplier). Several multipliers from different benchmarks were often used to estimate the same target population size. When several multipliers were implemented within the same study, the size estimates calculated from these

methods were often not in agreement with each other. The size of the population estimate derived using the multiplier method is determined by both the benchmark (the count of unique objects distributed or the number of people seen by a service, i.e., the numerator) and the multiplier (the proportion of respondents from the representative survey who received the object or participated in the service, i.e., the denominator). As a result, the validity of the method depends on the ability of selected services to maintain accurate records concerning unique clients of the target population seen by the service during a specified timeframe, or that the number of unique objects reported to be distributed is actually distributed. For example, if researchers believed 500 unique objects were distributed (the benchmark), but in reality only 300 unique objects were distributed, the calculated population size would be an overestimate, assuming the multiplier (the denominator) was accurate. Estimates from multiplier methods also showed variability with respect to the width of their 95% confidence intervals. The upper and lower bounds of the 95% confidence intervals were determined by the standard errors for the multiplier; that is, the proportion of respondents in the representative survey indicating they received the object or had participated in the specified program. These standard errors are calculated to account for the complex sampling design used to sample the target population, such as respondent-driven sampling (RDS), and allow inference to the target population based on the population sampled.

2.4 Discussion

Our review found no evidence of a single most rigorous population size estimation method among 21 separate methods held to side-by-side comparisons. In the absence of a gold standard against which to hold all methods, this judgment is based on consideration of multiple potential sources of biases that can result in severe over- or under- estimation, theoretical assumptions that cannot be shown to be met, and the high variability that each method engendered across multiple settings and populations. Nonetheless, a few notable patterns emerged when examining the relative rankings in size of estimates from studies that compared two or more methods simultaneously in the same population. The wisdom of the crowds method never produced the largest size estimate in publications that compared it to two or more other methods. This may result from the populations perceiving and articulating their sense of marginalization from the mainstream of society, a reason for their interest to public health researchers in the first place. Population-based surveys also were not the source of the highest estimate, with only one exception. This finding may also result from marginalization and stigmatization in that members of the population may not

acknowledge their status in home-based surveys. The reverse tracking method never produced the highest estimate when compared to others, a finding that may be due to the challenge in counting and distinguishing all persons who may be present at a particular hotspot or only counting those subsets of the population who are most visible. Potentially for similar reasons, direct enumeration did not produce the highest estimate when compared to other methods, with only one exception. Enumeration also did not produce the lowest estimate when compared to two or more other estimates. Using the published literature to estimate population size tended to fall in the middle when multiple methods were reported, never being the lowest and only once being the highest. Two potential biases may operate to produce this finding: a publication bias may prevent very (unacceptably) high or low estimates from being reported in the literature or the researchers pick a moderate estimate from those published in the literature to apply to their setting. The latter tendency may also explain why the Delphi method tended to produce a central estimate, particularly when conducted with an iterative process of presenting results from local studies and from the literature and allowing for revision of initial guesses. Similar to findings with the Delphi method, a substantial proportion of results from stakeholders who provide services and outreach to members of the target population tended towards a central estimate relative to the range produced by other methods. The majority of size estimates from stakeholders, however, were the highest among other methods reported. Rarely, did stakeholder estimates generate the lowest estimate of a target population. This pattern may be reflective of a desire to have a higher but credible official count of the target population size, which could be used to advocate for increased program funding.

For the majority of target populations included in this review, the different PSE methods produced estimates of the population size that were not in agreement with each other and had no overall pattern of being consistently the highest or lowest of multiple estimates. In the majority of side-by-side comparisons, 95% confidence intervals (or other defined upper and lower plausible bounds) around estimates produced by different PSE methods did not overlap, indicating that differences in the estimates were not occurring by chance but rather by systematic errors or biases in one or both methods. There are many possible sources of error. The multiplier methods, for example, have numerous reasons for over- or under- estimating the benchmark (e.g., duplicate client counts, incomplete recording of key population membership) and the multiplier (e.g., social desirability response bias, poor recall). Other differences may be inherent in who is reached by the different approaches. For example, if some PSE methods preferentially target certain members of the population (i.e., mapping-based methods are more likely to include the more visible) then these non-overlapping confidence intervals may be estimating the sizes of different

subsets of the target population. The methods to calculate confidence intervals may also be incorrect or not based on probability. For example, population-based surveys and the capture-recapture and reverse tracking methods provide calculations of standard errors that produce true confidence intervals. Upper and lower bounds for the wisdom of the crowds and Delphi methods, on the other hand, were often given as the range of guesses provided by the participants, with the minimum and maximum guesses providing the bounds. Bounds for both the service and unique object multipliers were determined using the standard error for the proportion from the representative survey who responded that they had received the service or object (the multiplier), accounting for the complex sampling design (for RDS studies or time-location sampling studies, for example). Recent studies of RDS estimators have questioned their validity in calculating appropriate weights and standard errors for population inference, especially if there is finite population bias.[12, 62] These studies add to the concern that the 2.5 and 97.5 percentiles calculated for the multipliers may be inaccurate, therefore resulting in inaccurate upper and lower bounds for the multipliers used to estimate population size.

Two PSE methods, capture-recapture and the network scale-up, offer alternate estimation approaches that may correct statistically for violations in their underlying assumptions. The most common assumption violations for the capture-recapture method are the source independence and the equal probability of capture assumptions; that is, an individuals presence in one source does not affect his/her probability of also being present in another source used in the analysis. To correct statistically for this assumption violation, investigators may use at least three sources and include interaction terms between a combination of sources in the regression modeling to indicate and control for source dependency. Investigators may also include individual-level covariates in the regression model or stratify by these covariates if they believe that the unequal probability of capture is due to individual-level characteristics. For example, Cruyff et al. estimated the number of opiate users in Rotterdam, the Netherlands using capture-recapture models that included gender, marital status, nationality, income, and age as covariates, among others. In this case, including these covariates in the regression modeling did not significantly affect the point estimate of the size of the population. However, assumptions about the distribution of the statistical model did noticeably affect the point estimate. In contrast to a Poisson distribution, using a negative binomial distribution, which includes a dispersion parameter to account for unobserved heterogeneity in capture probabilities, resulted in better model fit according to the Akaikes Information Criterion and log likelihood criteria. Alternative modeling approaches, such as mixture models or the Zelterman estimator, attempt to control for latent heterogenity in the capture probabilities. van der Heijden et al. estimated the population of problem drug

users in the Netherlands at risk for clinical hospital treatment, using the truncated Poisson regression (TPR) model and the Zelterman regression model, alternative estimators for a capture-recapture analysis using a single list. Statistical analysis of the TPR model indicated poor model fit. This suggests that despite adjustment for observed covariates, heterogeneity in the capture probabilities remained; that is, members of the target population had an unequal probability of being seen multiple times in a single source (violating a key assumption in the capture-recapture study design). In this example, the Zelterman regression, which is robust to unobserved heterogeneity, was determined to be the better model for estimating the unobserved population. Smit et al. noted significant differences in estimates produced by the Zelterman model and by other truncated Poisson models, with the former producing higher estimates; this result may indicate unobserved heterogeneity, in which case the Zelterman model should be preferred, as it is likely to be more accurate.

The network scale-up method attempts to addresses biases in population size estimation resulting from stigma. The size of the target population from the network scale-up method originates from a general population-based survey, not by asking respondents questions about their own behaviors, but rather asking about persons within their social networks. The method assumes that, on average, the people comprising an individuals network will be representative of the general population. That is, if 10% of the people in an individuals personal network are teachers, then teachers make up 10% of the general population. The method assumes that respondents are aware of their social contacts being members of the target population (i.e., no transmission error) and that members of the target population, on average, have the same personal network size as the general population. If these assumptions are violated, the generalized network scale-up can adjust for biases. This method requires a separate, representative survey of the target population in order to calculate two additional parameters; $\hat{\tau}$, the estimated information transmission rate to account for the transmission error, and $\hat{\delta}$, the estimated popularity rate to account for any difference in network size between the general population and the target population.[50]

The capture-recapture and generalized network scale-up methods are appealing because they offer the investigator a sense of control through the correction of biases after implementation of the study. However, in the absence of a gold standard, we do not know if such statistical adjustments are correcting bias or introducing bias into the estimation. When there are multiple statistical models to choose from when conducting a capture-recapture analysis with at least three sources, convention dictates selecting the model with the smallest value for the Akaikes information criterion (AIC).[26] However, a recent study by Jones et al. illustrates how this criterion for model selection can result in selecting the incorrect model to estimate the unobserved population size (Jones et al., 2014). Through simulation studies, the

authors showed how model selection based solely on statistical relationships within the data can lead to biased estimates. In this example, referrals of unique observations between capture sources induced interactions between sources. The statistical model that accurately described this relationship between sources was not the best model, according to the AIC. In fact, the model that would have been selected by the AIC convention underestimated the true population size (though the authors demonstrated that the bias from poor model selection can also result in an overestimation of the population size). Therefore, reliance on statistical models can result in unpredictable biases if knowledge of the real-world processes that generated the data are not taken into account.

The frequency with which the multiplier method was used is indicative of its appeal among public health researchers, likely due to the ease with which it can be incorporated into studies of hidden populations. Population size estimates can easily be calculated, given a count of clients from a service provider, together with a single question in a population-based survey about visiting that service provider. However, as noted in this review, the resulting estimates can be quite variable. Estimates from the multiplier method will be biased if either the benchmark is inaccurate (i.e., a different number of unique objects are distributed than are reported to have been distributed, service provider reports client visits as opposed to unique clients, or members of the target population are not distinguished from non-members), or the multiplier is inaccurate (respondents from the population survey may not recall receiving the unique object or visiting a particular service within the specified timeframe). Given the sensitivity of the multiplier method to violations of the assumptions, and the observed variability in estimates across studies observed in this review, researchers should be cautious when using a single multiplier to estimate the population size. Agreement in the estimated population size by multiple methods reduces concern when different methods produce similar estimates of the same population. However, this approach risks confusing the reliability of the estimates with the validity of the estimates. Estimates can be highly consistent with each other, and therefore reliable (i.e., produce similar results), and yet all of them may systematically under- or over- estimate the true population size, thus producing invalid estimates.

This review has several limitations. Although we used a combination of search strategies to find published articles that met our eligibility criteria, it is possible that our search strategy missed some articles that would have been eligible. Due to the immense volume of results from one PubMed query (PubMed2), we implemented a stopping rule, which ended the search after the reviewer screened an interval of 100 titles that were determined to be irrelevant to the review. At this point, the reviewer continued to screen an additional 100 titles to see if any eligible titles would have

been missed by using this stopping rule, and found none to be relevant. This finding, together with the duplicates identified using three separate search strategies, adds to our confidence in the comprehensiveness of this review. Still, while our initial search strategy erred on the side of sensitivity to increase the probability of finding all relevant articles, we recommend that future studies of population size estimation (or ones that include a PSE component) adhere to a standardized way of noting in the title or abstract that the study includes a population size estimation component.

We purposely restricted our review to peer-reviewed studies in order to consider only studies of sufficiently high quality to have passed review by public health researchers and to be readily accessible to other researchers to replicate our findings. As a result, we may have missed comparisons of PSE methods covered in the grey literature, such as technical reports. This exclusion may not only have led us to miss some comparisons, but also may have caused us to miss potentially novel methods not found in peer-reviewed articles. Furthermore, our literature search was conducted from February 6, 2015 to March 6, 2015; studies eligible for this review may have been published in the peer-reviewed literature after this time frame. For example, as of writing we are aware of at least two studies only recently published using the successive-sampling population size estimation (SS-PSE), a newly described method for the health literature that uses data on network size generated from RDS studies to estimate the population size.[29, 64] During the literature search we also came across novel PSE methods not identified in previous summaries, such as the multiple indicator method (which uses generalized linear regression modeling to extrapolate size estimates from geographic areas with data to geographic areas where data are absent) [19] and the Bayesian hierarchical model method, which uses Bayesian modeling to weight estimates from multiple PSE methods according to their susceptibility to bias and arrive at a single estimate.[3] However, these studies were not eligible for inclusion in this review due to either not including an additional PSE method for comparison or not being published in the peer reviewed literature.

Only one author [PW] conducted the literature search and assessed the eligible studies for review. Consequently, we were unable to calculate a reliability score to demonstrate the consistency with which the same studies would have been selected, had a second reviewer independently implemented the same search strategy. Due to the sensitivity in our search strategy, and the frequency of duplicates across different search strategies, we are confident that we identified and included most studies that were eligible according to the criteria for this review.

Currently, there is no evidence for a single best method to estimate the size of a hidden population. The evidence we have synthesized in this review highlights the variability in estimation of the same target population when using multiple PSE methods. Given the variability in population size estimates among the studies in-

cluded in this review, it is striking that only a fraction (7.3%) of PSE studies originally identified by title were eligible for inclusion in this review; the majority were ineligible because they did not include a second PSE method to compare size estimates. From this systematic review of the peer reviewed literature, it is clear that a single size estimate provided by any one PSE method is not likely to be sufficient for producing a reliable estimate of the target population size. At a minimum, multiple different PSE methods should be implemented within the same study, so as to communicate the degree of agreement (or disagreement) between different methods for estimating the size of the same population, thereby providing a more transparent indication of the certainty of the final size estimate. Moreover, multiple estimates based on different assumptions should reduce the risk of selecting a size estimate based on a single severely biased method. Given knowledge of the strengths and limitations of certain methods (e.g., mapping-based methods tend to produce estimates on the lower range because they are enumerating the more visible members of the target population), careful thought should be given to which strata of the target population are being enumerated and, given the limitations of the PSE methods, if any stratum remain essentially invisible to public health researchers and surveillance systems. Furthermore, using multiple estimation techniques, such as capture-recapture with covariates, may be revealing of underlying population characteristics, such as heterogeneity in capture probabilities due to measured or unmeasured characteristics (i.e., age, race, gender, etc.). We systematically reviewed the literature on population size estimation methods to assess their consistency, and provided evidence for trends in how generalizable methods perform when implemented in the field. Simulation studies, or a focused study of multiple PSE methods on a known population, should be implemented to evaluate the validity of each method and the role of bias in accurately enumerating the target population. Finally, while a gold standard census for many of the populations of interest to HIV research and public health may be distant or unobtainable, we advocate for rapid peer-reviewed publication of size estimates with methods clearly described to obtain a greater global or meta sense of the variability of methods and central tendencies of key populations representation in diverse societies.

Chapter 3

If you are not counted, you dont count: estimating the number of African-American men who have sex with men in San Francisco using a novel Bayesian approach

3.1 Introduction

[1]

Despite advances in treatment regimens and prevention strategies, HIV/AIDS remains a leading cause of morbidity and mortality worldwide.[42] Globally, key populations, such as men who have sex with men (MSM), female sex workers (FSW), and injection drug users (IDU), remain at increased risk for HIV infection. Due to biological, behavioral, and structural vulnerabilities, the prevalence of HIV infection is typically higher in these groups than in the general population. Targeting key populations for public health outreach is one of six strategies on the global agenda to achieve maximum effectiveness in the public health response to HIV.[45, 52] Investing in programs that focus on key populations is a component of the strategic global response to the HIV epidemic, and requires reliable estimates of the sizes of these

[1]This chapter has been published in the Journal of Urban Health: Wesson, P., Handcock, M.S., McFarland, W., & Raymond, H.F. (2015). If You Are Not Counted, You Don't Count: Estimating the Number of African-American Men Who Have Sex with Men in San Francisco Using a Novel Bayesian Approach. Journal of Urban Health. doi:10.1007/s11524-015-9981-0

populations, so that resources may be allocated efficiently and public health actions prioritized. Furthermore, enumeration of key populations allows epidemiologists to quantify the burden of disease and model the impact of targeted interventions. Finally, enumeration of these key populations contributes to the evaluation of programs with respect to reach, coverage, and intensity.

Even among key populations, disparities in disease burden exist. Studies have reported on the disparities in HIV and other STIs among African-American (AA) MSM.[22, 58] AA MSM are a key population for HIV infection in the San Francisco Bay Area. In 2010, AA men were reported to have the highest incidence of HIV infection for any racial group,[51] although recent data suggest a possible convergence between groups.[55] Among AA men living with AIDS in 2010, the majority (52%) were MSM. A previous study by Scott et al. also reported an increased burden of HIV and other STIs among AA MSM in the San Francisco Bay Area.[53]

Despite the observed relative disparities in disease burden among AA MSM, the true scope and scale of the epidemic in this population have not been quantified because the size of this population remains unknown. Although several methods are available, quantifying the size of many key populations remains a challenge in public health. Current population size estimation (PSE) methods widely used in public health require data that is very difficult to obtain or require assumptions that are difficult to meet or verify. For example, capture-recapture, which requires multiple data sources that each list members of a target population, traditionally assumes that these data sources are independent of each other and that each member of the target population has an equal probability of appearing on each list included in the analysis,[26, 27] assumptions that seem unlikely to be true. These assumptions may be relaxed, using log-linear models to specify the relationships between the data sources and allow the probabilities of appearance to vary. However, these modeling assumptions are subject to misspecification, resulting in biased estimates.[31] Similar to capture-recapture, the service multiplier method requires two sources of data; one source is a direct count of the target population participating in a service, while the other source is a representative sample of the target population. The multiplier method assumes that the two data sources are independent, and that one of the data sources is a representative sample of the population, an assumption that is difficult to verify for hidden populations.[30] Other PSE methods, such as network scale-up, require large population-based surveys and the addition of many questions that may not always be feasible.[50] PSE methods usually require planning in advance of study implementation in order to be carried out successfully.

A new PSE method has great appeal as it can be implemented using data routinely collected within respondent-driven sampling (RDS) surveys. We applied the new PSE method, referred to as the Sequential Sampling population size estimation

method (SS-PSE), to previously collected data from an RDS survey of AA MSM in San Francisco [10] to estimate the size of the citys AA MSM population. This method, which uses the network size question asked in RDS studies, has been tested in simulation studies,[17] but has less often been described empirically in key populations. Our aims, therefore, were to apply the SS-PSE method using data from the RDS study to estimate the number of AA MSM living in San Francisco and compare results to other estimates.

3.2 Methods

The Black Men Testing (BMT) data originate from a cross-sectional Integrated Bio-Behavioral Surveillance (IBBS) survey of AA MSM in San Francisco, California. The original study was implemented in 2009 by the San Francisco Department of Public Healths (SFDPH) HIV Prevention section for the purpose of using social networks as a channel to reach AA MSM for HIV testing.

Study participants were recruited through RDS, a peer-recruitment method commonly used worldwide to sample hard-to-reach populations.[20, 21] Sampling begins with members of the target population, referred to as seeds, purposefully selected by the research team. Each seed is given a pre-determined number of coupons to give to other target population members who are in their social network. The coupons themselves have no external monetary value; they are simply tokens that allow an individual to enroll in the study. The coupon allows potential participants to enter the study and tracks the waves and patterns of recruitment through a unique code that links the recruiter and recruit. Each study participant thereafter is given coupons to distribute within their social network, and this process of recruitment iterates until both sample size and sample stability (where the composition of the sample changes little with subsequent recruitment) are reached. In theory, with enough waves of recruitment, the final RDS study sample will be independent of the characteristics of the initial RDS seeds, and enough information is collected to adjust statistically for differential probability of being selected. A statistical assessment of RDS is given by Gile and Handcock.[12]

RDS seeds for the BMT study were selected to represent the diversity of AA MSM in San Francisco, according to age, neighborhood of residence, and education level. Each respondent was given three coupons to use to recruit other AA MSM at least 18 years of age and in his social network. The final sample size included 256 AA MSM. Details of the BMT study and main findings have been described elsewhere.[10]

The SS-PSE method models the total number of persons in the target population

using RDS data. The SS-PSE method is an adaptation of a similar model described and implemented by Nair and Wang [43] and by West [65] to estimate the size of untapped oil pools in an oil reserve based on the observed measures of size for already discovered oil pools in the reserve (volume, surface area, net pay, and depth). The model assumes a size-bias sampling in which larger oil pools are more likely to be discovered before smaller oil pools. A prior estimate of the total number of oil pools is included in the model, along with information on the measured parameters of the already discovered pools, to model the characteristics of the remaining oil pools (e.g., volume) in the reserve, should they exist. These parameters are modeled as a posterior distribution, expressing the probability of characteristics, such as volume of oil remaining in the reserve yet to be discovered.

In the human population application using RDS data, the SS-PSE method uses self-reported individual network size (i.e. the number of other members of the target population an individual respondent knows) as the informative measure of the target population. Just as the physical characteristics of the oil pools determine the probability that an individual oil pool will be observed, the SS-PSE assumes that the size of individuals social network with respect to the target population influences the probability that an individual will be observed during the RDS discovery process. The SS-PSE method assumes that respondents with larger network sizes, those more socially connected, are more likely to be discovered initially by RDS recruitment than respondents with smaller network sizes. Formally, the model assumes that ones probability of selection is proportional to that individuals network size. Over the period of recruitment, with sequential sampling without replacement, the probability of being sampled over time is proportional to the network size of the remaining members of the population. The model further assumes that the target population is uniform; when respondents report their network size, this number is in reference to the target population as a whole and is not restricted to specific subgroups within the target population. As an extension of this second assumption, the model implicitly assumes that respondents interpret the network size question in the same way.

The SS-PSE method uses a Bayesian approach to estimate the probable size of the target population. A prior estimate of the population size is used to represent previous knowledge about the target population and, if necessary, provide bounds on the population size estimate. The prior estimate, expressed as a measure of central tendency, is combined with the specified shape of the distribution to calculate the prior distribution of the population size. If very little is known about the prior size and distribution, a uniform distribution may be specified. For our informative prior, we used 4,450, based on a previous estimate by Scott et al.[53]

The SS-PSE method uses the prior estimate in combination with the specified distribution and the data (the self-reported network size) to calculate the posterior

population size estimate. Markov Chain Monte Carlo (MCMC) simulations are used to compute the posterior distribution. MCMC simulations use a directed random-walk algorithm to sample possible values of the parameter of interest.[14] While this process of sampling from the parameter space is random, some values will have a higher probability of being drawn than others, because the Markov chain is sampling from the more likely regions of the parameter space. The differential probability of sampling from the parameter space is determined by the information in the data (in this case, the network size) and the prior estimate for the population size. The entire distribution of the parameter of interest is then constructed from this (directed) random sampling. Consistently estimating the posterior distribution can be improved by increasing the MCMC settings, such as the number of samples taken from the parameter space. Additionally, the burn-in period may also be increased; the burn-in period refers to the number of samples initially taken to begin the Markov chain, but these samples do not contribute to the estimation of the posterior distribution. Any measure of central tendency can then be calculated to summarize the probability distribution of the population size. Full details of the SS-PSE method are described elsewhere.[17]

Three network size questions were included in the IBBS survey. We chose the most specific network size question (Of the [African American men who have sex with men, who live in San Francisco and are 18 years or older and you have seen in the past 30 days] how many do you think you could give a coupon to (like the one you brought in today) within the next four weeks?), as this was the most specific to an individuals probability of selection for the RDS study. Other questions were phrased more generally about the number of other AA MSM the respondent knows.

The SS-PSE also allows for the option of truncation. Truncation imposes bounds on the posterior probability distribution so that no probability is assigned to values outside defined bounds. The tail of the lower end of the probability distribution is always truncated at the sample size of the RDS sample because the estimated size of the target population cannot be less than the number of people sampled and included in the RDS data set. The user can specify upper truncation, although the default setting is no upper truncation for the posterior probability distribution. If the user has prior knowledge whereby it would be impossible for the population size to be above a certain value, the upper tail of the probability distribution may be truncated to avoid extending past a certain value (and therefore no probability is assigned to any value beyond this upper limit). The area under the curve of the region of the tail that would have extended past the upper truncation is then redistributed within the allowed bounds of the posterior probability distribution.

Analysis

Analyses were performed using STATA version 12,[54] R (version 3.1.1) [46] and RDS-Analyst (RDS-A) (version 1.7-16).[15] RDS-A allows for the selection of different RDS estimators to conduct population inference from the RDS sample. RDS-A includes the RDS estimator available in RDS Analysis Tool, as well as the Giles SS (Sequential Sampling), which accounts for finite population bias by using the reported individual network size and estimated population size to weight the sample, and does not assume sampling with replacement.[11] Recruitment trees were produced using RDS-A.

3.3 Results

The BMT data included 256 eligible AA MSM, recruited by ten seeds. Recruitment took place from February to September 2009. The reported network size ranged from 1 to 99. One network size reported as 0 was re-coded as 1 because we assumed that the respondent knew at least one other member of the target population (the person who recruited him, or the person he recruited). By the same logic, six respondents with a reported network size of 999 (Not Applicable) were also recoded to have a network size of 1. For this analysis, four non-seed participants were removed because they were not linked to any other participant in the dataset for unknown reasons. The final sample for this analysis included 252 respondents. Figure 1 shows the recruitment tree, with each node scaled to reflect reported network size. A slight decrease in network size over successive waves of recruitment is evident. Table 1 describes the demographic characteristics and key HIV-related variables in the study population with two RDS estimators to make population inference. Differences between the RDS-II adjusted estimates and the Giles sequential sampling (SS) adjusted estimates indicate that there is little finite population bias.

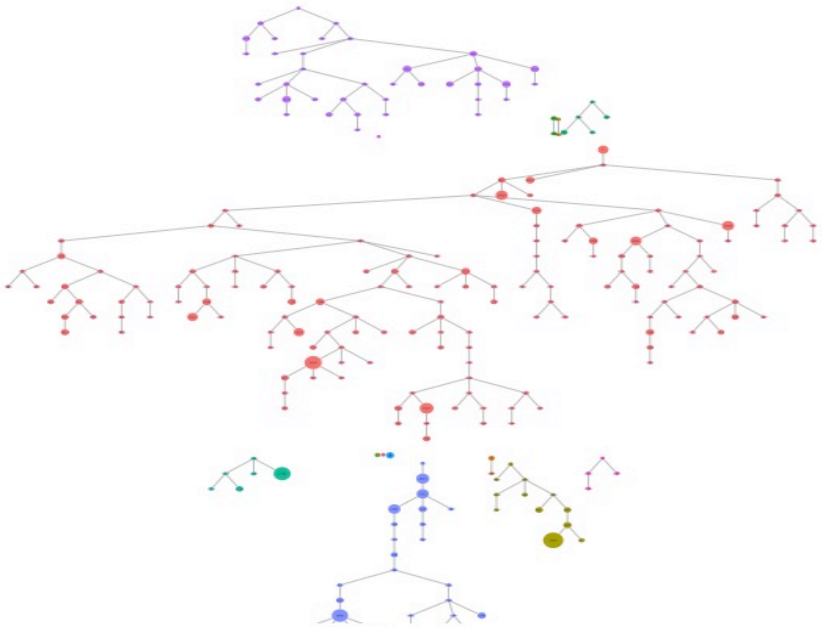


Figure 3.1: Recruitment tree (nodes scaled to network size) for African-American men who have sex with men (MSM) participating in the Black men testing (BMT) survey, San Francisco, 2009.

Characteristic	Crude county (%)	RDS-II weighted percent (95% CI)	Gile's SS weighted percent (95% CI)	Difference
Age group (years)				
18-20	2 (0.8)	3.1 (0.9, 5.3)	3.2 (-1.1, 7.4)	0.1
21-25	12 (4.8)	12.0 (7.6, 16.3)	12.5 (1.7, 23.8)	0.5
26-30	17 (6.8)	4.8 (-5.0, 14.5)	5.0 (1.5, 8.5)	0.2
31-35	17 (6.8)	4.6 (1.8, 7.3)	4.8 (1.6, 8.1)	0.2
36-40	25 (9.9)	11.8 (0.3, 23.2)	12.3 (2.7, 22.0)	0.5
41-45	58 (23.0)	17.1 (11.8, 22.4)	18.0 (11.2, 24.7)	0.9
46-50	58 (23.0)	19.9 (11.1, 28.6)	16.1 (8.5, 23.7)	-3.8
51+	63 (25.0)	26.9 (21.6, 32.1)	28.2 (17.2, 39.2)	1.3
Education				
<High school	37 (14.7)	18.9 (8.6, 29.2)	14.8 (7.8, 21.9)	-4.1
High school	98 (38.9)	47.9 (35.5, 60.4)	50.2 (38.2, 62.2)	2.3
>High school	117 (46.4)	33.2 (22.6, 43.8)	35.0 (24.4, 45.6)	1.8
Annual income				
0-10k	106 (42.1)	55.1 (42.2, 68.1)	52.6 (41.9, 63.2)	-2.5
11-20k	66 (26.2)	20.7 (11.2, 30.3)	22.0 (14.5, 29.4)	1.3
21-30k	36 (14.3)	9.1 (-1.5, 19.7)	9.7 (4.2, 15.1)	0.6
31k+	44 (17.5)	15.0 (6.8, 23.2)	15.8 (8.5, 23.2)	0.8
Ever injected drugs (yes)	92 (36.5)	32.8 (20.9, 44.8)	29.5 (21.2, 37.9)	-3.3
Injected drugs in last 6 months (yes)	40 (15.9)	11.2 (5.2, 17.2)	11.8 (6.0, 17.7)	0.6
Ever tested for HIV (yes)	235 (93.3)	93.3 (86.6, 99.9)	92.9 (88.3, 97.5)	-0.4
Diagnosed with HIV prior to survey (of 229 respondents)	68 (29.7)	21.7 (12.7, 30.7)	27.0 (15.8, 38.2)	5.3
Positive HIV test result during survey (of 245 tested)	79 (32.2)	25.9 (14.2, 37.6)	34.0 (22.2, 45.9)	8.1

Table 3.1: Demographic characteristics, injection drug use (IDU), and HIV status of African-American men who have sex with men (MSM) participating in the Black men testing (BMT) respondent-driven sampling (RDS) survey, San Francisco, 2009 (N=252)

Population size estimates using the SS-PSE method are shown in Figure 3.2 and detailed in Table 3.2. Combining the prior distribution, based on a prior median estimate of 4,450, and the network size distribution from the BMT dataset, the model calculated a posterior median estimate of 5,708 (95% CI: 1,381-25,799; Model 1). Increasing the burn-in period, interval, and sample size for the Markov Chain Monte Carlo simulation settings reduced this median estimate to 4,917 (95% CI: 1,267 28,771; Model 2), which was more consistently estimated over repeated simulations. The American Community Survey indicates that for 2010, the year following the BMT survey, 20,824, AA men 18 years or older were living in San Francisco.[4] Truncating the upper bound of the prior distribution for the population size to a conservative 15,000 (i.e., that MSM are far less than 72% of adult men) resulted in a median estimate of 4,518 (95% CI: 1,330-13,051; Model 3). Using a flat prior distribution, specifying no prior knowledge of the population size, the SS-PSE estimated the median posterior estimate of the AA MSM population living in San Francisco to be 1,875 (95% CI: 910-2,461; Model 4). Increasing the specified prior median from 4,450 to 10,000 and again truncating the prior distribution at 15,000 resulted in a posterior median estimate of 6,762 (95% CI: 1,994-13,863; Model 5).

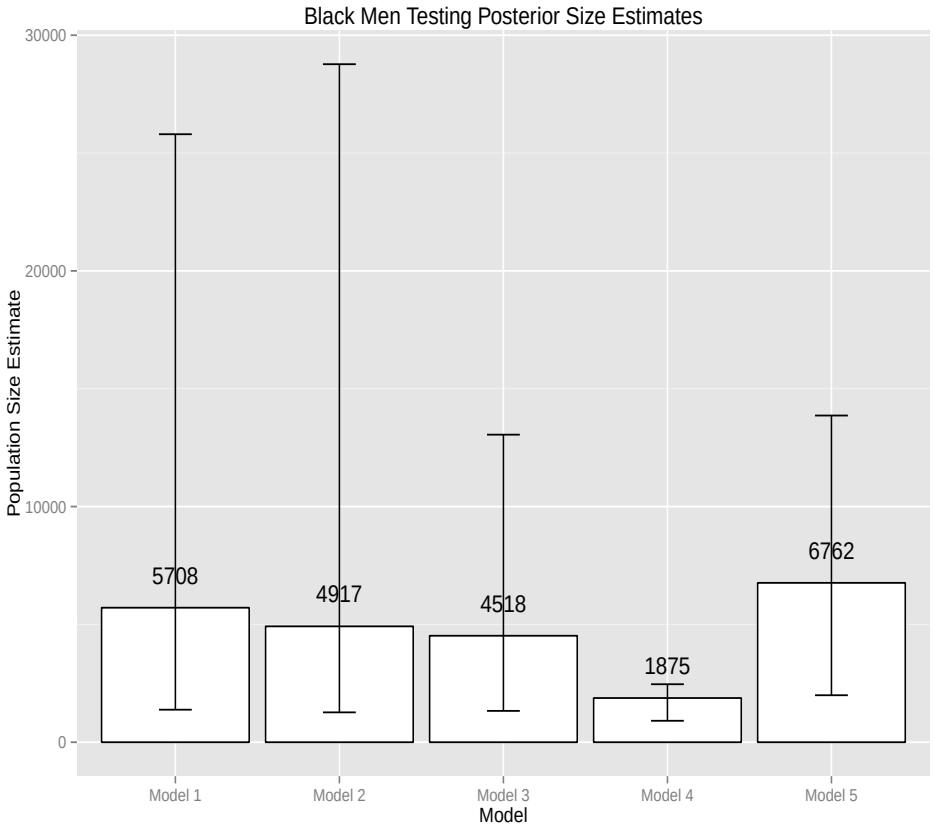


Figure 3.2: Bar plots comparing posterior population size estimates of the number of African-American men who have sex with men (MSM) by different prior inputs using the sequential sampling (SS-PSE) method, San Francisco, 2009.

Parameters	Model 1	Model 2	Model 3	Model 4	Model 5
Prior estimate	4,450	4,450	4,450	None	10,000
Prior distribution	Beta	Beta	Beta	Uniform	Beta
Burn-in period	5,000	10,000	10,000	10,000	10,000
Upper truncation	None	None	15,000	None	15,000
Lower truncation	252	252	252	252	252
Posterior Median	5,708	4,917	4,518	1,875	6,762
95% Credible Interval	1,381-25,799	1,267-28,771	1,330-13,051	910-2,461	1,994-13,863

Table 3.2: Posterior population size estimates of the number of African-American men who have sex with men (MSM) by different prior inputs using the sequential sampling size (SS-PSE) method, San Francisco, 2009.

For comparison, we examined other AA MSM population size estimates using different methods (Table 3.3). Previous size estimation exercises performed by the SFDPH estimated 66,487 total MSM living in San Francisco as of December 2010.(Raymond et al., 2013) In 2008, the National HIV Behavioral Surveillance (NHBS) survey done by time-location sampling estimated 6.5% of all MSM in San Francisco to be AA.[55] Applying the NHBS proportion to the estimated count from SFDPH yields an estimated 4,320 AA MSM. Two multiplier method adaptations were also possible. In 2009, 1,170 AA MSM diagnosed with HIV infection were reported to the SFDPH surveillance system by the time of the BMT survey. Meanwhile, 17.3% of the respondents in the BMT survey were HIV positive and aware of their status. Taking the BMT prevalence as the prevalence of diagnosed HIV cases, those that would be seen in the HIV surveillance system, we applied this proportion as the multiplier to the benchmark estimate from the HIV surveillance system, to yield an estimate of 6,763 (95% CI: 4,415 - 11,142) AA MSM in San Francisco.

Alternatively, the San Francisco HIV case reporting system indicated there were 1,186 HIV-positive AA MSM at the time of the 2008 NHBS survey of MSM. Using the 25% of HIV-positive cases among respondents to the BMT survey who previously did not know they were HIV positive, we adjust the number from the surveillance system to 1,581 total HIV cases among African-American MSM. NHBS estimates 25% of AA MSM to be HIV positive. Assuming 1,581 to be 25% of the total number of AA MSM, we project the population size to be 6,325.

Method	Source 1 (benchmark)	Source 2 (multiplier)	Population size estimate
SS-PSE	NA	NA	4917
Simple proportion	Estimated total San Francisco MSM population size [4] • 66,687	NHBS (2008) estimated proportion of MSM who are AA [55] • 6.5%	4320
Multiplier	AA MSM living with HIV in surveillance system • 1170	BMT proportion of African-Americans diagnosed HIV [30] • 17.3%	6763
Multiplier	Estimated number of AA MSM living with HIV (surveillance data accounting for unrecognized infection from BMT) • 1581	NHBS (2008) prevalence of HIV among AA MSM [55] • 25%	6325

Table 3.3: "External validation": other methods to estimate the number of African-American men who have sex with men (MSM) in San Francisco, 2009

3.4 Discussion

We estimated the size of the AA MSM population in San Francisco to be nearly 5,000 (4,917; 95% CI: 1,267 28,771). Taking into account the size of the total AA adult male population in San Francisco and truncating the prior distribution to maximum plausible upper value, refined this size estimate to 4,518 (95% CI: 1,330-13,051). This estimate is highly consistent with our prior estimate of 4,450, based on Scott et al.s projection, which used data from the 2004 National HIV Behavioral Surveillance MSM 1 study to estimate the number of AA MSM in San Francisco.[53]

We note several factors that affect the precision and consistency of this estimate, or are sensitive parameters using the SS-PSE method. The associated 95% probabil-

ity intervals for these estimates are quite wide. The median posterior estimate when using a flat prior distribution was nearly 2,000 (1,875; 95% CI: 910-2,461), which appears implausible as it is close to the number of AA MSM known to be living with HIV. Using 10,000 as a prior median (roughly twice the size of our informative prior) resulted in a median posterior estimate of 6,762 (95% CI: 1,994-13,863). This estimate is nearly identical to the estimated population size using the multiplier method of the HIV case reporting system and HIV prevalence in the BMT survey (6,763). Using a uniform-flat prior and a large (relatively uninformative) prior could provide an acceptable lower and upper bound to the estimated population size, respectively. This approach is especially helpful in settings that lack external data sources with counts of the target population.

Results from the BMT RDS survey suggest a 17.3% prevalence of HIV infection. Using the estimated population size from the SS-PSE method as the denominator, we estimate the prevalence of recognized HIV infection among AA MSM in 2009 to be 24%. This figure is consistent with the Giles SS estimate (27%; 95% CI: 16% - 38%) (Table 3.1). Extending the Giles SS estimator to diagnosed HIV infection within the BMT survey, again using our estimated population size of 4,917, we estimate the prevalence of HIV infection among AA MSM to be 34% (95% CI: 22% - 46%). This places AA MSM as an extremely vulnerable population for HIV infection, following MSM IDU (47.4% prevalence), transgender IDU (44.4% prevalence), and transgender women (35.5% prevalence).[51]

As with all size estimation approaches, the SS-PSE method depends on meeting underlying assumptions. These assumptions are challenging to verify. First, the model assumes that the probability of selection at any point is proportional to an individuals network size. That is, during recruitment, the probability of being sampled at a given point in time is proportional to their network size relative to the still unsampled members of the population at that point. Visualizing this size bias phenomenon through a recruitment tree with nodes scaled to reported network size does not show a clear decreasing trend in reported network size with subsequent waves of recruitment (Figure 3.1). The attempted crude visualization with these plots may not be sufficient to check the first assumption for the SS-PSE. The subtle signal may be observed only with a more sophisticated model that plots the likelihood of observing each participant at the moment he is observed, given the distribution of the remaining network sizes in the target population. A second assumption for the SS-PSE method is that the target population is uniform, such that the respondents reported network size is specific to the target population as a whole, and not to a specific subgroup. Unfortunately, there does not seem to be an empirical test for this assumption. A close approximation may be to explore homophily (i.e., similarity in characteristics between recruiter and recruit) in the data set, but this would

be limited to participants recruiting behavior, and not the composition of their social network with respect to the target population. While these assumptions seem reasonable and more likely to be met than the assumptions for other population size estimation methods (e.g., source independence in capture-recapture analysis), it is unclear if the BMT data set meets these assumptions and, if not, what would be the resulting direction of the bias.

Our model implicitly assumes that an individuals reported network size is an appropriate proxy for his probability of being recruited. In RDS studies, out-degree (the number of target population members a participant knows and can recruit) is used as a proxy for in-degree (the number of target population members who could have recruited the participant) to estimate the participants probability of selection into the study. RDS estimators use the network size measurement to weight participant observations in order to make inference to the characteristics of the larger target population from the study population. Investigators have previously noted difficulties in accurately measuring network size and have included different ways of asking the question in the same or separate surveys. Even if measured accurately, reported network size may not accurately reflect a participants probability of selection, due to other covariates that may influence recruitment behavior.[49] For example, an individual may report a network size of five, but only three of the people they had in mind would also consider the individual to be a part of their social network and would recruit him (reciprocity). In this case, the individual has over-estimated his true network size and therefore his probability of selection. If individuals believe they have a large network size, they may choose to round rather than indicate the exact size of their network. In the BMT survey, we observed digit preference behavior, whereby higher network sizes were reported in factors of five (e.g., 30, 35, 40, 45, 50, etc.). For other network size questions, the range extended past 100. While it is possible for someone to know 769 other AA MSM in San Francisco (the maximum reported network size for one of the network size questions asked in BMT), this is likely a generalization for having a large network size. We observed that reported network sizes greater than 100 may lead to convergence problems for the SS-PSE method.

According to RDS theory, with enough waves of recruitment, the characteristics of the final RDS sample will be independent of the characteristics of the seeds, and the sample will be representative of the target population. Unfortunately, without a gold standard for comparison, this assumption cannot be confirmed (especially when studying hidden populations). Previous research on MSM in Forteleza, Brazil has challenged the validity of this claim. Although the RDS sampling process succeeded in reaching otherwise inaccessible members of the MSM population in Brazil, the sample over-represented lower socioeconomic MSM compared to the other sampling

approaches.[33] It is possible that in our sample of AA MSM, the RDS sampling process only reached the lower socioeconomic portion of the AA MSM population in San Francisco. In that case, our inference only applies to this segment of the target population, and the size of the target population is therefore larger than that which we estimated.

The SS-PSE R package, sspse,[16] is under ongoing development. Fixes to bugs in the program can affect the estimation of the posterior distribution. The results presented in this paper are output from model runs in July 2014. Bug fixes since then may have an impact on the replicability of these exact results. To improve the precision of the posterior estimation, we increased the MCMC settings. All results described here used a burn-in period of 10,000, a sample size of 1000, and an interval sampling of 100. Increasing the MCMC settings improves the precision of posterior estimates, at the cost of noticeably increasing computation time.

The SS-PSE method provides a simple and appealing tool to rapidly produce estimates of the size of high-risk populations - a fundamental public health measure that has been scarce for much of the HIV epidemic. Under the above outlined conditions, the SS-PSE method produced reasonable estimates for the size of the AA-MSM population in San Francisco, including lower and upper acceptable bounds. The model has the potential to be a useful addition to the repertoire of population size methods available to epidemiologists and other public health practitioners. The model especially has appeal because of its reasonable assumptions and seamless integration into RDS studies, which are commonly implemented to study hidden populations around the world.[38] The method has it limitations: First, the amount of information about population size in the RDS data is modest, so that the posterior distribution can have high variance. Secondly, the SS-PSE model will somewhat misspecify the actual RDS process, mainly due to the SS approximation to the RDS process. This will lead to some error of the posterior (as compared to the posterior based on the unknown RDS process). In addition, current concerns with regard to replication of results and manipulation of parameter inputs to adjust posterior estimates could make investigators vulnerable to confirmation bias. As a result, the appeal of this method should not obviate the planning for and use of multiple PSE methods to triangulate the most plausible size estimate for the target population. Combining multiple methods, as is often done in practice,[61] could balance and reduce the impact of bias on any one particular method. As the SS-PSE method produces a posterior distribution, it can be used as prior input to other methods using Bayesian inference.

Chapter 4

Evaluating the completeness of HIV surveillance using capture-recapture models, Alameda County, California

4.1 Introduction

Surveillance systems permit counties and states to estimate the incidence and prevalence of infectious diseases in the population and describe their local epidemiologic features by characterizing the populations most affected. Appropriate resource allocation, priority settings, and public health strategies are informed by accurate information from such surveillance systems. Within California, Alameda County ranked among the top five counties for cumulative number of AIDS cases, and among the top ten counties for cumulative number of persons living with HIV as of 31 December, 2013.[5] The Alameda County Public Health Department (ACPHD) estimated there were 5,649 people living with HIV/AIDS (PLWHA) in the county as of December 2013. They also estimated 230 new diagnoses annually. The incidence rates of HIV infection in Alameda County are highest among African Americans, men, and adults between the ages of 20-29 and 40-49 years. Eighty percent of HIV infections among men in Alameda County are associated with same-sex sexual contact. Half of HIV infections among women are associated with either high-risk heterosexual contact or injection drug use; the mode of HIV transmission for the remaining half is unknown.[35]

Laboratory reporting of HIV-infected individuals is a core component of HIV

surveillance for local and state health departments throughout the US. In Alameda County, laboratory tests (e.g., HIV antibody, CD4 cell count, and HIV viral load) are ordered by health care providers and sent to a laboratory. The results of these tests are sent back to the health care provider for clinical decision-making and disclosing to the patient, as well as to a surveillance clerk at ACPHD. By statute, health care providers must also report HIV-positive individuals to the ACPHD. When a laboratory report indicative of an HIV diagnosis is reported to ACPHD, the surveillance clerk will check the county database to determine if the individual was previously known to the county. If the individual has not previously been reported, the surveillance clerk will check the enhanced HIV/AIDS Reporting System (eHARS), the California Department of Public Health's (CDPH) database of known cases within the state. If the individual is not in eHARS, a public health investigator will follow up with the health care facility that ordered the laboratory test, gather additional information, update the county registry, and report the case to CDPH. This method of surveillance should capture all HIV-infected individuals within Alameda County for whom a laboratory test is ordered.

Although California statute mandates reporting all test results indicative of HIV infection to the county health department, the extent of underreporting and, by extension, the completeness of the surveillance system, is unknown. Not determining and accounting for potential under-reporting can result in biased estimates of disease burden. Additionally, such biases may under- or over- estimate disparities in accessing and retention in HIV care for population groups (e.g., sex, race, risk behavior, etc.). A formal evaluation of such surveillance systems is necessary both to describe accurately the epidemiologic features of HIV infection and to plan equitable distribution of health resources.

We used capture-recapture methods to estimate the universe of new diagnoses and PLWHA for whom ACPHD is responsible for conducting laboratory or case reporting to CDPH in 2013. This target population includes Alameda County residents receiving HIV-related health care services within the county, as well as residents of other counties who received HIV-related health care services within the county. The target population does not include Alameda County residents who received HIV-related health care services outside of Alameda County. By estimating the size of this target population, we can evaluate for the first time the completeness of the ACPHD HIV surveillance system. We also collected demographic information for the individuals on each list in order to estimate the size of population subgroups and determine if segments of the population were systematically underrepresented in the HIV surveillance system. We restricted our analysis to calendar year 2013 because it is recent enough to be relevant to describing the current population of PLWHA for which ACPHD is responsible, but enough time has passed to limit reporting delays.

4.2 Methods

Capture-recapture is a population size estimation method, originating in wildlife biology, now applied in public health to estimate the size of hidden and hard-to-reach populations.[25, 34, 56, 57, 7] The method is based on the amount of overlap in two samples the greater the overlap the smaller the population due to higher probability of capture in both samples. In its theoretically simplest form, a population is randomly sampled on two occasions, and those captured on either sampling occasion are marked in such a way that they can be identified if they appear on both capture occasions, or uniquely on only one capture occasion. Using the Lincoln-Petersen estimator, where n_{01} is the number of captures uniquely in the first source, n_{10} is the number of captures uniquely in the second source, and n_{11} is the number of captures from both sources, the estimated number not caught in either capture occasion can be calculated:[26]

$$\hat{n_{00}} = \frac{n_{01}n_{10}}{n_{11}} \tag{4.1}$$

The total population size is then calculated by summing the number of unique individuals observed on any capture occasion with the estimated count for the unobserved population. This method relies on four assumptions: (1) the target population is closed (there are no entries or losses during the study period); (2) there is no loss of tags (matching of cases that appear in different sources is complete); (3) for any single source, each case in the population has the same catchability or probability of ascertainment; and (4) for at least two sources, ascertainment of any case by each of the sources is independent.[23, 26]

In public health, separate lists of the target population (e.g., hospital registry, disease registry) can be used as capture occasions.[27] Such lists (or sources) may be correlated. For example, people with an advanced stage of disease may be more likely to appear on a list from a specialized clinic if their primary care physicians preferentially refer them to such facilities. Public health applications of capture-recapture are often at risk of bias, due to violations in the 3rd (capture homogeneity) and 4th (source independence) assumptions. Positive dependence between two sources will underestimate the target population size; negative dependence will overestimate the target population.[26, 63] Modeling approaches have been developed and applied to quantify and adjust for lack of source independence. For example, log-linear regression models can account for source dependency when three or more sources are included in the capture-recapture analysis. The bias due to source dependency is controlled by including interaction terms for parameters corresponding to individual

sources in the regression model. When moving from a two-source capture-recapture to a three-source capture-recapture analysis, the additional source/list provides additional degrees of freedom with which to estimate these parameters, and the log-linear regression model takes the form:[6]

$$\begin{aligned} logE(Z_{ijk}) = u + u_1 I(i=1) + u_2 I(j=1) + u_3 I(k=1) + u_{12} I(i=j=1) \\ + u_{13} I(i=k=1) + u_{23} I(j=k=1) + u_{123} I(i=j=k=1) \end{aligned} \tag{4.2}$$

Where $u_{\cdot}$ is the log expected count for the capture profile, indexed by the subscripts (e.g., u_1 refers to the log expected count for the number of individuals uniquely on list 1, u_{12} refers to the log expected count for the number of individuals on both list 1 and list 2). The intercept for this model, u, is the log expected count for the number of people not captured on any list. The equation above can be extended when including additional lists for the analysis and, for a given model, there are $2^n - 1$ degrees of freedom available to estimate the model parameters, where n refers to the number of lists. There are not enough degrees of freedom to estimate all possible model terms. By convention, researchers assume no n-way interaction. The number of potential models to fit (combinations of main terms and interaction terms) follows an exponential relationship with the number of lists available. For three lists, eight different models are possible; four lists result in 113 possible models to fit. Traditionally, the best fitting model is identified as the one with either the lowest Akaikes Information Criterion (AIC) or the lowest Bayesian Information Criterion (BIC), that is, statistics that balance the fit of the model to the observed data with a penalization for including more parameters in the model.[26]

Data Sources

We obtained six lists representing diverse segments of the target population to conduct our capture-recapture analysis.

Source 1 (S1) - A private hospital and part of a large HMO network in the San Francisco Bay Area.

Source 2 (S2) - A private hospital and tertiary care center; the list includes all patients from its HIV care clinic.

Source 3 (S3) - A Public Health hospital that serves as a safety net within the county. Patients are not required to have insurance to receive care. The hospitals list includes patients from the Emergency Department and the HIV care clinic. This hospital is also a part of the Alameda County Medical Center (ACMC) network, a network of public health hospitals. As such, it includes patients from HIV care

clinics at other hospitals within the network. Finally, this hospital serves as an AIDS Drug Assistance Program (ADAP) enrollment site; ADAP is a program that provides access to HIV medication for PLWHA who are uninsured or under-insured.

Source 4 (S4) - The ACPHD HIV surveillance list. A record of all HIV patients for whom a laboratory test was ordered within Alameda County and reported to the public health department. The list serves as a record of diagnosed HIV cases within Alameda County.

Source 5 (S5) - The Electronic Death Reporting System (EDRS). This list includes people who died in Alameda County and Alameda County residents who died outside of the county. Cases were included if there was any mention of HIV or AIDS under the causes of death (four fields for data entry) or under Significant Conditions.

Source 6 (S6) - ACPHD-funded HIV testing sites. Lists from three ACPHD-funded HIV testing sites were combined into a single list of unduplicated clients.

Each source provided a list of individual patients seen for HIV-related services from January 1, 2013 to December 31, 2013. At our request, lists included the following information: Patient name (first and last), Date of birth, Sex/Gender, Race and Ethnicity, and Patient HIV Risk History. These variables are among the standard patient data required for reporting a case on the AIDS Case Reporting Form to CDPH. The six individual lists were combined into a single, aggregated data set.

Record Linkage

Record linkage was achieved through a combination of manual and semi-automated matching algorithms on the aggregated data set. Manual record linkage was done using Microsoft Excel to sort observations by Patient name, Sex/Gender and Date of Birth to identify and link matches between lists. We used FRIL (Fine-grained Record Linkage), free open-source software for record linkage, to perform semi-automated matching.[32] Patient name and Date of birth were used for the semi-automated matching. A combination of exact match, distance matching, and Soundex (a phonetic algorithm to account for misspelled names) matching algorithms was used to identify possible matches. Matches identified by software were manually reviewed before confirming the match. After matches were identified and confirmed, patient identifiers (e.g., Name) were removed from the aggregated data set.

Capture-recapture analysis

R statistical software was used to perform the capture-recapture analysis.[46] We applied the Lincoln-Petersen estimator to pairwise combinations of lists to estimate

the unobserved population size of diagnosed cases of PLWHA for whom ACPHD was responsible for conducting laboratory or case reporting to ACPHD in 2013. We used the R package Rcapture [48] to fit log-linear regression models, controlling for potential source dependencies, and selected the best fitting models according to the lowest AIC and BIC. Associated confidence intervals were calculated using the profile likelihood.

We used the R package DGA [28] to calculate the population size using a Bayesian model averaging approach. DGA, Decomposable Graphs Approach, estimates a posterior probability distribution for the possible values of the population size for each decomposable graph, a model that specifies a dependency structure between lists.[36] The posterior probability distributions are averaged together and weighted by their marginal likelihoods to calculate a single posterior probability distribution for the population size. From this single posterior probability distribution, we calculated the mean and 95% credible interval to estimate the size of the target population.

Subgroup Analysis

To estimate the size of population subgroups, we stratified the data by demographic variables and applied the Bayesian model averaging approach within these strata. Using population size estimates from the subgroup analysis, we estimated four additional parameters to assess quantitatively whether any population subgroups were systematically under-represented in the HIV surveillance system.

Within each demographic category (e.g., Sex), Surveillance Ratios compare the number of individuals with a given characteristic (e.g., Females) on the laboratory list to the number of individuals with the reference characteristic (e.g., Males) on the laboratory list. Target Population ratios calculate ratios of the same subcategories as Surveillance Ratios, but use the estimated number in the target population, taken from the results of the DGA model. Detection Ratios are ratios of the proportion of individuals from a target population with a given characteristic who are captured on the surveillance list (e.g., Females) relative to the same proportion for the reference characteristic (e.g., Males).

Ascertainment-corrected adjusted Detection Ratios (ACADR) calculate the probability that the HIV surveillance system will detect an individual from the target population with a given characteristic (e.g., Females) relative to the reference characteristic (e.g., Males), controlling for all other measured characteristics. To calculate this parameter, the marginal distribution in the target population was determined for each measured characteristic, after subtracting from the total (estimated) population the portion accounted for by the laboratory-based surveillance list. These marginal distributions were used to calculate sampling weights, which were then

applied to the analytic data set, excluding individuals observed on the laboratory-based surveillance list (regardless of whether or not they were also observed on any combination of the facility-based lists). The subset of the aggregated data set not accounted for by the laboratory-based surveillance list is then weighted to look like the portion of the target population not explicitly captured by the surveillance list. A modified Poisson regression,[66] using generalized estimating equations (GEE) with an exchangeable correlation structure and robust standard errors, was implemented to model the probability that an individual with a given demographic characteristic would be on the laboratory list (detected by the laboratory) relative to the reference category for that demographic category, holding constant all other measured characteristics.

Sensitivity Analysis

As a sensitivity analysis, we followed the recommendation of Cormack et al.[9] that removing the list with the most complete coverage of the target population results in more plausible population size estimates. We fit log-linear regression models using S1, S2, and S3, accounting for all combinations of source dependencies.

Positivity violations (i.e., zero cell counts for several list intersections) make a subset of the 113 possible log-linear regression models for a four-source capture-recapture analysis unidentifiable. We manually fit log-linear regression models for all the remaining possible models, assessing their point estimate for the population size, 95% confidence intervals, and model fit according to the AIC.

Ethics Statement

The study received ethical review and approval from the University of California, Berkeley Office for the Protection of Human Subjects.

4.3 Results

5,376 unique individuals were identified from the capture-recapture sampling of the ACPHD HIV laboratory-based surveillance list (S4) and the three facility-based lists (S1-S3). An additional 16 individuals were uniquely included on the EDRS list (S5), and 12 were uniquely included on the HIV testing list (S6). Due to small sample size and data quality concerns, we excluded the EDRS and HIV testing sites from the statistical analysis. The largest proportion of the study population was accounted for by the laboratory-based surveillance list (n=4,979); 80% of individuals in this

aggregated data set were males, while 42% were black and 33% were white. Nearly half (47%) of the individuals in the data set were 50 years of age. Over half (58%) had male-male sexual contact (i.e., MSM) as the HIV transmission risk. Table 4.1 describes the demographic characteristics of the study population, stratified by the four lists; p-values for the chi-squared statistic for each of the demographic characteristics indicate heterogeneity between lists with respect to composition. Figure 4.1 illustrates the four-list capture-profile as a Venn diagram. Each oval represents either the laboratory-based surveillance list or one of the three facility-based lists; numbers within list intersections indicate the number of unique individuals identified from that combination of lists. The facility-based lists revealed 397 unique individuals who were not previously identified by the laboratory-based surveillance list.

	Source 1 (%)	Source 2 (%)	Source 3 (%)	Source 4 (%)	Chi-squared
Sex					p=9.49e-11
Male	766 (0.84)	707 (0.75)	771 (0.74)	4,040 (0.81)	
Female	144 (0.15)	237 (0.25)	266 (0.26)	939 (0.19)	
Race/Ethnicity					p=2.2e-16
NH[1]White	395 (0.43)	347 (0.37)	171 (0.16)	1,660 (0.33)	
NH Black	342 (0.38)	410 (0.43)	569 (0.55)	2,105 (0.42)	
Hispanic	113 (0.12)	114 (0.12)	218 (0.21)	818 (0.16)	
Asian	41 (0.05)	36 (0.04)	47 (0.05)	202 (0.04)	
Am. Ind.	<10 (<0.01)	<10 (<0.01)	<10 (<0.01)	17 (<0.01)	
Pac. Isl.	14 (0.02)	15 (0.02)	21 (0.02)	91 (0.02)	
Other	<10 (<0.01)	<10 (<0.01)	<10 (<0.01)	69 (0.01)	
Unknown	<10 (<0.01)	12 (0.01)	<10 (<0.01)	<17 (<0.01)	
Age Cat. (years)					p=2.2e-16
<19	<10 (<0.01)	10 (0.01)	0 (<0.01)	57 (0.01)	
20-29	43 (0.05)	131 (0.14)	84 (0.08)	451 (0.09)	
30-39	88 (0.10)	115 (0.12)	206 (0.20)	688 (0.14)	
40-49	272 (0.30)	231 (0.25)	346 (0.33)	1,388 (0.14)	
50-59	312 (0.34)	294 (0.31)	288 (0.28)	1,575 (0.28)	
60+	193 (0.21)	163 (0.17)	113 (0.11)	820 (0.16)	
HIV Risk					p=2.2e-16
Het. contact	138 (0.15)	226 (0.24)	298 (0.29)	800 (0.16)	
MSM[2]	624 (0.69)	482 (0.51)	443 (0.43)	2,969 (0.60)	
IDU[3]	30 (0.03)	96 (0.10)	119 (0.11)	421 (0.08)	
MSM & IDU	64 (0.07)	75 (0.08)	90 (0.09)	395 (0.08)	
Medical	<10 (<0.01)	<10 (<0.01)	<10 (<0.01)	24 (<0.01)	
Other/Unknown	47 (0.05)	56 (0.06)	82 (0.08)	370 (0.07)	
Total	910	944	1,037	4,979	

Table 4.1: Demographic characteristics and HIV risk history of persons living with HIV/AIDS, stratified by reporting source, from four sources in Alameda County, California, 2013

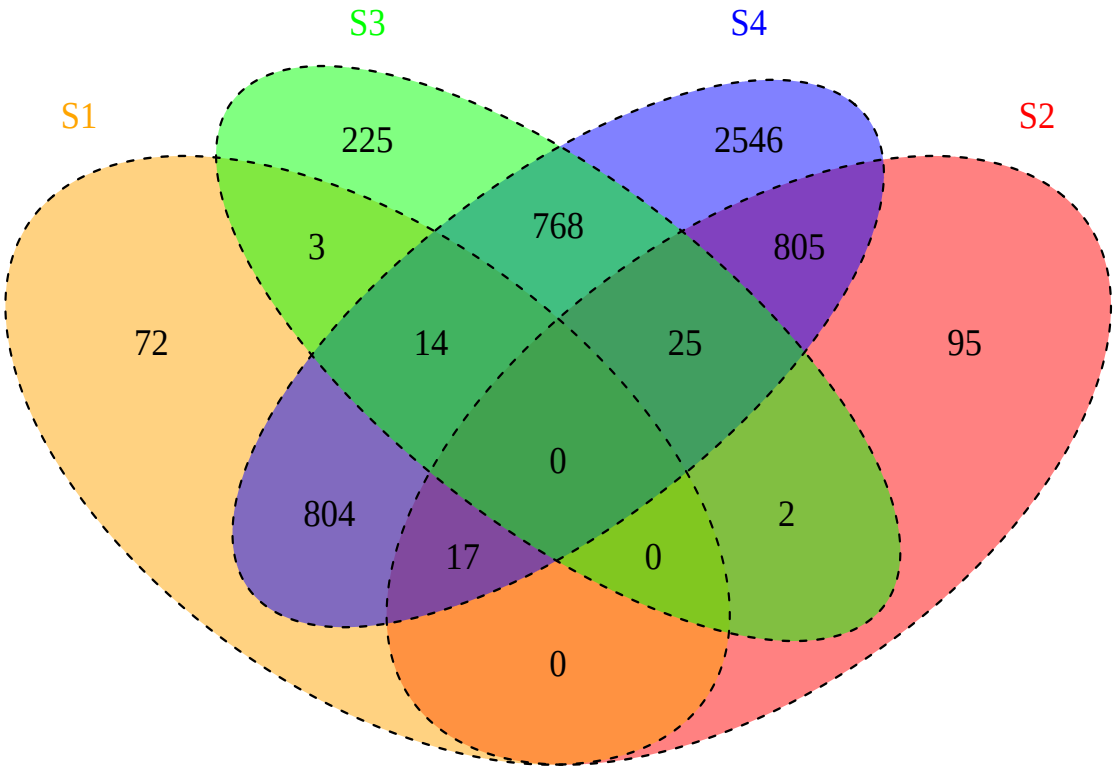


Figure 4.1: Four-source capture profile, persons living with HIV/AIDS, Alameda County, CA, 2013

Table 4.2 provides estimates of the unobserved population size and the total population size using the Lincoln-Petersen estimator. Two-source capture-recapture analysis using any combination of the facility-based lists with each other indicated negative source dependence given the magnitude of the estimated size for the unobserved population (S1*S2: $\hat{N}$=48,695; S1*S3: $\hat{N}$=53,580; S2*S3: $\hat{N}$=34,303). Two-source capture-recapture analysis between the laboratory-based surveillance list and

[1]non-Hispanic
[2]men who have sex with men
[3]injection drug users

S1, S2, or S3 estimated the total size of the target population to be 5,426, 5,549, and 6,398, respectively.

Source A	Source B	n_{10}	n_{01}	n_{11}	n_{00}	$\hat{N}$
S1	S2	893	927	17	48,695	50,532
S1	S3	893	1,020	17	53,580	55,510
S2	S3	917	1,010	27	34,303	36,257
S4	S1	4,144	75	835	372	5,426
S4	S2	4,132	97	847	473	5,549
S4	S3	4,172	230	807	1,189	6,398

Table 4.2: Estimate of the unobserved population size, total size of the diagnosed PLWHA population under Alameda County, CA, public health jurisdiction in 2013 (Lincoln-Petersen estimator)

Log-linear regression models were used to incorporate information from all four lists and to model the source dependencies (Table 4.3). The log-linear model assuming independence between the four lists estimated the target population size to be 5,943 (95% CI: 5,867-6,023). The best fitting models, determined by each models AIC, are listed in Table 3. According to the AIC criterion, the best fitting model estimates the total population size to be 6,124 (95% CI: 6,003 - 6,256), indicating the laboratory-based surveillance system to be 81.3% complete (number of unique individuals listed on the laboratory-based surveillance list/estimated size of the target population). The remaining four best fitting log-linear models provide similar estimates of the total population size, ranging from 6,092 to 6,124), with the fifth best fitting model estimating the population size at 5,604 (95% CI: 5,544 - 5, 670).

Model[4]	$\hat{N}$	95% CI	AIC	df	%Complete (laboratory)
Observed counts	5,376		–		92.6%
Independence	5,943	5,867-6,023	974.48	10	83.8%
Base[5]+ S1*S2 + S1*S3 + S1*S4 + S2*S3 + S2*S4	6,124	6,003-6,256	97.92	5	81.3%
Base + S1*S2 + S1*S4 + S2*S4 + S1*S2*S4 + S1*S3 + S2*S3	6,122	6,001-6,254	98.7	4	81.3%
Base + S1*S2 + S1*S3 + S2*S3 + S1*S2*S3 + S1*S4 + S2*S4	6,124	6,003-6,256	99.85	4	81.3%
Base + S1*S2 + S1*S3 + S1*S4 + S2*S3 + S2*S4 + S3*S4	6,092	5,654-7,212	99.91	4	81.7%
Base + S2*S3 + S2*S4 + S3*S4 + S2*S3*S4 + S1*S2 + S1*S3	5,604	5,544-5,670	100.89	4	88.8%
DGA	5,720	5,587-6,190	–	–	87.0%

Table 4.3: Log-linear regression and DGA model estimates of the total size of the diagnosed PLWHA population under Alameda County, CA, public health jurisdiction in 2013, and completeness of laboratory surveillance list (four-source capture-recapture model)

A Bayesian model averaging approach using all models identified by decomposable graphs and weighted by their marginal likelihood provided a single estimate for the size of the target population ($\hat{N}$=5,720) and the corresponding 95% credible interval (5,587 - 6,190) (Table 4.3, Figure B.1). The DGA model revealed a bimodal posterior probability distribution for the estimated population size. Peaks in the posterior probability distribution indicate likely values of the parameter of interest, given the data; a bimodal distribution indicates two values for the population size with high probability. 81% of the posterior probability distribution was attributed to a population size estimate of 5,638; 17% of the posterior probability distribution was attributed to a population size estimate of 6,123. Results from the DGA model indicated that the laboratory-based surveillance system was 87% complete.

Using the Bayesian model averaging approach, stratified population size estimates were calculated for the measured demographic characteristics. Table 4.4 describes the target population according to these demographic characteristics, comparing the number observed in the laboratory-based surveillance system to the estimated size, according to the DGA model. Females in the estimated target population were 12% more likely than males to be detected by the surveillance system (ACADR 1.12, 95% CI: 1.08 - 1.17). Non-Hispanic Blacks were 4% more likely than non-Hispanic Whites to be detected by the surveillance system (ACADR 1.04, 95% CI: 1.02 - 1.06), whereas Hispanics were 4% less likely to be detected by the surveillance system compared to non-Hispanic Whites (ACADR 0.96, 95% CI: 0.94 - 0.99). Increasing age categories were positively correlated with ACADRs relative to the 29 age group holding other variables constant, although a statistically significant association was found only for the 60 years age group (ACADR 1.21, 95% 1.15 - 1.27). All measured HIV risk groups were significantly more likely to be detected by the surveillance list, relative to transmission via heterosexual contact.

[4]The top five best fitting hierarchical models by AIC criterion are also among the top ten best fitting models according to BIC criterion

[5]"Base" = S1+S2+S3+S4

Stratified population	Observed (laboratory)	$\hat{N}$ (95% CI)	Surveillance Ratio	Target Population Ratio	Detection Ratio	Ascertainment-corrected adjusted Detection Ratio (95% CI)
Sex						
Male	4,040	4,658 (4,503-5,072)	REF	REF	REF	REF
Female	939	1,090 (1,067-1,126)	0.23	0.23	0.99	1.12 (1.08-1.17)
Race						
White	1,660	1,864 (1,802-2,069)	REF	REF	REF	REF
Black	2,105	2,354 (2,312-2,405)	1.27	1.26	1.00	1.04 (1.02-1.06)
Hispanic	818	963 (941-988)	0.49	0.52	0.96	0.96 (0.94-0.99)
Asian	293	345 (328-366)	0.18	0.19	0.95	1.02 (0.98-1.06)
Other	103	275 (188-368)	0.06	0.15	0.42	0.94 (0.87-1.02)
Age Cat.						
≤ 29	508	604 (576-642)	REF	REF	REF	REF
30-39	688	849 (808-940)	1.35	1.41	0.96	0.98 (0.94-1.02)
40-49	1,388	1,612 (1,582-1,648)	2.73	2.67	1.02	1.00 (0.96-1.04)
50-59	1,575	1,744 (1,713-1,780)	3.10	2.89	1.07	1.04 (1.00-1.07)
≥ 60	820	880 (863-906)	1.61	1.46	1.11	1.04 (1.01-1.08)
HIV Risk						
Het. contact	800	1,213 (1,072-1,285)	REF	REF	REF	REF
MSM	2,969	3,186 (3,092-3,243)	3.71	2.63	1.41	1.33 (1.26-1.40)
IDU	421	438 (432-457)	0.53	0.36	1.46	1.25 (1.19-1.31)
MSM & IDU	395	421 (407-437)	0.49	0.35	1.42	1.34 (1.27-1.42)
Other	394	454 (431-569)	0.49	0.37	1.32	1.21 (1.15-1.27)

Table 4.4: DGA model-based estimates of the size and detectability of demographic and HIV risk subpopulations among the diagnosed PLWHA population under Alameda County, CA, public health jurisdiction, 2013[6]

Sensitivity analyses using other models did not generate plausible size estimates. These models are given as supplementary appendices (Table B.1, Figure B.2, Table B.2).

4.4 Discussion

We estimate Alameda Countys laboratory-based HIV surveillance system to be 87% complete. That is, there were 5,720 persons diagnosed with HIV and receiving treatment in Alameda County, with 4,979 reported in the laboratory-based system. Estimates from individual models calculated using the DGA R package were consistent with the best fitting log-linear regression models. That is, a high probability was attributed to a population size estimate of 5,638 (81%) and to a size estimate of 6,123 (17%). Selecting a single best fitting model according to the AIC or BIC criterion does not account for the likelihood of that model, relative to the likelihood of competing models. The DGA package allows each model to contribute to the final estimate by calculating and weighting each model by its marginal likelihood when averaged together into a single posterior distribution. This approach accounts for multiple likely estimates of the true population size, and the uncertainty in model selection, while the AIC/BIC model selection criterion does not.

We identified 397 unique individuals included on one or more of the facility-based lists, but not included on the HIV surveillance list. There are several potential reasons why the HIV surveillance list may have missed 13% of the target population. First, not all diagnosed cases of HIV infection are documented with a laboratory test. Health care providers may forego a laboratory test for visits by PLWHA for a number of reasons; for example, a patient may already be actively engaged in care or be visiting the health care facility only to renew a prescription. Second, at some health care facilities, laboratory tests are not conducted in the same location as the visit with the health care provider. Patients may not always follow through with the ordered test. Third, not all visits to healthcare facilities are medical visits; they may be for social services, for example. Finally, laboratory tests ordered in the context of a clinical trial are exempt from mandated reporting to the surveillance system. In these scenarios, PLWHA known to healthcare facilities remain virtually invisible to the HIV surveillance system.

We also found that as a sampling mechanism, the laboratory-based surveillance system captures for the most part a representative cross-section of the population

[6] "Asian" Race category includes Asian and Pacific Islander categorizations. "Other" Race category includes Other, Unknown, and American Indian categorizations. "Other" HIV Risk category includes Medical and Other categorizations.

of PLWHA in the county. Nonetheless, our analyses suggest that groups at high risk for HIV infection and groups known to carry a disproportionate burden of HIV infection, such as racial minorities, MSM, and injection drug users (IDU), have a higher probability of inclusion. Women and persons of older age were also were more likely to be reported. For example, although there are four times as many males as females on both the surveillance list and in the estimated target population, females were 12% more likely to be detected by the HIV surveillance list compared to males. The mechanism supporting the HIV surveillance system laboratory reporting of HIV infected individuals - depends on health care providers ordering laboratory tests for their patients. Differences in the probability of detection between demographic subgroups may reflect a higher propensity of health care providers to order laboratory tests for patients who are in certain demographic categories (e.g., females, blacks, older age, MSM, IDU),. This apparent over-sampling of minority and marginalized populations, who have been known to bear a disproportionate burden of HIV infection, could reflect the success of public health programs with respect to reaching high risk groups.

Contrary to previous recommendations by Cormack et al. to remove the list with the most complete coverage [9] we found that removing the HIV surveillance list resulted in improbably large estimates of the size of the target population (Table B.1). Seven of the eight possible log-linear models estimated the size of the population to be greater than 35,000 individuals, with the majority of models estimating a population size ten times as large as the number on the HIV surveillance list. The eighth model, the worst fitting according to the AIC, estimated the population size to be 2,830, or 57% of what was observed on the HIV surveillance list and therefore not plausible. In our study, including the list with the most complete coverage was necessary to produce plausible estimates of the population size. As previously discussed, all pairwise combinations of the three facility-based lists indicated negative source dependency. In a three-source capture-recapture analysis, in which all lists are negatively dependent with respect to one another, there were not enough degrees of freedom to control for all source dependencies. The addition of the HIV surveillance list, a list that statistically appears to be independent of the facility-based lists, provided additional degrees of freedom to control the source dependencies adequately. Our experience suggests that at least one source, perhaps the source with the most complete coverage, must be independent of the other sources in order to calculate plausible population size estimates. Investigators must carefully consider the source dependencies during the study design stage when selecting sources and check their assumptions about source dependencies by applying the Lincoln-Peteresen estimator to all pair-wise combinations. Of note, the Bayesian model averaging approach yielded a low point estimate of 3,208 for the population size. Although this was 64%

of what was observed on the HIV surveillance list, the 95% credible interval (2,830 - 6,370) included our best estimate from the four-source model.

Also contrary to past recommendations,[9, 24] we found that the saturated model, the model with all possible interaction terms to model source dependencies, was not optimal. Because there was no overlap of HIV positive cases on S1 and S2, models including the S1*S2 interaction were not identifiable. For the same reason, models including the S1*S2*S3 interaction were also not identifiable As a sensitivity analysis, we calculated the 41 remaining identifiable models (out of 113 possible models for a four-source capture-recapture analysis) (Figure B.2, Table B.2). As previously discussed, source dependence was only indicated for each pair of the facility-based lists (S1*S2, S1*S3 and S2*S3, with S1*S2 not identifiable due to zero overlap). Modeling additional source dependencies, thereby moving closer to a fully saturated model, generally increased both the bias and the variance of the estimated population size, compared to the DGA model, our best estimate. Before modeling all possible interactions between sources, scientists should assess the empirical evidence for source dependencies, and let that knowledge inform which parameters are explicitly modeled.

Similar to the source independence assumption, violations of the capture homogeneity assumption may be addressed analytically. The capture homogeneity assumption assumes that within a given source, individuals have the same probability of capture; that is, individual characteristics do not influence the probability that someone will be observed on a particular list. The distribution of demographic characteristics differs between lists in this study, and affects the probability of detection by a particular source. One strategy to control for heterogeneity in capture probabilities is to stratify the data by variables that are believed to be the source of the heterogeneity and proceed with the capture-recapture analysis, estimating the population size within homogenous groups. For our study, we stratified the data by single demographic characteristics (e.g., sex) and estimated the population sizes within these subgroups. The sum of the population size of the mutually exclusive demographic subgroups was consistent with the total estimated population size. This approach does not control simultaneously for all measured sources of capture heterogeneity. Stratifying on all combinations of measured demographic covariates, however, would have resulted in cells that were too sparse to use in the analysis. Furthermore, such an approach assumes that the measured demographic variables (sex, race, age, HIV risk) were the only sources of heterogeneity in the capture probabilities. It is likely that other variables for which we were not able to collect information (e.g., insurance status) also influence the individual capture probability within a given source. In this scenario, grade of membership and other latent class modeling techniques, which identify homogenous latent groups (pure classes) within the

population and allow individuals to have partial membership in all classes, may be appropriate analytic strategies to control for both measured and unmeasured sources of capture heterogeneity.[2, 39, 40]

Unlike violations of non-independence and capture heterogeneity, there is unfortunately no statistical solution for inaccuracies in record linkage. We used a combination of manual and semi-automated procedures to identify the same individual on multiple lists. The HIV surveillance list included AKAs, alternate names for individuals, which improved our ability to identify the same individuals on multiple lists. Several investigators were involved in cross-checking and confirming the record linkage, bolstering the accuracy of the matching process. Despite these efforts, it is still possible that our team did not successfully identify all matches between lists. This potential failure to identify all matches is of particular concern if AKAs are not documented on either the HIV surveillance list or another registry accessible to ACPHD. Anecdotal evidence from individual sites indicates that false identifiers occur, especially for patients who are undocumented foreigners or transgender individuals. For this analysis, we must assume that if a patient gives a false identifier, that same false identifier is consistently given if they visit multiple sites. Evidence of the negative source dependencies between health care facilities included in this analysis relaxes this concern to some degree, as individuals on one facility-based list are very unlikely to appear on another facility-based list.

The closed population assumption assumes that there are no changes to the target population, which may result in some people having a zero probability of inclusion on a given list. In our study design, we attempted to meet this assumption by focusing on a well-defined time window within which to estimate the population size (calendar year 2013). The death registry indicates that ours is not a true closed population; 57 PLWHA died in 2013, 16 of whom did not appear on any other list included in this analysis. However, there were no temporal constraints on any of the lists included in this analysis. Each list was active in sampling from the target population throughout the 2013 calendar year. Therefore, at any given time during this study period, individual members of the target population had a non-zero probability of being sampled by any of the lists included in this analysis. For this reason, we do not believe that the dynamics of this target population violate the closed population assumption in such a way that would bias our estimates of the population size.

Capture-recapture analysis implicitly assumes that the unobserved population is similar to the observed population with respect to both measured and unmeasured characteristics. The results of our analysis apply to the population of diagnosed PLWHA engaging with the health care system within Alameda County in 2013. Our analysis benefits from accessing four diverse lists covering overlapping, but sometimes different, segments of the target population. The lists differ with respect to

the distribution of race/ethnicity, insurance status, and reason for detection (e.g., medical visit with or without an accompanying laboratory test, social services visit, prescription renewal). Therefore, we are confident that our aggregated data set is highly representative of the target population. However, if there are certain types of people who have a zero probability of appearing on any of these lists, essentially making them invisible to the public health surveillance system, they would not be included in the estimation of the size of the target population.

In addition to collecting diverse lists, our study benefited from collecting information on measured demographic characteristics, making subgroup analyses possible. Results from the subgroup analysis, however, are dependent upon accurate data collection at the health care facilities. Agreement in recorded information between lists, observed during the record linkage stage of the study, suggests that basic demographic information (e.g., date of birth, race) is accurately recorded. Information that is more sensitive, such as HIV transmission risk, may be recorded with less accuracy if patients are concerned about stigma (e.g., homosexual contact or injection drug use). If this is true, as suggested anecdotally at some sites, then our data may be subject to misclassification with regard to transmission risk categories, such as MSM and IDU being misclassified as transmission due to heterosexual contact. This would result in an overestimation of the population subgroup infected with HIV through heterosexual contact. Such concerns reinforce the importance of accurate data collection by the health care provider at the point of contact with the patient.

Laboratory testing and reporting of test results are commonly used for HIV surveillance by public health agencies throughout the United States.[13] Our study suggests that, while not a complete census of the target population, this method of surveillance presents a mostly representative sample of the target population. Parameters such as the ACADR can be useful when exploring the existence of underserved populations, which may have implications for the equitable provision of and access to public health resources. Routine evaluations of laboratory-based surveillance systems are important for accurately documenting the local HIV disease burden. Public health should invest in resources that facilitate such evaluations, such as user-friendly software common to all local health care facilities to collect data on their patient population.

Chapter 5

Conclusion

Globally, HIV infection is often concentrated among hidden populations, populations existing on the margins of society due to stigma and discrimination. Population size estimation of these key populations has become an emergent priority for the global public health community to manage the current epidemic and to marshal resources to prevent new infections. Knowing the size of the target population is a necessary first step in understanding the epidemiologic features of disease in different settings. Population size estimation methods are an important set of tools for epidemiologists and other public health researchers to accomplish this fundamental task in public health surveillance. As demonstrated in this dissertation, these methods can be used to not only enumerate a population, but also to recover plausible population characteristics despite biased sampling and the absence of a sampling frame. The breadth of size estimation methods available allows epidemiologists to estimate population sizes under various scenarios with respect to sampling strategy and study design. However, the lack of consistent agreement in size estimates between population size estimation methods is of great concern and can have negative implications for public health professionals and community members who rely on these population estimates to set targets for HIV testing and treatment, and to advocate for sufficient resources to control the local burden of disease. As discussed and demonstrated in this dissertation, there is no universal best method to estimate a population size. Until the nature of the differences in estimates from different methods applied to the same population is fully understood, population size estimation studies should employ multiple different methods (or modeling techniques) to qualify the reliability of reported estimates of the population size.

Bibliography

[1] Abu S Abdul-Quader, Andrew L Baughman, and Wolfgang Hladik. "Estimating the size of key populations: current status and future possibilities." In: *Current opinion in HIV and AIDS* 9 (2014), pp. 107–14. ISSN: 1746-6318. DOI: 10.1097/COH.0000000000000041.

[2] Dankmar Baehning. "Editorial Recent Developments in Capture-Recapture Methods and Their Applications". In: 50 (2008), pp. 954–956.

[3] Le Bao, Adrian E Raftery, and Amala Reddy. "Estimating the Size of Populations at High Risk of HIV in Bangladesh Using a Bayesian Hierarchical Model". In: (2010).

[4] U.S. Census Bureau. *SEX BY AGE Universe : Total population 2006-2010 American Community Survey Selected Population Tables*. 2014.

[5] HIV/AIDS Surveillance Section California Department of Public Health, Office of AIDS. *HIV/AIDS Surveillance in California*. Tech. rep. Sacramento, CA: California Department of Public Health, 2014, pp. 1–8.

[6] A Chao et al. "The applications of capture-recapture models to epidemiological data." In: *Statistics in medicine* 20.20 (Oct. 2001), pp. 3123–57. ISSN: 0277-6715.

[7] Anton W Moll van Charante and Paul G Mulder. "Reporting of Industrial Accidents in the Netherlands". In: *American journal of epidemiology* 148.2 (1998), pp. 182–190.

[8] Office of the U.S. Global AIDS Coordinator. *PEPFAR 3.0 - Controlling the Epidemic: Delivering on the Promise of an AIDS-free Generation*. Tech. rep. 2014, pp. 1–32.

[9] Richard M Cormack, Yue-fang Chang, and Gordon S Smith. "Estimating deaths from industrial injury by capture-recapture : a cautionary tale". In: (2000), pp. 1053–1059.

[10] Vincent Fuqua et al. "Using social networks to reach Black MSM for HIV testing and linkage to care." In: *AIDS and behavior* 16.2 (Feb. 2012), pp. 256–65. ISSN: 1573-3254. DOI: 10.1007/s10461-011-9918-x.

[11] Krista J. Gile. "Improved Inference for Respondent-Driven Sampling Data With Application to HIV Prevalence Estimation". In: *Journal of the American Statistical Association* 106.493 (Mar. 2011), pp. 135–146. ISSN: 0162-1459. DOI: 10.1198/jasa.2011.ap09475.

[12] Krista J Gile and Mark S Handcock. "Respondent-Driven Sampling: An Assessment of Current Methodology." In: *Sociological methodology* 40.1 (Aug. 2010), pp. 285–327. ISSN: 0081-1750. DOI: 10.1111/j.1467-9531.2010.01223.x.

[13] H Irene Hall et al. "Assessing the completeness of reporting of human immunodeficiency virus diagnoses in 2002-2003: capture-recapture methods." In: *American journal of epidemiology* 164.4 (Aug. 2006), pp. 391–7. ISSN: 0002-9262. DOI: 10.1093/aje/kwj216.

[14] Ghassan Hamra, Richard MacLehose, and David Richardson. "Markov chain Monte Carlo: an introduction for epidemiologists." In: *International journal of epidemiology* 42.2 (Apr. 2013), pp. 627–34. ISSN: 1464-3685. DOI: 10.1093/ije/dyt043.

[15] Gile Krista J Handcock Mark S, Fellows Ian E. *RDS Analyst: Software for the Analysis of Respondent-Driven Sampling Data, Version 0.42*. 2014.

[16] MS Handcock and KJ Gile. *sspse: estimating hidden population size using respondent driven sampling data.* 2015.

[17] MS Handcock, KJ Gile, and CM Mar. "Estimating Hidden Population Size using Respondent-Driven Sampling Data". In: *arXiv preprint arXiv:1209.6241* (2012).

[18] MS Handcock, KJ Gile, and CM Mar. "Estimating the size of populations at high risk for HIV using respondene-driven sampling data". In: *Biometrics* (2015).

[19] Gordon Hay et al. "Capture–recapture and anchored prevalence estimation of injecting drug users in England: national and regional estimates." In: *Statistical methods in medical research* 18.4 (Aug. 2009), pp. 323–39. ISSN: 0962-2802. DOI: 10.1177/0962280208094687.

[20] DD Heckathorn. "Respondent-driven sampling: a new approach to the study of hidden populations". In: *Social problems* 44 (1997), pp. 174–199.

[21] DD Heckathorn. "Respondent-driven sampling II: deriving valid population estimates from chain-referral samples of hidden populations". In: *Social problems* 49.1 (2002), pp. 11–34.

[22] T G Heckman et al. "HIV risk differences between African-American and white men who have sex with men." In: *Journal of the National Medical Association* 91.2 (Feb. 1999), pp. 92–100. ISSN: 0027-9684.

[23] E B Hook and R R Regal. "Capture-recapture methods in epidemiology: methods and limitations." In: *Epidemiologic reviews* 17.2 (Jan. 1995), pp. 243–64. ISSN: 0193-936X.

[24] E B Hook and R R Regal. "Validity of methods for model selection, weighting for model uncertainty, and small sample adjustment in capture-recapture estimation." In: *American journal of epidemiology* 145.12 (June 1997), pp. 1138–44. ISSN: 0002-9262.

[25] Ernest B Hook and Agneta Lindsjo. "Down Syndrome in Live Births by Single Year Maternal Age Interval in a Swedish Study : Comparison with Results from a New York State Study". In: *American Journal of Human Genetics* 30 (1978), pp. 19–27.

[26] International Working Group for Disease Monitoring and Forcasting. "Capture-recapture and multiple-record systems estimation. I: History and theoretical development". In: *American Journal of Epidemiology* 142.10 (1995), pp. 1047–1058.

[27] International Working Group for Disease Monitoring and Forcasting. "Capture-Recapture and Multiple-Record Systems Estimation II: Applications in Human Diseases". In: *American Journal of Epidemiology* 142.10 (1995), pp. 1059–1068.

[28] James Johndrow, Kristian Lum, and Patrick Ball. *dga: Capture-Recapture Estimation using Bayesian Model Averaging. R package version 1.2*. 2015. URL: `http://cran.r-project.org/package=dga`.

[29] Lisa G. Johnston et al. "Estimating the Size of Hidden Populations Using Respondent-driven Sampling Data". In: *Epidemiology* 26.6 (2015), p. 1. ISSN: 1044-3983. DOI: `10.1097/EDE.0000000000000362`.

[30] Lisa G Johnston et al. "Incorporating the service multiplier method in respondent-driven sampling surveys to estimate the size of hidden and hard-to-reach populations: case studies from around the world." In: *Sexually transmitted diseases* 40 (2013), pp. 304–10. ISSN: 1537-4521. DOI: `10.1097/OLQ.0b013e31827fd650`.

[31] H. E. Jones et al. "Recapture or Precapture? Fallibility of Standard Capture-Recapture Methods in the Presence of Referrals Between Sources". In: *American Journal of Epidemiology* 179.11 (Apr. 2014), pp. 1383–1393. ISSN: 0002-9262. DOI: 10.1093/aje/kwu056.

[32] Pawel Jurczyk et al. *FRIL: Fine-Grained Records Integration and Linkage Tool v. 3.2*. 2009. URL: http://fril.sourceforge.net/.

[33] Carl Kendall et al. "An empirical comparison of respondent-driven sampling, time location sampling, and snowball sampling for behavioral surveillance in men who have sex with men, Fortaleza, Brazil". In: *AIDS and Behavior* 12.SUPPL. 1 (2008), pp. 97–104. ISSN: 10907165. DOI: 10.1007/s10461-008-9390-4.

[34] A Larson, A Stevens, and G Wardlaw. "Indirect estimates of 'hidden' populations: capture-recapture methods to estimate the numbers of heroin users in the Australian Capital Territory." In: *Social science & medicine (1982)* 39.6 (Sept. 1994), pp. 823–31. ISSN: 0277-9536.

[35] Richard Lechtenberg, Neena Murgai, and HIV Epidemiology and Surveillance Unit. "HIV in Alameda County". In: *Annual Epidemiology Data Presentation to the CCPC*. Oakland, CA, 2015. URL: http://hivccpc.org/wp-content/uploads/2016/01/Epi-data-presentation-to-the-CCPC-July-22-2015.pdf.

[36] David Madigan and Jeremy York. "Bayesian methods for estimation of the size of a closed population". In: *Biometrika* 84.1 (1997), pp. 19–31.

[37] Robert Magnani et al. "Review of sampling hard-to-reach and hidden populations for HIV surveillance." In: *AIDS (London, England)* 19 Suppl 2 (May 2005), S67–72. ISSN: 0269-9370.

[38] Mohsen Malekinejad et al. "Using respondent-driven sampling methodology for HIV biological and behavioral surveillance in international settings: a systematic review." In: *AIDS and behavior* 12.4 Suppl (July 2008), S105–30. ISSN: 1090-7165. DOI: 10.1007/s10461-008-9421-1.

[39] Daniel Manrique-Vallier and Stephen E Fienberg. "Population size estimation using individual level mixture models." In: *Biometrical journal. Biometrische Zeitschrift* 50.6 (Dec. 2008), pp. 1051–63. ISSN: 1521-4036.

[40] Daniel Manrique-vallier and Stepthen E Fienberg. "Modeling heterogeneity using the Grade of Membership model The Grade of Membership model". In: *Biometrical Journal* (2008), pp. 1–19.

[41] Nicky McCreesh et al. "Evaluation of respondent-driven sampling." In: *Epidemiology (Cambridge, Mass.)* 23.1 (Jan. 2012), pp. 138–47.

[42] Christopher J L Murray et al. "Disability-adjusted life years (DALYs) for 291 diseases and injuries in 21 regions, 1990-2010: a systematic analysis for the Global Burden of Disease Study 2010." In: *Lancet* 380.9859 (Dec. 2012), pp. 2197–223. ISSN: 1474-547X.

[43] Vijayan N Nair and Paul CC Wang. "Maximum Likelihood Estimation Under a Successive Model Discovery Sampling". In: *Technometrics* 31.4 (1989), pp. 423–436.

[44] Jerry Okal 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." In: *Sexually transmitted infections* 89.5 (Aug. 2013), pp. 366–71. ISSN: 1472-3263.

[45] Nancy S Padian et al. "HIV prevention transformed: the new prevention research agenda". In: *Lancet* 378.9787 (July 2011), pp. 269–78. ISSN: 1474-547X.

[46] R Core Team. *R: A Language and Environment for Statistical Computing*. Vienna, Austria, 2014. URL: `http://www.r-project.org`.

[47] H Fisher Raymond et al. "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." In: *Sexually transmitted infections* 89 (2013), pp. 383–7. ISSN: 1472-3263. DOI: `10.1136/sextrans-2012-050675`.

[48] Louis-Paul Rivest and Sophie Baillargeon. *Rcapture: Loglinear Models for Capture-Recapture Experiments. R package version 1.4-2*. 2014.

[49] Abby E Rudolph, Crystal M Fuller, and Carl Latkin. "The importance of measuring and accounting for potential biases in respondent-driven samples." In: *AIDS and behavior* 17.6 (July 2013), pp. 2244–52. ISSN: 1573-3254. DOI: `10.1007/s10461-013-0451-y`.

[50] Matthew J Salganik et al. "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." In: *American journal of epidemiology* 174.10 (Nov. 2011), pp. 1190–6. ISSN: 1476-6256. DOI: `10.1093/aje/kwr246`.

[51] San Francisco HIV Epidemiology Section. "HIV / AIDS Epidemiology Annual Report HIV / AIDS Epidemiology Annual Report". In: (2010).

[52] Bernhard Schwartlander et al. "Towards an improved investment approach for an effective response to HIV/AIDS." In: *Lancet* 377.9782 (June 2011), pp. 2031–41. ISSN: 1474-547X.

[53] Hyman M Scott et al. "Racial/ethnic and sexual behavior disparities in rates of sexually transmitted infections, San Francisco, 1999-2008." In: *BMC public health* 10 (Jan. 2010), p. 315. ISSN: 1471-2458. DOI: 10.1186/1471-2458-10-315.

[54] StataCorp. *Stata Statistical Software*. College Station, TX.

[55] May Sudhinaraset, H Fisher Raymond, and Willi McFarland. "Convergence of HIV prevalence and inter-racial sexual mixing among men who have sex with men, San Francisco, 2004-2011." In: *AIDS and behavior* 17.4 (May 2013), pp. 1550–6. ISSN: 1573-3254. DOI: 10.1007/s10461-012-0370-3.

[56] Seymour Sudman, Monroe G Sirken, and Charles D Cowan. "Sampling Rare and Elusive Populations". In: *Science* 240.4855 (1988), pp. 991–996.

[57] K Tilling, J a Sterne, and C D Wolfe. "Estimation of the incidence of stroke using a capture-recapture model including covariates." In: *International journal of epidemiology* 30.6 (Dec. 2001), 1351–9, discussion 1359–60. ISSN: 0300-5771.

[58] Lucia V Torian et al. "Department of Health Sexually Transmitted Disease Clinics , A Decade of Serosurveillance Finds that Racial Disparities and Associations Between HIV and Gonorrhea Persist". In: *Sexually Transmitted Diseases* 29.2 (2002), pp. 73–78.

[59] UNAIDS. *Guidelines on Estimating the Size of Populations Most at Risk to HIV*. Tech. rep. Geneva, Switzerland, 2010, p. 51.

[60] Joint United Nations Programme on HIV/AIDS (UNAIDS). *GLOBAL REPORT: UNAIDS report on the global AIDS epidemic 2013*. Tech. rep. Geneva, Switzerland, 2013.

[61] UNAIDS/WHO Working Group on Global HIV/AIDS and STI Surveillance. *Guidelines on Estimating the Size of Populations Most at Risk to HIV*. Tech. rep. Geneva, Switzerland, 2011.

[62] Ashton M. Verdery et al. "Respondent-driven Sampling Estimators Under Real and Theoretical Recruitment Conditions of Female Sex Workers in China". In: *Epidemiology* 26.5 (2015), pp. 661–665. ISSN: 1044-3983. DOI: 10.1097/EDE.0000000000000335. URL: http://content.wkhealth.com/linkback/openurl?sid=WKPTLP:landingpage%7B%5C&%7Dan=00001648-201509000-00006.

[63] Giuseppe Verlato and Michele Muggeo. "Capture-Recapture Method in the Epidemiology of Type 2 Diabetes". In: *Diabetes Care* 23.6 (2000), pp. 759–764.

[64] Paul Wesson et al. "If You Are Not Counted, You Don't Count: Estimating the Number of African-American Men Who Have Sex with Men in San Francisco Using a Novel Bayesian Approach". In: *Journal of Urban Health* (2015). ISSN: 1099-3460. DOI: 10.1007/s11524-015-9981-0.

[65] Mike West. "Inference in successive sampling discovery models". In: *Journal of Econometrics* 75.1 (Nov. 1996), pp. 217–238. ISSN: 03044076. DOI: 10.1016/0304-4076(95)01777-1.

[66] Guangyong Zou. "A Modified Poisson Regression Approach to Prospective Studies with Binary Data". In: 159.7 (2004), pp. 702–706. DOI: 10.1093/aje/kwh090.

Appendix A

Appendix to Chapter 2

Scale	Author	Study Location	Population	PSE Method	Population Size Estimate	Lower-Upper bounds
0-5,000	Johnstson 2013	Mauritius	Female Sex Workers	Service Multiplier	424	241-610
				Service Multiplier 2	254	141-370
				Service Multiplier 3	945	745-1148
				Service Multiplier 4	910	702-1119
	Johnston 2013	Chiang Mai, China	People Who Inject Drugs	Service Multiplier	600	40-1161
				Service Multiplier 2	1500	59-2942
	Kimber 2008	Sydney, Australia	People Who Inject Drugs	Capture-recapture	2671	–
				Capture-recapture-TP	1288	–
				Capture-recapture-Open	877	–
				Service Multiplier	1146	–
				Service Multiplier 2	600	–
				Service Multiplier 3	675	–

Scale	Author	Study Location	Population	PSE Method	Population Size Estimate	Lower-Upper bounds
				Service Multiplier 4	750	–
				Service Multiplier 5	1000	–
	Luan 2005	Chengdu, China	Men who have Sex with Men	Capture-recapture	1408	1116-1908
				Capture-recapture 2	1207	932-1712
				Capture-recapture 3	949	757-1272
				Service Multiplier	877	–
	Mutagoma 2014	Rwanda	Female Sex Workers	Capture-recapture	3205	2998-3412
				Mapping & Enumeration	3348	–
				Service Multiplier	2253	1916-2524
	Sawitri 2012	Bali	People Who Inject Drugs	Service Multiplier	769	–
				Service Multiplier 2	723	–
				Service Multiplier 3	700	–

Scale	Author	Study Location	Population	PSE Method	Population Size Estimate	Lower-Upper bounds
				Nominator	529	–
	Smit 2002	Utrecht, Netherlands	Homeless Population	Capture-recapture (Chao)	261	213-356
				Capture-recapture (Zelterman)	345	278-455
	Xu 2014	Greater Victoria, British Columbia, Canada	People Who Inject Drugs	Capture-recapture	3329	2246-5078
				Capture-recapture (Huggin)	3342	2254-5098
				Capture-recapture (Pledger)	3330	2246-5078
	Khalid 2014	Unguja Island, Zanzibar	People Who Inject Drugs	Unique Object Multiplier	3381	2623-4138
				Capture-recapture	2819	2122-3516
				Literature Review	4000	–
				Delphi	3000	200-5000
				Panel of Experts	3000	–

Scale	Author	Study Location	Population	PSE Method	Population Size Estimate	Lower-Upper bounds
				Wisdom of the Crowds	475	15-1500000
	Vadivoo 2008	Dimapur, India	Female Sex Workers	Programme	2400	–
				Unique Object Multiplier	2683	1833-3492
				Programme	1000	–
	Vadivoo 2008	Kolhapur, India	Female Sex Workers	Unique Object Multiplier	313	183-617
				Reverse Tracking Method	248	–
				Service Multiplier	645	–
	Vadivoo 2008	Parbhani, India	Female Sex Workers	Programme	2100	–
				Unique Object Multiplier	2795	1758-4739
				Programme	3100	–
	Vadivoo 2008	Prakasam, India	Female Sex Workers	Unique Object Multiplier	1055	912-1247
				Reverse Tracking Method	1757	1749-1765

Scale	Author	Study Location	Population	PSE Method	Population Size Estimate	Lower-Upper bounds
	Vadivoo 2008	Pune, India	Female Sex Workers (street-based)	Programme	1400	–
				Unique Object Multiplier	2263	1745-2942
				Reverse Tracking Method	343	–
				Service Multiplier	1240	1745-2942
	Vadivoo 2008	Pune, India	Female Sex Workers (brothel-based)	Programme	3248	–
				Unique Object Multiplier	579	266-1341
				Reverse Tracking Method	1455	1448-1463
	Vadivoo 2008	Visakhapatnam, India	Female Sex Workers	Programme	4200	–
				Unique Object Multiplier	1149	963-1412
				Reverse Tracking Method	2616	2586-2646
				Service Multiplier	161	153-173
	Vadivoo 2008	Warangal, India	Female Sex Workers	Programme	3500	–
				Reverse Tracking Method	1010	1004-1017

Scale	Author	Study Location	Population	PSE Method	Population Size Estimate	Lower-Upper bounds
	Vadivoo 2008	Yevatmal, India	Female Sex Workers	Programme Unique Object Multiplier Reverse Tracking Method Service Multiplier	1000 192 275 554	– 136-313 – –
	Vadivoo 2008	Churachandpur, India	People Who Inject Drugs	Programme Unique Object Multiplier	2500 1493	– 1031-2083
	Vadivoo 2008	Phek, India	People Who Inject Drugs	Programme Unique Object Multiplier	1800 2439	– 1754-3846
	Vadivoo 2008	Wokha, India	People Who Inject Drugs	Programme Unique Object Multiplier	4800 2941	– 1852-6250
	Vadivoo 2008	Coimbatore, India	Men who have Sex with Men	Programme Unique Object Multiplier Service Multiplier Reverse Tracking Method	1500 4673 2642 2849	– 3546-6289 2484-2857 2819-2878

Scale	Author	Study Location	Population	PSE Method	Population Size Estimate	Lower-Upper bounds
	Vadivoo 2008	East Godavari, India	Men who have Sex with Men	Programme	1800	–
				Unique Object Multiplier	1196	862-1712
				Service Multiplier	2577	2086-3262
				Reverse Tracking Method	1141	1126-1155
	Vadivoo 2008	Madurai, India	Men who have Sex with Men	Programme	1400	–
				Unique Object Multiplier	3676	2747-5102
				Service Multiplier	1752	1527-2098
				Reverse Tracking Method	937	919-956
	Vadivoo 2008	Pune, India	Men who have Sex with Men	Programme	3200	–
				Unique Object Multiplier	334	292-379
				Service Multiplier	335	–
				Reverse Tracking Method	316	–

Scale	Author	Study Location	Population	PSE Method	Population Size Estimate	Lower-Upper bounds
0-10,000	Vadivoo 2008	Salem, India	Men who have Sex with Men	Programme	2500	–
				Unique Object Multiplier	3401	2445-4975
				Service Multiplier	2033	1745-2499
				Reverse Tracking Method	1416	1385-1447
	Vadivoo 2008	Visakhapanam, India	Men who have Sex with Men	Programme	1600	–
				Unique Object Multiplier	746	619-919
				Service Multiplier	910	721-1168
				Reverse Tracking Method	950	940-959
	Cruyff 2008	Rotterdam, The Netherlands	Opiate users	Capture-recapture (Poisson, no covariates)	2937	2834-3040
				Capture-recapture (Poisson, with covariates)	2992	2879-3105

Scale	Author	Study Location	Population	PSE Method	Population Size Estimate	Lower-Upper bounds
				Capture-recapture (NB, no covariates)	5213	4056-6470
				Capture-recapture (NB, with covariates)	5001	4601-5402
				Unique Object Multiplier	2251	1462-3030
	Khalid 2014	Unguja Island, Zanzibar	Female Sex Workers	Service Multiplier	5000	3924-6076
				Capture-recapture	4622	3625-5608
				Literature Review	1225	980-10528
				Delphi	900	200-4000
				Panel of Experts	3958	–
				Wisdom of the crowds	200	5-500000
				Unique Object Multiplier	3811	2659-5319
	Khalid 2014	Unguja Island, Zanzibar	Men who have Sex with Men	Service Multiplier	2157	1528-2785
				Capture-recapture	6160	5268-7052

Scale	Author	Study Location	Population	PSE Method	Population Size Estimate	Lower-Upper bounds
				Literature Review	2775	2313-9250
				Delphi	850	370-6000
				Panel of Experts	2157	–
				Wisdom of the crowds	1200	10-70000
	Livak 2013	South Side of Chicago, USA	Young Black Men who have Sex with Men	Service Multiplier	6340	–
				Survey	2286	1241-3379
				Delphi	8109	–
				Triangulated Estimate	5578	–
	Vadivoo 2008	Chitoor, India	Female Sex Workers	Programme	7900	–
				Unique Object Multiplier	871	773-1004
				Reverse Tracking Method	2436	2424-2448
	Vadivoo 2008	Coimbatore, India	Female Sex Workers	Service Multiplier	4300	4098-4592
				Programme	1600	–
				Unique Object Multiplier	7692	5714-10526
				Reverse Tracking Method	4221	4174-4268

Scale	Author	Study Location	Population	PSE Method	Population Size Estimate	Lower-Upper bounds
				Service Multiplier	3462	3101-3942
				Programme	4800	–
	Vadivoo 2008	Dharmapuri, India	Female Sex Workers	Unique Object Multiplier	1563	1389-1812
				Reverse Tracking Method	1841	1806-1876
				Service Multiplier	8711	8061-9501
				Programme	7200	–
	Vadivoo 2008	East Godavari, India	Female Sex Workers	Unique Object Multiplier	1462	1147-1931
				Reverse Tracking Method	4517	4466-4569
				Programme	10400	–
	Vadivoo 2008	Guntur, India	Female Sex Workers	Unique Object Multiplier	1176	1006-1404
				Reverse Tracking Method	3726	3704-3748
				Service Multiplier	1179	1122-1261

Scale	Author	Study Location	Population	PSE Method	Population Size Estimate	Lower-Upper bounds
	Vadivoo 2008	Hyderbad, India	Female Sex Workers	Programme	7500	–
				Reverse Tracking Method	3334	3313-3356
	Vadivoo 2008	Karimnagar, India	Female Sex Workers	Programme	2500	–
				Reverse Tracking Method	9749	9640-9857
	Vadivoo 2008	Salem, India	Female Sex Workers	Programme	5100	–
				Unique Object Multiplier	6757	4310-11111
				Reverse Tracking Method	2421	2383-2458
				Service Multiplier	4750	4312-5390
	Vadivoo 2008	Chennai, India	Men who have Sex with Men	Programme	9100	–
				Unique Object Multiplier	5517	4040-7619
				Reverse Tracking Method	4955	4889-5002
				Service Multiplier	5085	4587-5479
	Vadivoo 2008	Hyderabad, India	Men who have Sex with Men	Programme	5000	–
				Unique Object Multiplier	4902	2315-10870

Scale	Author	Study Location	Population	PSE Method	Population Size Estimate	Lower-Upper bounds
				Reverse Tracking Method	4076	3951-4201
0-20,000	Archibald 2001	Montreal, Canada	People Who Inject Drugs	Survey Delphi	4300	800-20800
					12500	10000-15000
				Capture-recapture	11680	8640-16460
	Kuhnert 2009	Bangkok, Thailand	Illicit Drug Users	Capture-recapture (CAMCR)	18367	–
				Capture-recapture (Zelterman)	12796	–
				Capture-recapture (Zelterman 2)	10312	–
				Capture-recapture (Chao)	10743	–

Scale	Author	Study Location	Population	PSE Method	Population Size Estimate	Lower-Upper bounds
	Okal 2013	Nairobi, Kenya	People Who Inject Drugs	Service Multiplier	22727	–
				Service Multiplier 2	13251	–
				Service Multiplier 3	5031	–
				Service Multiplier 4	5652	–
				Wisdom of the crowds	3000	–
				Literature Review	6562	–
				Delphi	6107	–
	Okal 2013	Nairobi, Kenya	Men who have Sex with Men	Service Multiplier	10417	–
				Service Multiplier 2	3261	–
				Service Multiplier 3	22222	–
				Service Multiplier 4	11667	–
				Wisdom of the crowds	10000	–
				Literature Review	12265	–

Scale	Author	Study Location	Population	PSE Method	Population Size Estimate	Lower-Upper bounds
	Quaye 2015	Cape Coast, Ghana	Men who have Sex with Men	Delphi	11042	–
				Literature Review	2058	–
				Mapping & Enumeration	1095	–
				Service Multiplier	4897	3455-6342
				Unique Object Multiplier	2460	2106-3853
				Wisdom of the crowds	300	5-20000
	Quaye 2015	Koforidua, Ghana	Men who have Sex with Men	Literature Review	587	–
				Mapping & Enumeration	1132	–
				Service Multiplier	8848	1260-16437
				Unique Object Multiplier	2885	687-5084
				Wisdom of the crowds	200	5-10000

Scale	Author	Study Location	Population	PSE Method	Population Size Estimate	Lower-Upper bounds
	van der Heijden 2013	The Netherlands	Problem Drug Users	Capture-recapture (TP)	6695	5779-7611
				Capture-recapture (Zelterman)	10415	8400-12429
				Service Multiplier	17700	–
	Guo 2013	Chongqing, China	People Who Inject Drugs	Network scale-up	7488	6523-8452
				Network scale-up (Respect factor adjusted)	14975	13047-16904
	Guo 2013	Chongqing, China	Men who have Sex with Men	Network scale-up	7452	6490-8414
				Network scale-up (Respect factor adjusted)	16767	14602-18932
	Vadivoo 2008	Chennai, India	Female Sex Workers	Programme	17400	–
				Unique Object Multiplier	3810	2909-5128
				Reverse Tracking Method	10634	10485-10783
				Service Multiplier	11809	9104-15832

Scale	Author	Study Location	Population	PSE Method	Population Size Estimate	Lower-Upper bounds
	Vadivoo 2008	Madurai, India	Female Sex Workers	Programme	15400	–
				Unique Object Multiplier	3650	2817-4878
				Reverse Tracking Method	6519	6395-6643
				Service Multiplier	7668	6907-8816
	Vadivoo 2008	Mumbai, India	Female Sex Workers (street-based)	Programme	4000	–
				Unique Object Multiplier	3879	2286-6737
				Reverse Tracking Method	2622	2612-2630
				Service Multiplier	11910	8293-17600
	Vadivoo 2008	Thane, India	Female Sex Workers (brothel-based)	Programme	5300	–
				Unique Object Multiplier	435	371-516
				Reverse Tracking Method	1244	1242-1246
				Service Multiplier	11344	8955-14553

Scale	Author	Study Location	Population	PSE Method	Population Size Estimate	Lower-Upper bounds
0-30,000	Vadivoo 2008	Thane, India	Female Sex Workers (street-based)	Programme	3200	–
				Unique Object Multiplier	789	625-1019
				Reverse Tracking Method	945	932-958
				Service Multiplier	16174	11330-23495
	Vadivoo 2008	Guntur, India	Men who have Sex with Men	Programme	2800	–
				Unique Object Multiplier	4167	2049-8621
				Reverse Tracking Method	685	681-688
				Service Multiplier	10087	4784-21091
	Vadivoo 2008	Mumbai Thane, India	Men who have Sex with Men	Programme	20800	–
				Unique Object Multiplier	870	610-1266
				Reverse Tracking Method	1182	1164-1200
	Archibald 2001	Vancouver, Canada	People Who Inject Drugs	Service Multiplier	17500	13900-23100
				Survey	6400	1100-28800

Scale	Author	Study Location	Population	PSE Method	Population Size Estimate	Lower-Upper bounds
				Delphi	10500	8000-13000
				Capture-recapture	11670	10300-13420
	Archibald 2001	Vancouver, Canada	Men who have Sex with Men	Service Multiplier	15900	11500-22100
				Survey	7000	1000-24200
				Delphi	15000	10000-20000
				Survey 2	26500	–
	Johnston 2011	Mauritius	People Who Inject Drugs	Unique Object Multiplier	6329	4566-9772
				Service Multiplier	10444	6266-23500
				Service Multiplier 2	5699	4301-8009
				Service Multiplier 3	10315	7153-17500
				Service Multiplier 4	7182	5679-9614
				Service Multiplier 5	8900	7974-9980

Scale	Author	Study Location	Population	PSE Method	Population Size Estimate	Lower-Upper bounds
	Quaye 2015	Accra, Ghana	Men who have Sex with Men	Literature Review	13966	–
				Mapping & Enumeration	975	–
				Service Multiplier	20822	15449-26198
				Unique Object Multiplier	4187	2878-5499
				Wisdom of the crowds	500	4-30000
	Quaye 2015	Kumasi, Ghana	Men who have Sex with Men	Literature Review	5024	–
				Mapping & Enumeration	785	–
				Service Multiplier	5484	3409-7502
				Unique Object Multiplier	2994	1772-4219
				Wisdom of the crowds	500	10-30000
	Vadivoo 2008	Mumbai, India	Female Sex Workers (brothel-based)	Programme	6900	–
				Unique Object Multiplier	4152	2585-6850
				Reverse Tracking Method	7167	7016-7318

Scale	Author	Study Location	Population	PSE Method	Population Size Estimate	Lower-Upper bounds
	Johnston 2005	San Francisco, CA	People Who Inject Drugs	Service plier Multi-	17358	12718-24009
				Service plier Multi-	23779	16027-31534
				Service plier 2 Multi-	10158	3273-17044
				Service plier 3 Multi-	20909	14960-26861
				Service plier 4 Multi-	10130	2998-17263
	Archibald 2001	Toronto, Canada	People Who Inject Drugs	Service plier Multi-	17700	13275-23900
				Survey	12300	3600-32000
				Delphi	12500	10000-15000
				Capture-recapture	13360	10460-17880
0-50,000	Archibald 2001	Montreal, Canada	Men who have Sex with Men	Survey Delphi	18500 40000	4900-46500 30000-50000
				Survey 2	37000	–

Scale	Author	Study Location	Population	PSE Method	Population Size Estimate	Lower-Upper bounds
	Archibald 2001	Toronto, Canada	Men who have Sex with Men	Service Multiplier	39100	28700-53000
				Survey Delphi	18800 35000	5300-45000 25000-45000
				Survey 2	35000	–
	Hope 2005	London, United Kingdom	Crack Cocaine users	Capture-recapture	20972	13093-42960
				Ratio Estimation	23000	–
	Vadivoo 2008	Mumbai, India	Female Sex Workers	Programme	32600	–
				Unique Object Multiplier	6250	3030-16667
	Guo 2013	Chongqing, China	Drug Users	Network scale-up	18979	17444-20515
				Network scale-up (Respect factor adjusted)	37959	34888-41030
	Guo 2013	Chongqing, China	Female Sex Workers	Network scale-up	31576	29595-33556

Scale	Author	Study Location	Population	PSE Method	Population Size Estimate	Lower-Upper bounds
				Network scale-up (Respect factor adjusted)	28418	26636-30201
0-50,000+	Raymond 2013	San Francisco, CA, USA (2010)	Men who have Sex with Men	Service Multiplier	40944	–
				Service Multiplier 2	67633	–
				Service Multiplier 3	66900	–
				Service Multiplier 4	37692	–
				Service Multiplier 5	110682	–
				Service Multiplier 6	51000	–
				Survey	77999	–
				Survey 2	60981	–
				Survey 3	74753	–
				Lieb's ratio method	60804	–
				Lieb's ratio method 2	69889	–

Scale	Author	Study Location	Population	PSE Method	Population Size Estimate	Lower-Upper bounds
	Johnston 2013	San Francisco, CA, USA (2009)	People Who Inject Drugs	Service Multiplier	45315	24576-66057
				Service Multiplier 2	5200	1003-9398
				Service Multiplier 3	18250	4675-31826
				Service Multiplier 4	12857	–
				Service Multiplier 5	81500	658-162343
				Service Multiplier 6	7000	2945-11056
	Guo 2013	Chongqing, China	Clients of Female Sex Workers	Network scale-up	73513	70491-76535
				Network scale-up (Respect factor adjusted)	163199	156490-169908
	Chen 2013	Changsha, China	Men who have Sex with Men	Modified LMS	65657	–
				Capture-recapture	2636	–
	Okal 2013	Nairobi, Kenya	Female Sex Workers	Service Multiplier	54467	–
				Service Multiplier 2	16550	–

Scale	Author	Study Location	Population	PSE Method	Population Size Estimate	Lower-Upper bounds
				Service Multiplier 3	649000	–
				Wisdom of the crowds	10000	–
				Literature Review	29494	–
				Delphi	29494	–
Population Proportion	Jing 2014	Taiyuan, China	Men who have Sex with Men	Network scale-up	0.149	0.131-0.167
				Generalized Network scale-up	2.927	2.566-3.288
				Median of Multipliers	0.486	–
	Salganik 2011	Curitiba, Brazil	Heavy Drug Users	Survey (2004)	0.3	0-0.7
				Survey (2010)	0.6	0-1.6
				Service Multiplier	0.6	0.3-3.2
				Network scale-up	3.3	2.7-4.1
				Generalized Network scale-up	6.3	4.5-8

Appendix B

Appendix to Chapter 4

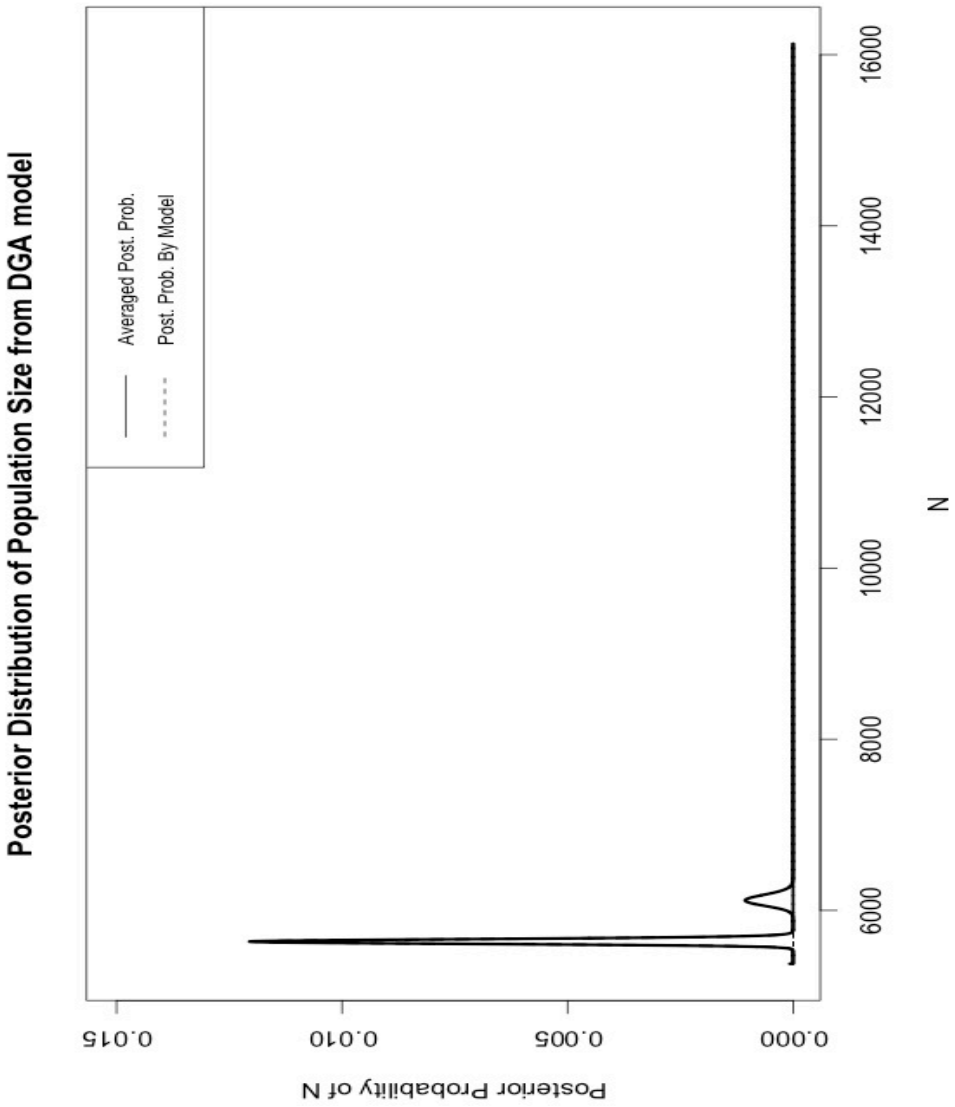


Figure B.1: Posterior probability distribution of the size of the diagnosed PLWHA population under Alameda County, CA, public health jurisdiction in 2013 from DGA model, 2013

Model	$\hat{N}$	95% CI	AIC	df	%Complete (laboratory)
Independence	45,278	35,842-58,420	51.62	3	11.0%
Base + S2*S3	52,298	38,402-73,994	51.53	2	9.5%
DGA	3,208	2,830-6,370	–	–	155%

Table B.1: Estimate of unobserved population size, total population size, and completeness of laboratory surveillance list from the best-fitting log-linear regression modeling and DGA model (three-source capture-recapture analysis), Alameda County, CA, 2013

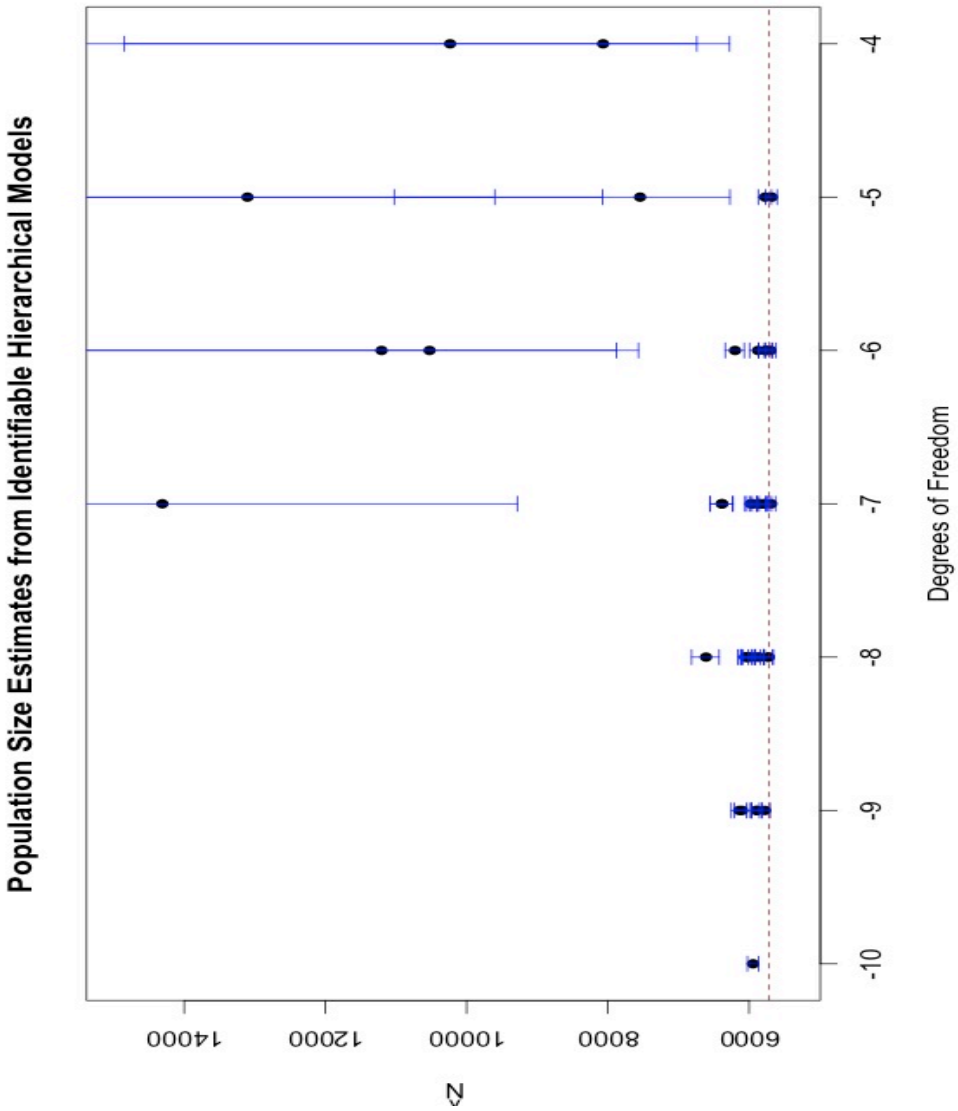


Figure B.2: Population Size Estimates of the diagnosed PLWHA population under Alameda County, CA, public health jurisdiction from identifiable hierarchical models, 2013[1]

[1]Due to plot scaling, two model estimates were excluded from this plot (including the best fitting model with three degrees of freedom remaining). Horizontal dotted line corresponds to estimated from DGA model (5,720).

Model	$\hat{N}$	95% CI	AIC	df[2]
Base	5,943	5,867-6,023	974.5	10
Base + S1*S3	5,891	5,820-5,965	715.1	9
Base + S1*S3 + S2*S3	5,846	5,780-5,915	485.8	8
Base + S1*S3 + S2*S3 + S3*S4	5,684	5,625-5,747	417.7	7
Base + S1*S3 + S1*S4 + S2*S3 + S2*S4	6,198	6,069-6,338	391.2	6
Base + S1*S3 + S1*S4 + S2*S3 + S2*S4 + S3*S4	7,543	6,270-11,026	387.4	5
Base + S1*S3 + S1*S4 + S3*S4 + S1*S3*S4 + S2*S3 + S2*S4	10,234	6,739-30,225	382.75	4
Base + S1*S3 + S1*S4 + S3*S4 + S1*S3*S4 + S2*S3 + S2*S4 + S3*S4 + S2*S3*S4	1.113e10	8,062-3.339e10	379.9	3

Table B.2: Population Size Estimates of the diagnosed PLWHA population under Alameda County, CA, public health jurisdiction from Identifiable Models, 2013

[2]Degrees of freedom remaining. Selected models are the best-fitting model (lowest AIC) within strata of models with equivalent degrees of freedom remaining.