From the holdings of the Department of Special Collections , Princeton University Library, Princeton, NJ

These documents are either in the public domain, or can only be used for educational and research purposes (Fair use) as per U.S. Copyright law (text below). By accessing this file, all users agree that their use falls in the public domain or within fair use as defined by the copyrightlaw. We do not charge any permission or usage fees for the publication of images of material inour collections, including those provided by Princeton University Library, via our website, catalog, or directly from staff.For more information regarding copyright, credit, and citation please review the information onthe Copyright, Credit and Citation Guidelines page of our website.
https://library.princeton.edu/special-collections/policies/copyright-credit-and-citation-guidelines. If you have any questions, including how to order higher-quality images from the collection, contact Special Collections Public Services via our Ask Us! Form. https://library.princeton.edu/special-collections/ask-us

---

Doctoral Dissertations, 2011-2026 Population Studies

2015

Network Reporting Methods

Feehan, Dennis Michael

Princeton, NJ : Princeton University

---

http://arks.princeton.edu/ark:/88435/dsp014m90dx89q

U.S. Copyright law: The copyright law of the United States (Title 17, §108, United States Code) governs the making of photocopies or other reproductions of copyrighted material. Under certain conditions specified in the law, libraries and archives are authorized to furnish a photocopy or other reproduction. One of these specified conditions is that the photocopy or reproduction is not to be used for any purpose other than private study, scholarship, or research. If a user makes a request for, or later uses, a photocopy or reproduction for purposes in excess of fair use that user may be liable for copyright infringement. This institution reserves the right to refuse to accept a copying order if, in its judgment, fulfillment of the order would involve a violation of copyright law. https://www.copyright.gov/title17/title17.pdf

Network reporting methods

Dennis Michael Feehan

A Dissertation
Presented to the Faculty
of Princeton University
in Candidacy for the Degree
of Doctor of Philosophy

Recommended for Acceptance
by the Program in
Population Studies
Adviser: Matthew J. Salganik

September 2015

© Copyright by Dennis Michael Feehan, 2015.

All rights reserved.

Abstract

This dissertation is about how the science of sampling and survey research can be generalized beyond the individual: while traditional surveys ask respondents to report about themselves, we study *network reporting* surveys in which respondents are asked to report about others. Despite many technical challenges, understanding how to design and analyze network reporting studies is worth the effort: network reporting surveys can be used to study many important rare and hidden populations for which traditional survey methods are inadequate.

We begin by introducing the network reporting framework, a general toolkit that can help researchers develop new estimators and new data collection methods. The framework can also help researchers better understand many existing estimators. We apply the framework to analyze network scale-up, an existing method for studying epidemiologically important hidden populations like sex workers and drug injectors. We derive the precise conditions required for the basic scale-up estimator to have desirable statistical properties; we also introduce a new, *generalized scale-up* estimator, which we expect to outperform the existing estimator in many settings.

Next, we turn to a key question for all network reporting studies: which personal network should respondents be asked to report about? We conjecture that there may be a trade-off between the quantity and the quality of information obtained from different personal networks. We test this conjecture by embedding an experiment in a large, nationally-representative household survey that we conducted in Rwanda. Our results show that there may indeed be a trade-off, but future work is required to understand this possibility in more detail.

Finally, we apply the network reporting framework to a critical, unsolved problem in demography: estimating adult death rates in countries that lack complete vital registration systems. We introduce a new estimator and data collection procedure called *network survival*, and we test our new approach using the Rwanda survey.

Our results demonstrate that a network reporting study is feasible in an environment where adult mortality estimates are sorely needed, and the estimated network survival rates have plausible levels and age-patterns. However, more work is required before we can fully assess the accuracy of this new approach.

Acknowledgements

I could not have finished this dissertation without the help and support of many, many people.

Matt Salganik was a dedicated and inspiring adviser, as well as a coauthor and friend. This project certainly would likely never have happened if I had not TAed for him several years ago, in my first year of graduate school. That experience led to our work on the scale-up method, the Rwanda project, and beyond.

My committee, Noreen Goldman, Scott Lynch, and Doug Massey read the work and gave me many helpful comments; Noreen Goldman, especially, took the time to read and comment on multiple drafts of each chapter, and even showed up in the audience for one of my job talks.

For mentoring me through my general exams, and for general support and encouragment as I was finding my way in graduate school, I am grateful to Tom Espenshade, Georges Reniers, and Simon Levin. I would also like to thank Jessica Metcalf, who kindly agreed to serve on my examination committee despite my mid-summer defense date.

I also owe a debt of gratitude to the wonderful staff of the Office of Population Research. The administrative, computing, and library support team at OPR are the best in the business: thanks go especially to Wayne Appleton, Elana Broch, Nancy Cannuli, Chang Chung, Jennifer Curatola, Mary Lou Delaney, Joanne Donatiello, Jennifer Flath, Lynne Johnson, Joyce Lopuh, Judie Miller, and Robin Pispecky.

I learned a tremendous amount from my collaborators on the Rwanda survey; in addition the Matt Salganik, they include Mary Mahy, Wolfgang Hladik, Aline Umubyeyi, and Bernard Barrere. I am grateful that we had the opportunity to work together, and hope we can collaborate again in the future. I am also grateful to each of the 4,669 people in Rwanda who took the time to respond to the survey, and to the team of interviewers and supervisors who made the data collection happen.

More personally, I am incredibly fortunate to have been surrounded by the fantastic group of students at OPR. I will be forever indebted to Laura and to Piali, who were enormous sources of emotional and intellectual support and inspiration. For the students in or approximately in my cohort: your friendship, intellectual insights, and tireless sense of fun kept me going through the marathon that this degree turned into, and I am looking forward to staying in touch with you for the rest of my life. You are too numerous to mention individually, but you know who you are. Thanks!

Finally, and most importantly, none of this would ever have happened without the love and support of my family, especially Mom, Dad, Noah, and Harlo.

To my parents.

Contents

Abstract iii

Acknowledgements v

List of Tables xiii

List of Figures xv

1 Introduction **1**

2 Network reporting **8**

2.1 Network reporting in a nutshell 10

2.2 Network reporting in detail: Sampling and imperfect reporting 13

2.2.1 Aggregate visibility estimators 14

2.2.2 Reporting errors 16

2.2.3 Individual visibility estimators 19

2.3 Conclusion 20

3 Generalized network scale-up **22**

3.1 The generalized scale-up estimator 24

3.1.1 Estimating N_H from sampled data 24

3.2 Comparison between the generalized and basic scale-up estimators . . 30

3.3 Variance estimation 34

3.4 Recommendations for practice 36

3.4.1 Estimation with samples from F and H 38

3.4.2 Estimation with only a sample from F 39
3.5 Conclusion and next steps 41

4 Quantity vs quality: A survey experiment to improve the network scale-up method 45

4.1 Introduction 46
4.2 Methods 49
4.2.1 The network scale-up method 49
4.2.2 Tie definition 50
4.2.3 Survey experiment for evaluating tie definitions 52
4.2.4 Internal consistency checks for evaluating tie definitions 53
4.2.5 Linear blending for size estimates 53
4.2.6 Framework for sensitivity analysis 55
4.2.7 Data collection and processing 57
4.3 Results 58
4.3.1 Internal consistency checks 59
4.3.2 Size estimates for hidden populations and sensitivity analysis . 60
4.4 Discussion 62
4.4.1 Limitations 63
4.4.2 Conclusion 64

5 The network survival estimator: evidence from Rwanda 66

5.1 Introduction and overview 66
5.2 Background 69
5.2.1 The problem: estimating death rates 69
5.2.2 Sibling survival 70
5.3 The network survival method 73
5.3.1 Estimating the amount of exposure, E_α 75

5.3.2 Estimating the number of deaths, O_α 75
5.3.3 Estimating respondents' network sizes 80
5.3.4 Adjustment factors for $\widehat{O}_\alpha$ 83
5.3.5 Putting it all together to estimate death rates, $\widehat{M}_\alpha$ 84
5.4 Results 85
5.5 Comparison to other estimates 93
5.5.1 Sibling survival from the 2010 RDHS 93
5.5.2 Estimates from international organizations 96
5.6 Discussion 97

6 Conclusion **103**

A Notation **107**

B Network reporting appendices **113**
B.1 Estimation with false positive reports 113
B.2 Estimates with a sample from F 115
B.2.1 Requirements for sampling designs from F 115
B.2.2 Estimating the total number of out-reports, $y_{F,H}$ 115
B.2.3 Reporting about multisets 116
B.2.4 Network degree and the known population method for estimating $\bar{d}_{F,F}$, $\bar{d}_{F,U}$, and $\bar{d}_{U,F}$ 117
B.2.5 Estimating the frame ratio, ϕ_F 122
B.3 Estimates with samples from F and H 124
B.3.1 Requirements for sampling designs from H 124
B.3.2 Data collection 125
B.3.3 Estimation using aggregated relational data from the hidden population 126
B.3.4 Estimating the average visibility, $\bar{v}_{H,F}$ 128

B.3.5 Guidance for choosing the probe alters for the game of contacts, $\mathcal{A}$ . . . 135
B.3.6 Term-by-term: δ_F and τ_F . . . 136
B.3.7 Estimating the size of the hidden population, N_H . . . 141
B.4 Sensitivity . . . 143
B.4.1 Generalized scale-up . . . 143
B.4.2 Basic scale-up . . . 144
B.5 Approximate unbiasedness of compound ratio estimators . . . 144
B.5.1 Overview . . . 144
B.5.2 The general case . . . 146
B.5.3 Applying Result B.5.1 to scale-up . . . 150
B.6 Variance estimation . . . 152
B.6.1 Variance estimation with a sample from F . . . 152
B.6.2 Variance estimation with sample from F and H . . . 157
B.7 Simulation study . . . 159
B.7.1 Overview . . . 159
B.7.2 Simulation design . . . 160
B.7.3 Results . . . 163

C Tie definition appendices **166**

C.1 Data collection and evaluation . . . 166
C.1.1 Groups of known size . . . 166
C.1.2 Balance checks . . . 168
C.1.3 Manipulation checks . . . 169
C.2 Tie strength and reporting accuracy: formal models . . . 169
C.2.1 Parametric model . . . 173
C.3 Comparisons with other estimates . . . 176
C.3.1 UNAIDS benchmark estimates . . . 176

C.3.2 Previous studies of key populations at risk for HIV in Rwanda 177
C.3.3 Comparison with UNAIDS benchmarks and previous Rwanda studies . . . 179
C.4 Linear blending . . . 181
C.5 Framework for sensitivity analysis . . . 183
C.5.1 Producing estimates under alternative assumptions . . . 185
C.6 Bootstrap procedure . . . 187

D Rwanda mortality appendices **191**
D.1 Estimating personal network size . . . 191
D.1.1 Probability samples . . . 191
D.1.2 Reporting about multisets . . . 192
D.1.3 Adapted known population estimator . . . 193
D.2 The network survival estimator . . . 195
D.2.1 Estimating the amount of exposure, E_α . . . 196
D.2.2 Estimating the number of deaths, O_α . . . 197
D.2.3 Combined estimator for M_α . . . 200
D.3 Robustness framework . . . 204
D.3.1 Inaccuracies in reports about deaths, y_{F,O_α} . . . 204
D.3.2 Decomposition framework . . . 206
D.4 Results for both sexes and tie definitions . . . 208
D.5 Sibling survival estimates . . . 208
D.5.1 Background . . . 208
D.5.2 Direct sibling survival estimates . . . 214
D.6 ${}_{45}q_{15}$ comparisons . . . 216
D.7 Data collection and network survival survey instrument . . . 218

Bibliography **225**

List of Tables

3.1 Network scale-up studies that have been completed. 25

3.2 Summary of conditions needed for basic and generalized scale-up estimators. 44

4.1 The two definitions of a personal network connection (also called a tie) that were used in the Rwanda survey. 51

4.2 The known populations used to estimate network sizes. 54

5.1 Estimated ${}_{45}q_{15}$ values, by tie definition and sex, along with summaries of the estimated sampling distribution for each estimator. 91

A.1 Notation used in this dissertation. 107

A.1 Notation used in this dissertation. 108

A.1 Notation used in this dissertation. 109

A.1 Notation used in this dissertation. 110

A.1 Notation used in this dissertation. 111

B.1 Responses collected during the game of contacts for each respondent i and each group A_j. 127

B.2 Summary of estimators' robustness to imperfect assumptions. 145

B.3 Description of the general form of the nonlinear estimators we propose. 147

B.4 Estimated coefficients of variation for the average degree from 3 different scale-up surveys. 152

B.5 Estimated coefficients of variation for quantities derived from a sample from the hidden population. 152
B.6 Estimated pairwise correlations for quantities derived from a sample from the hidden population. 153
B.7 Approximate relative bias in the estimates of the nonlinear quantities using data taken from the Curitiba study, the point estimates produced by the Curitiba study, and the estimated implied absolute bias. . . . 153

C.1 Comparison between households assigned to the acquaintance tie definition and households assigned to the meal tie definition. 169
C.2 UNAIDS benchmark ranges for the populations most at-risk for HIV/AIDS from UNAIDS (2014, pg. 22). 177
C.3 Population estimates used to produce UNAIDS benchmark values, and also to produce hidden population size estimates. 177
C.4 Benchmark estimates for hidden populations in Rwanda. 178
C.5 Question wording used to estimate the size of four key populations at risk for HIV in Rwanda. 180
C.6 Quantities needed for sensitivity analysis. 185
C.7 Possible assumptions a researcher might make about ϕ_F, δ_F, τ_F, and η_F for estimating the number of female sex workers from each arm of the study. 185

List of Figures

2.1 Illustration of the network reporting framework. 10

3.1 Assessing variance estimation procedures using scale-up studies in the United States, Rwanda, and Curitiba, Brazil 37

3.2 Recommended schematic of inputs and outputs for a study using the generalized scale-up estimator. 40

4.1 Hypothesized relationship between tie strength and error. 52

4.2 Estimates of network size distributions produced from the acquaintance tie definition and from the meal tie definition. 59

4.3 Internal consistency check for the 22 groups of known size in the Rwanda study. 60

4.4 Estimated hidden population size using each tie definition and the blended estimator. 61

4.5 Sensitivity of blended estimates for each hidden population. 62

4.6 Recipe for a survey experiment with blending. 65

5.1 How the network survival estimator converts the number of reports about deaths into an estimate for the number of deaths. 77

5.2 The components of network survival estimates for the death rate in a particular demographic group. 84

5.3 Estimating components of age-specific death rates for Rwandan Males for 12 months prior to our survey using responses from the meal tie definition. 88

5.4 Estimated age-specific death rates for Rwandan Males for 12 months prior to our survey using responses from the meal tie definition. . . . 89

5.5 Comparison between estimated conditional probabilities of death at adult ages ($_{45}q_{15}$) computed from death rate estimates for the two types of personal network investigated with our survey (Acquaintance and Meal), and direct sibling survival. 90

5.6 Comparison between network survival death rate estimates for the two types of personal network investigated with our survey, and direct sibling survival death rates estimates from the 2010 Rwanda Demographic and Health Survey. 92

5.7 Estimated age-specific death rates for Rwandan males using the meal definition under several different possible violations of the decedent network condition and the accurate reporting condition. 94

5.8 Average number of deaths reported from each interview in Rwanda using the acquaintance and meal tie definitions from the network survival study, and using the sibling history module of the 2010 Rwanda Demographic and Health Survey. 101

5.9 Estimated $_{45}q_{15}$ for Rwanda from six different sources. 102

A.1 Illustration of several important pieces of notation used throughout this thesis. 112

B.1 Example of a game board that could be used in the game of contacts interviewing procedure if the hidden population was people who inject drugs and the frame was made up of adults. 126

B.2 Data sources used for estimating adjustment factors. 137
B.3 Schematic of the bootstrap procedure to put a confidence interval around $\widehat{N}_H$ when there is a sample from the frame s_F. 153
B.4 Schematic of the bootstrap procedure to put a confidence interval around $\widehat{N}_H$ when there is a sample from the frame s_F and a sample from the hidden population s_H. 158
B.5 The mixing matrix used to generate a random network using the stochastic block model. 162
B.6 Estimated size of the simluated hidden population for the generalized and basic scale-up estimators. 164
B.7 Bias and predicted bias in the basic scale-up estimates for the same parameter configurations depicted in Figure B.6. 165

C.1 Average number of connections by tie definition for 22 groups of known size. 170
C.2 Comparison between our estimates, the UNAIDS benchmarks, and the estimates from other studies in Rwanda. 179
C.3 Illustration of the procedure used to estimate the sampling uncertainty in the blended estimates. 189
C.4 Three summaries of the errors in estimates from the internal consistency checks: mean squared error, average relative error, and mean absolute error. 190

D.1 Estimating components of age-specific death rates for Rwandan females for 12 months prior to our survey using responses from the meal tie definition. 209

D.2 Estimating components of age-specific death rates for Rwandan males for 12 months prior to our survey using responses from the acquaintance tie definition. 210
D.3 Estimating components of age-specific death rates for Rwandan females for 12 months prior to our survey using responses from the acquaintance tie definition. 211
D.4 Comparison between sibling estimates based on deaths reported 12 months and 60 months before the interview. 223
D.5 Comparison between ${}_{45}q_{15}$ approximated from estimated age-specific death rates under two different assumptions. 224

Chapter 1

Introduction

Surveys have traditionally been based on the idea that researchers can estimate characteristics of a population by obtaining a sample of individuals and asking them to report about themselves. For example, in order to estimate support for various candidates in an upcoming election, researchers might obtain a random sample of eligible voters and ask each one a question like "which candidate are you planning to vote for?" A large body of research in statistics and social science has produced extensive insights about how to randomly sample people, how to conduct interviews, and how to analyze the results. This research program has been very successful, and surveys are now a cornerstone of quantitative social science.

This dissertation is about how the science of sampling and survey research can be generalized beyond the individual: while traditional surveys ask respondents to report about themselves, we study surveys where respondents are asked to report about others. We call this approach *network reporting*, because it is based on the premise that survey respondents have useful information about people they are connected to through their personal networks. For example, a network reporting study might ask respondents "Please list the people who live in the same household as you" and then, for each household member, "Which candidate is this person planning to vote for?"

This example reveals that asking respondents to report about others introduces some technical challenges. First, which set of other people—that is, which personal network—should respondents be asked to report about? People are embedded in all sorts of different networks, and it is not obvious what trade-offs might be at play in deciding which network to use. In the example above, a respondent's personal network is defined by the people in her household, but perhaps we would be able to produce more accurate estimates by asking respondents to report about their close friends or family instead. Second, what if respondents' reports about others are inaccurate? For example, a respondent might omit a household member from her reports, or she might incorrectly report a household member's political views. Third, how should the structure of the network that respondents report about be taken into consideration? For example, if it is possible for more than one person from the same household to end up in the sample, then there is a risk of over-counting people in households with multiple survey respondents. Presumably, this risk of over-counting must be accounted for when researchers use network reports to estimate population quantities.

Despite these challenges, there are several important reasons why understanding how to design and analyze surveys that ask respondents to report about others is worth the effort. First, many critical scientific and policy problems require information about so-called hidden populations. Members of these hidden populations are difficult or impossible to sample and interview, making them ill-suited to being studied using traditional survey methods. These hidden populations can include people who have died, people who migrate, men who have sex with men, people who inject drugs, sex workers, and many other groups. Although it is not usually possible to study hidden populations using traditional survey methods, it may be possible to learn about members of a hidden population by asking survey respondents to report about others. For example, researchers who study mortality cannot sample and in-

terview people who have died; however, researchers can ask living people to report about acquaintances or family members who have died.

Second, many groups that are important for science and policy are relatively rare, meaning that they are small in proportion to the total size of the population. Unfortunately, the number of people researchers are able to interview—the sample size—is typically constrained by logistical factors such as time and money. These constraints mean that it may not be feasible to obtain a sample that is large enough to provide much information about rare populations using traditional survey methods. In this situation, asking respondents to report about members of their personal networks can have the advantage of increasing the amount of information obtained from each interview: for example, interviewing a respondent with 10 people in her personal network can, in principle, provide information about 10 people from just one interview.

Finally, network reporting studies enable researchers to study the networks themselves. People everywhere are embedded in a wide range of different personal networks. These networks are the result of the structure and patterns that govern every aspect of our lives: personal and familial relationships, interpersonal interactions, neighborhoods, professional affiliations and so forth. Many of these networks are of considerable scientific importance: their structure is related to the spread of diseases and ideas; people's access to critical resources like job opportunities and loans; people's physical and mental health; and so on. Unfortunately, many important social networks are very difficult to study because it is extremely rare or impossible to be able to completely observe them. Survey respondents' reports about their personal networks provide a sampled snapshot of these network connections that can be used to make inferences about certain global network properties. For example, respondents' network reports can be used to estimate the distribution of personal network sizes, which is an important part of network structure.

Throughout this dissertation, we adopt a design-based approach to making statistical inferences from sampled data. This approach has the advantage of freeing researchers from any need to make distributional assumptions. This freedom is appealing because very little is known about the structure of most types of social network, and so it is not typically possible to justify or even check model assumptions. Moreover, the design-based perspective often leads to estimators with a simple form, making it possible for us to analytically understand the sensitivity of the estimators to violations of the assumptions they rely upon. Of course, as we will see, the techniques we develop will still rely on various conditions being satisfied. But part of our contribution is to clarify exactly what these conditions are, and what happens to estimates when they are not met.

The remainder of this dissertation starts with Chapter 2, which introduces the network reporting framework. The network reporting framework provides researchers with a general toolkit for reasoning about survey respondents' network reports. The framework enables researchers to derive mathematical identities than can form the basis for developing new ways to collect and analyze network reports. Further, because the framework enables researchers to derive estimators from first principles, it is possible to understand precisely what assumptions must be made for estimators to have good statistical properties such as consistency and unbiasedness. The framework also enables researchers to develop methods for assessing the sensitivity of their network reporting estimates to the assumptions the estimators rely upon. Finally, the network reporting framework helps researchers understand and connect innovations in collecting and analyzing survey data that have emerged independently in several different fields, including anthropology, sociology, statistics, and demography.

The next two chapters, Chapter 3 and Chapter 4, study a technique called *network scale-up* which was originally developed by anthropologists to estimate the size of hidden populations (Bernard et al., 1989; Killworth et al., 1998b). Although the

scale-up method is appealing for many practical reasons, scale-up researchers worried about the simplicity of the model it was based upon, and about many potential sources of bias. It has not been clear how these biases relate to one another or how the biases and violations of the model assumptions would impact scale-up estimates. Chapter 3 employes the network reporting framework to analyze the scale-up method, leading to a more detailed and model-free understanding of the original estimator. This more detailed understanding includes a precise statement of the conditions that must hold for the estimator to have statistically desirable qualities such as unbiasedness and consistency. We also develop analytic expressions that reveal how failure to meet these required conditions will impact scale-up estimates. Further, we use the network reporting framework to derive an entirely new, generalized scale-up estimator. This new estimator takes advantage of an innovative data collection method that was introduced in Salganik et al. (2011b). We expect the generalized scale-up estimator to outperform the basic scale-up estimator in many settings, but our results enable researchers to compare the assumptions required by each method in order to make that decision for themselves.

Next, we turn to the main source of empirical results that the remainder of the dissertation is focused on: original data collected using a large, nationally-representative household survey from Rwanda. Chapter 4 uses the results of this survey to make two main contributions: one substantive and one methodological. Substantively, we use the scale-up method to estimate the size of four epidemiologically important hidden populations: female sex workers, clients of female sex workers, men who have sex with men, and people who inject drugs. These four hidden populations are critical to Rwanda's efforts to combat the HIV epidemic. Since these size estimates rely upon the assumptions of the basic scale-up method, we also use the framework for sensitivity analysis developed in Chapter 3 to provide a range of size estimates.

Methodologically, Chapter 4 uses the Rwanda scale-up survey to investigate a critical question that is relevant to all network reporting studies: which personal network should respondents be asked to report about? People everywhere are embedded in a wide array of different types of personal networks, ranging from close family and household members to distant acquaintances and professional contacts. In principle, respondents could be asked about the people they are connected to through any of these various networks. We conjecture that there is a trade-off between the quantity and the quality of information that can be obtained from respondents, and that this trade-off can be manipulated by varying the strength of the tie that respondents are asked to report about. In order to provide a first test of this conjecture, we embedded an experiment in the Rwanda survey: respondents were randomly assigned to report about one of two possible networks, one relatively strong and one relatively weak. Our results show that there may indeed be a trade-off at play, but future work is required to understand this possibility in more detail. Our study design can be used as a template that future scale-up studies can follow.

Chapter 5 applies the network reporting framework to a central problem of demography: estimating adult mortality in countries that lack adequate death registration. Death rates at adult ages are critical for many important scientific and policy questions; without adult mortality estimates, researchers cannot measure life expectancy, assess socioeconomic differences in health, or make the population projections required to plan for social welfare programs, school enrollment, labor force needs, and so on. Unfortunately, accurately estimating adult death rates in the world's poorest countries—including nearly all of Sub-Saharan Africa and much of Asia—has proven to be an extremely difficult problem. In Chapter 5, we develop a new data collection method and a new estimator for adult mortality rates based on respondents' reports about people in their networks who have died. We test this method on the Rwanda survey, enabling us to produce two sets of estimated adult death rates, one from

each arm of the survey experiment. We compare our estimates to the sibling survival estimate from a Demographic and Health Survey that was conducted in Rwanda at almost the same time. Our results are promising: they demonstrate that a network reporting study is feasible in an environment where adult mortality estimates are sorely needed, and the estimated network survival rates have plausible levels and age-patterns. However, more work is required before we can fully assess the accuracy of this new approach.

Finally, Chapter 6 concludes with a brief summary and a discussion of a few of the many directions that future research on network reporting could take.

Chapter 2

Network reporting

Network reporting is a general toolkit that researchers can use to help design and analyze surveys that ask respondents to report about people they are connected to in a wide range of different types of social network. The idea is to focus on the network of *reports*, rather than the underlying social network that was used to generate the reports. Under ideal conditions, when reporting is accurate, the structure of the network of reports will be closely related to the structure of the underlying social network. However, when reporting is inaccurate, the two networks may differ. The network reporting framework can then be used to help researchers understand the potential sources of error in their estimates. In some cases, this analysis may lead to improved estimators and new methods for collecting data. This analysis also enables researchers to develop analytical expressions that can be used to evaluate the impact different potential sources of error will have on substantive inferences. The chapters in the remainder of this dissertation illustrate both of these possibilities.

Researchers who design a network reporting study make two main decisions: first, researchers must decide which network respondents will be asked to report about (i.e., what will they tell respondents it means to "know" someone?). Second, researchers must decide how they will use respondents' network reports to estimate the quantity

of interest. The first decision—choosing a network tie definition—is the subject of Chapter 4. The second decision—choosing an estimator—is the subject of this chapter and the next.

Any method that tries to use network reports to estimate the size of a hidden population must contend with the problem of overcounting: if everyone's reported connections to drug injectors are simply added up, the resulting total will be larger than the total number of drug injectors because each drug injector can be reported more than once. Below, we will see that there are two main strategies for solving this problem. The first is to adjust for overcounting at the aggregate level, by dividing the total number of reports by the average number of times each drug injector is reported. And the second is to adjust for overcounting at the individual level, by weighting each report to account for the number of times that individual drug injector could have been reported.

The rest of the chapter begins by introducing the reporting graph and its close relative, the bipartite reporting graph. These two graphs are important conceptual tools that enable researchers to derive network reporting estimators. At first, our discussion will be intuitive and non-technical: we use a simple example to illustrate two identities, each of which leads to a different approach to estimating the size of a hidden population. Next, we introduce mathematical notation that will be used throughout the rest of this dissertation. Using this notation, we develop the tools necessary to account for features of real network reporting studies, like sampling and imperfect reporting. We conclude with a discussion of the potential directions that future work could take.

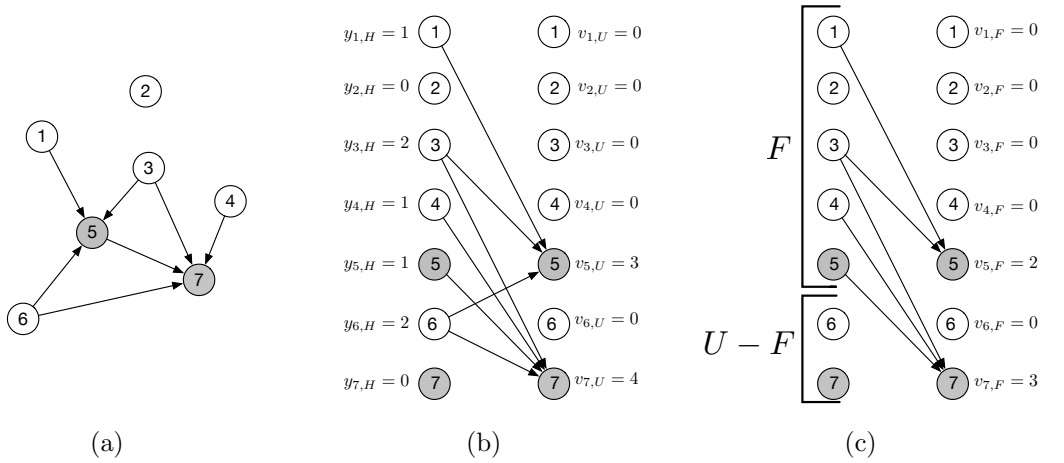


Figure 2.1: Illustration of the network reporting framework. Panel (a) shows a *reporting graph* consisting of a population of 7 people, 2 of whom are drug injectors (shown in grey). A directed edge $i \to j$ indicates that i counts j as a drug injector when answering the question "How many drug injectors do you know?" Panel (b) shows the same population, but redrawn as a *bipartite reporting graph* so that each person now appears twice: as a source of out-reports, on the left, and as a recipient of in-reports, on the right. This arrangement shows that total out-reports and total in-reports must be equal. Panel (c) shows the same population again, but now some of the people are in the frame population F and some are not. In real network reporting studies, we can only learn about out-reports from the frame population.

2.1 Network reporting in a nutshell

Consider a population of 7 people, 2 of whom are drug injectors (Figure 5.1(a)). Two people are connected by a directed edge $i \to j$ if person i would count person j as a drug injector when answering the question "How many drug injectors do you know?" Whenever $i \to j$, we say that i makes an *out-report* about j and that j receives an *in-report* from i. Figure 5.1(a) is called a *reporting graph*.

In this overview, we will assume that everyone in the population is interviewed, and that reporting is perfectly accurate; in the more technical sections below, we carefully extend these arguments to situations where respondents are sampled from a sampling frame and reporting is not perfect.

Each person in the reporting graph can be viewed as both a source of out-reports and a recipient of in-reports. In order to emphasize this point, Figure 5.1(b) shows the

population with each person represented twice: on the left as a sender of out-reports and on the right as a receiver of in-reports. This visual representation is called a *bipartite reporting graph*; any reporting network can be turned into a bipartite reporting graph, and vice-versa.

The bipartite reporting graph can be used to illustrate two simple but important identities; each of these two identities leads to a different estimator for hidden population size. First, the *aggregate reporting identity* says that

$$\text{total out-reports} = \text{total in-reports}. \tag{2.1}$$

Equation 2.1 is a consequence of the fact that each report leaves the left-hand side of the bipartite reporting graph and arrives on the right-hand side. Below, we will see that this aggregate identity can be manipulated so that it reads

$$\text{size of H} = \frac{\text{\# out-reports}}{\text{\# number of in-reports per member of H}}, \tag{2.2}$$

where H is the hidden population (e.g., drug injectors). The denominator—the number of in-reports per member of the hidden population —is called the *visibility* of H. Equation 2.2 shows that one strategy for estimating the size of the hidden population is to combine an estimator for the total number of out-reports, and an estimator for the number of in-reports per hidden population member.

Intuitively, if we simply used the total number of out-reports to estimate hidden population size, our estimate would be too large because members of the hidden population can be reported more than once. To account for this fact, Equation 2.2 says to divide the total number of out-reports by the average number of times each hidden population member gets reported (i.e., the visibility of H).

Second, the *individual reporting identity* says that, for each specific hidden population member j,

$$\text{total out-reports about } j = \text{total in-reports to } j. \tag{2.3}$$

This individual identity can be manipulated to develop a different estimator from the one above. Equation 2.3 can be rearranged to read

$$\frac{\text{total out-reports about } j}{\text{total in-reports to } j} = 1, \tag{2.4}$$

provided that the total in-reports and out-reports are not 0. Now, since the size of the hidden population is $\sum_j 1$ where j ranges over all hidden population members, this means that

$$\text{size of H} = \sum_j \frac{\text{total out-reports about } j}{\text{total in-reports to } j}. \tag{2.5}$$

Since the total reports about a hidden population member j is $\sum_{i \to j} 1$, where $i \to j$ ranges over all people i who report j, Equation 2.5 can also be written as

$$\text{size of H} = \sum_i \sum_{i \to j} \frac{1}{\text{total in-reports to } j}. \tag{2.6}$$

Therefore, an alternative strategy for estimating the size of the hidden population is to (i) weight each particular report about j, $i \to j$, by (1/# times j is reported); and, (ii) sum all these weighted reports up. The sum of these weighted reports will then be 1 for each particular hidden population member j and, over all j, the sum will be the size of the hidden population.

The intuition behind Equation 2.6 can be illustrated with an example: a hidden population member j who gets reported by 10 people will contribute 10 to the total

number of *reports* about hidden population members; however, this hidden population member j only adds 1 to the *size* of the hidden population. Therefore, each of the reports about j is divided by 10; j's total contribution to the sum in Equation 2.6 is then 1. In this sense, Equation 2.6 is based on adjusting for overcounting at the individual level, rather than the aggregate level: while Equation 2.2 suggests adding the reports up and then divide by the aggregate visibility, Equation 2.6 suggests dividing reports by the individual visibilities, and then adding them up.

These two identities allow the network reporting framework to connect existing research from a wide range of different disciplines. The aggregate network reporting identity is closely related to research from anthropology and sociology on the *network scale-up* method (Killworth et al., 1998b; Bernard et al., 2010); in fact, we use the aggregate reporting identity to propose a new network scale-up estimator at the end of this chapter. The individual network reporting identity leads directly to an approach called *multiplicity estimation* (Sirken, 1970; Thompson, 2002) and its modern relative *indirect sampling* (Lavallée, 2007). Network reporting provides a unified way to link together these historically distinct, but closely related approaches. Network reporting also enables researchers to analyze estimators that are based on survey respondents reporting about others, but that have not traditionally been thought of as network problems. For example, network reporting can help researchers better understand the sibling survival method (Chapter 5).

2.2 Network reporting in detail: Sampling and imperfect reporting

Having introduced the bipartite reporting graph and the two identities that it leads to, our goal is now to formalize the concepts developed in the previous section, and

then to incorporate features of real-world network reporting studies, such as sampling and imperfect reporting.

We will need some notation to describe different parts of the reporting graph. Let U be the entire population, and let $H \subset U$ be a hidden population of size N_H. Let $y_{i,H}$ be the total number of out-reports from person i (i.e., person i's answer to the question "How many drug injectors do you know?"). For example, Figure 5.1(b) shows that person 5 would report knowing 1 drug injector, so $y_{5,H} = 1$. Let $v_{i,U}$ be the total number of in-reports to i if everyone in U is interviewed; that is, $v_{i,U}$ is the *visibility* of person i to people in U. For example, Figure 5.1(b) shows person 5 would be reported as a drug injector by 3 people so $v_{5,U} = 3$.

2.2.1 Aggregate visibility estimators

Using this notation, we can express the aggregate network reporting identity from Equation 2.1 as

$$y_{U,H} = v_{U,U}, \tag{2.7}$$

where $y_{U,H} = \sum_{i \in U} y_{i,H}$ and $v_{U,U} = \sum_{i \in U} v_{i,U}$. Multiplying both sides of this equation by N_H and then re-arranging, we have

$$N_H = \frac{y_{U,H}}{v_{U,U}/N_H}. \tag{2.8}$$

Equation 2.8 is the most general version of Equation 2.2 which we derived in the previous section. Equation 2.8 is an expression for the size of the hidden population that does not depend on any assumptions about network structure or reporting accuracy; it is just a different way of expressing the identity that the total number of out-reports must equal the total number of in-reports. If we could estimate the two

terms on the right side of Equation 2.8—one term related to out-reports ($y_{U,H}$) and one term related to in-reports ($v_{U,U}/N_H$)—then we could estimate N_H.

However, in order to make the identity in Equation 2.8 useful in practice we need to modify it to account for an important logistical requirement of survey research. In real scale-up studies, researchers do not sample from the entire population U, but instead they sample from a subset of U called the frame population, F. For example, in almost all scale-up studies the frame population has been adults. In standard survey research, restricting interviews to the frame population does not cause problems because respondents report about themselves: the group making the reports is the same as the group that inferences are being made about. However, with the scale-up method respondents report about others, so it is not necessarily the case that the reports are the same as the group that inferences are being made about. As we show in Section 3.4.2 of Chapter 3, failure to consider this fact requires the introduction of an awkward adjustment factor. Rather than attempting to make an adjustment after the fact, we derive an identity explicitly in terms of the frame population. Restricting our attention to out-reports coming from people in the frame population, it must be the case that

$$N_H = \frac{y_{F,H}}{v_{U,F}/N_H}, \tag{2.9}$$

where $y_{F,H} = \sum_{i \in F} y_{i,H}$ and $v_{U,F} = \sum_{i \in U} v_{i,F}$. The only difference between Equation 2.8 and Equation 2.9 is that Equation 2.9 restricts out-reports and in-reports to come from people in the frame population (Figure 2.1(c)). The identity in Equation 2.9 is extremely general: it does not depend on any assumptions about the relationship between the entire population U, the frame population F, and the hidden population H. For example, it holds if no members of the hidden population are in the frame population, if there are barrier effects, and if there are transmission errors.

Thus, if we could estimate the two terms on the right side of Equation 2.9—one term related to out-reports ($y_{F,H}$) and one term related to in-reports ($v_{U,F}/N_H$)—then we could estimate N_H under very general conditions.

Unfortunately, despite repeated attempts, we were unable to develop a practical method for estimating the term related to in-reports ($v_{U,F}/N_H$). However, if we make an assumption about respondents' reporting behavior, then we can re-express Equation 2.9 as an identity made up of quantities that we can actually estimate. Specifically, if we assume that the out-reports from people in the frame population only include people in the hidden population, then it must be the case that the visibility of everyone not in the hidden population is 0: $v_{i,F} = 0$ for all $i \notin H$. We call this reporting assumption the *no false positives assumption* for reasons that will become clear in a moment. If the no false positives assumption holds, we can re-write Equation 2.9 as

$$N_H = \frac{y_{F,H}}{v_{U,F}/N_H} = \frac{y_{F,H}}{\bar{v}_{H,F}}, \tag{2.10}$$

where $\bar{v}_{H,F} = v_{H,F}/N_H$. Equation 2.10 is a mathematical version of Equation 2.2, which we saw above. Chapter 3 introduces the *generalized network scale-up estimator*, which is based on the identity in Equation 2.10.

2.2.2 Reporting errors

To understand the no false positives assumption substantively, consider the two possible types of reporting errors: false positives and false negatives. *False negatives* occur when a respondent is connected to someone in the hidden population, but does not report this fact. This possibility has been discussed at length in research about the scale-up method (Bernard et al., 2010, see also Chapter 3). Since hidden populations like drug injectors are often stigmatized, it is reasonable to suspect that false negatives will be a serious problem for estimators based on network reports. Fortu-

nately, Equation 2.10 holds even if there are false negative reporting errors. (However, false negatives can cause problems if the underlying social network is being used to estimate visibility; see Chapter 3 and Chapter 5 for examples.)

False positives occur when a respondent erroneously reports a connection to someone who is not actually in the hidden population. For example, a respondent who is not connected to any drug injectors might incorrectly report that one of her acquaintances is a drug injector. These false positive reports are not accounted for in the identity in Equation 2.10; if false positives exist, then the quantity $y_{F,H}/\bar{v}_{H,F}$ will be larger than N_H. To see this, note that each individual i's out-reports can be divided into two groups: true positives, which actually connect to the hidden population ($y^+_{i,H}$); and false positives, which do not connect to the hidden population ($y^-_{i,H}$; see also Figure A.1). Therefore,

$$y_{i,H} = y^+_{i,H} + y^-_{i,H}, \tag{2.11}$$

and, in aggregate,

$$y_{F,H} = y^+_{F,H} + y^-_{F,H}, \tag{2.12}$$

where $y^+_{F,H} = \sum_{i \in F} y^+_{i,H}$ and $y^-_{F,H} = \sum_{i \in F} y^-_{i,H}$. Since the total number of true-positive out-reports must equal the total number of true-positive in-reports, an aggregate reporting identity must hold:

$$y^+_{F,H} = v_{H,F}. \tag{2.13}$$

Dividing both sides by $v_{H,F}$, and then multiplying both sides by N_H produces

$$N_H = \frac{y^+_{F,H}}{(v_{H,F}/N_H)} = \frac{y^+_{F,H}}{\bar{v}_{H,F}}. \tag{2.14}$$

Equation 2.14 holds even if there are false positive reporting errors. Now, returning to Equation 2.10 and re-writing it in terms of false positives and true positives produces

$$\frac{y_{F,H}}{\bar{v}_{H,F}} = \frac{y^{+}_{F,H} + y^{-}_{F,H}}{\bar{v}_{H,F}} \geq N_H, \tag{2.15}$$

where equality holds only if $y^{-}_{F,H} = 0$; i.e., if the no false positive assumption holds.

If we could develop estimators for $y^{+}_{F,H}$ instead of $y_{F,H}$, we could avoid having to make the no false positives assumption. Unfortunately, any attempt to estimate $y^{+}_{F,H}$ directly from F would produce information about $y_{F,H}$ instead. Therefore, we propose that researchers collect information about $y_{F,H}$ and then estimate an adjustment factor that relates $y_{F,H}$ to $y^{+}_{F,H}$. This approach leads us to introduce a new quantity called the *precision of out-reports*, η_F:

$$\eta_F = \frac{y^{+}_{F,H}}{y_{F,H}}. \tag{2.16}$$

The precision is useful because it relates the observed out-reports, $y_{F,H}$ to the true positive out-reports, $y^{+}_{F,H}$. It varies from 0, when none of the out-reports are true positives, to 1, when the out-reports are perfect. Even if false-positive reports are possible, the precision allows us to derive an identity that relates out-reports to N_H:

$$N_H = \frac{\eta_F \; y_{F,H}}{\bar{v}_{H,F}}. \tag{2.17}$$

Appendix B.1 discusses the challenges we encountered in trying to devise a way to measure the precision. We expect that the measurement of the precision of out-reports will pose a major challenge for future network reporting research, and we hope that practical solutions to this problem can be found. For the time being, we recommend that researchers show the impact that different values of the precision of out-reports would have on size estimates.

2.2.3 Individual visibility estimators

Following our analysis of the aggregate visibility estimator, we will begin by expressing the individual reporting identity in the most general situation using our mathematical notation. We will then modify the expression to account for the frame population, and then we will address the issue of reporting errors.

We can express the individual reporting identity in Equation 2.3 as

$$y_{U,j} = v_{j,U}, \tag{2.18}$$

which implies $y_{U,j}/v_{j,U} = 1$, as long as $v_{j,U} > 0$. Summing over all of the hidden population members yields

$$\sum_{j \in H} \frac{y_{U,j}}{v_{j,U}} = N_H. \tag{2.19}$$

Restricting reports so that they can only come from the frame population produces the individual reporting identity

$$y_{F,j} = v_{j,F}, \tag{2.20}$$

and the corresponding expression for N_H

$$\sum_{j \in H} \frac{y_{F,j}}{v_{j,F}} = N_H. \tag{2.21}$$

Equation 2.21 is an expression for N_H that is written in terms of a sum over the members of H. In order to turn an expression for N_H into a useful estimator, we must frame it in terms of reports from F. However, like the aggregate visibility estimator, it turns out that we must first make the assumption that there are no false positive reports. To see why, we will write true positive reports as $i \xrightarrow{+} j$, and false

positive reports as $i \stackrel{-}{\rightarrow} j$. Since $N_H = \sum_{j \in H} 1$, it is always the case that

$$\sum_{i \in F} \sum_{i \stackrel{+}{\rightarrow} j} \frac{1}{v_{j,F}} = \sum_{j \in H} \frac{y_{F,j}}{v_{j,F}} = N_H. \tag{2.22}$$

Equation 2.22 relies on the fact that only true positive reports from F ($i \stackrel{+}{\rightarrow} j$) contribute to the sum. More generally,

$$\sum_{i \in F} \sum_{i \rightarrow j} \frac{1}{v_{j,F}} = \sum_{i \in F} \left(\sum_{i \stackrel{+}{\rightarrow} j} \frac{1}{v_{j,F}} + \sum_{i \stackrel{-}{\rightarrow} j} \frac{1}{v_{j,F}} \right) \geq N_H. \tag{2.23}$$

From Equation 2.22, it is clear that equality will only hold here if the sum over $i \stackrel{-}{\rightarrow} j$ contributes nothing to the total; that is, if there are no false positive reports.

2.3 Conclusion

In this chapter, we introduced the network reporting framework. Network reporting is a general toolkit that researchers can use to help design and analyze surveys that ask respondents to report about people they are connected to in a wide range of different types of social network. Using the bipartite reporting graph, we saw that the network reporting framework can be used to deduce two important types of estimator: one based on adjusting for overcounting at the aggregate level, and one based on adjusting for overcounting at the individual level. The framework also enables researchers to reason about realistic features of surveys such as the sampling frame and reporting errors. This reasoning may lead to improved estimators and new methods for collecting data (Chapter 3), and it also enables researchers to develop analytical expressions that can be used to evaluate the impact different potential sources of error will have on substantive inferences (Chapters 3–5).

There are many important directions for future work, and we will only mention a few of these here. First, in order to be used in practice, both aggregate and individual estimators currently require that researchers make the assumption that there are no false positive reports. For some hidden populations (like deaths), this seems reasonable. For other hidden populations (like men who have sex with men), this may be less reasonable. It would be extremely helpful to understand how to estimate the extent of false positive reporting, even in specialized situations.

Second, researchers would benefit from having more guidance about when aggregate estimators should be preferred and when individual estimators should be preferred. For example, are there situations that would guarantee that estimates produced with one type of estimator will have smaller standard errors? Or, are there empirical or mathematical reasons to expect one type of estimator or the other to generally be more robust to reporting errors? Individual estimators require more information, since respondents have to be able to report about the number of times each hidden population member could have been reported (v_j). Intuition suggests that extra information should be an advantage, but it is not obvious that this is the case.

Third, our focus in this dissertation is on estimating the size of a hidden population. Future work could explore how network reporting approaches can be extended to estimate categorical or continuous quantities as well.

Finally, although this dissertation is focused on a model-free, design-based perspective, future work could use insights from the network reporting framework as a foundation for building models of hidden population size and social network structure from sampled data.

Chapter 3

Generalized network scale-up

We now illustrate the network reporting framework by using it to analyze an existing approach called network scale-up. The core insight behind the network scale-up method is that ordinary people have embedded within their personal networks information that can be used to estimate the size of hidden populations, if that information can be properly collected, aggregated, and adjusted (Bernard et al., 1989, 2010). In a typical scale-up survey, randomly sampled adults are asked about the number of connections they have to people in a hidden population (e.g., "How many people do you know who inject drugs?") and a series of similar questions about groups of known size (e.g., "How many widowers do you know?"; "How many doctors do you know?"). Responses to these questions are called *aggregate relational data* (McCormick et al., 2012a).

To produce size estimates from aggregate relational data, previous researchers have begun with the *basic scale-up model*, which makes three important assumptions: (i) social ties are formed completely at random (i.e., random mixing); (ii) respondents are perfectly aware of the characteristics of their alters (i.e., the people they are connected to); and (iii) respondents are able to provide accurate answers to survey questions about their personal networks. From the basic scale-up model Killworth

et al. (1998b) derived the *basic scale-up estimator*. This estimator, which is widely used in practice, has two main components. For the first component, the aggregate relational data about the hidden population are used to estimate the number of connections that respondents have to the hidden population. For the second component, the aggregate relational data about the groups of known size are used to estimate the number of connections that respondents have in total. For example, a researcher might estimate that members of her sample have 5,000 connections to people who inject drugs and 100,000 connections in total. The basic scale-up estimator combines these pieces of information to estimate that 5% $(5,000/100,000)$ of the population injects drugs. This estimate is a sample proportion, but rather than being taken over the respondents, as would be typical in survey research, the proportion is taken over the respondents' alters. Researchers who desire absolute size estimates multiply the alter sample proportion by the size of the entire population, which is assumed to be known.

Unfortunately, the three assumptions underlying the basic scale-up model have all been shown to be problematic. Scale-up researchers call violations of the random mixing assumption *barrier effects* (Killworth et al., 2006; Zheng et al., 2006; Maltiel et al., 2015); they call violations of the perfect awareness assumption *transmission error* (Shelley et al., 1995, 2006; Killworth et al., 2006; Salganik et al., 2011b; Maltiel et al., 2015); and they call violations of the respondent accuracy assumption *recall error* (Killworth et al., 2003, 2006; McCormick and Zheng, 2007; Maltiel et al., 2015). In order to address some of these three problems, Maltiel et al. (2015) developed a more complex, flexible version of the basic scale-up model, but the quality of estimates from that model is still unknown.

Rather than depending on the basic scale-up model (or its variants), we use the network reporting framework to reveal that one of the two main components of the basic scale-up estimator is problematic. Therefore, we propose a new estimator—the

generalized scale-up estimator—that combines the aggregate relational data traditionally used in scale-up studies with similar data collected from the hidden population. Collecting data from the hidden population is a major departure from current scale-up practice, but it enables researchers to estimate a critical quantity directly. If no hidden population sample is available, then researchers must make several assumptions about this critical quantity in order to use the basic scale-up method. Our study clarifies precisely what those basic scale-up assumptions are, and how violating them impacts hidden population size estimates.

In the next section, we derive the generalized scale-up estimator, and we describe the data collection procedures needed to use it. In Section 3.2, we compare the generalized and basic scale-up estimators. This comparison highlights the fact that the generalized scale-up estimator is consistent in a much wider range of situations than the basic scale-up estimator. We then propose a decomposition that separates the difference between them into three measurable and substantively meaningful factors (Equation 3.10). In Section 3.3, we propose a new variance estimation procedure for both basic and generalized scale-up estimators, and we demonstrate that it outperforms the existing variance estimator. In Section 3.4 we make practical recommendations for the design and analysis of future scale-up studies, and in Section 3.5, we conclude with a summary and an outline of next steps. Appendices A - G provide technical details and supporting arguments.

3.1 The generalized scale-up estimator

3.1.1 Estimating N_H from sampled data

The generalized scale-up estimator is based on an aggregate visibility identity (Chapter 2), under the assumption that there are no false positive reports. Equation 2.10 relates our quantity of interest, the size of the hidden population (N_H), to two other

Hidden population(s)	Location	Citation
Mortality in earthquake	Mexico City, Mexico	(Bernard et al., 1989)
Rape victims	Mexico City, Mexico	(Bernard et al., 1991)
HIV prevalence, rape, and homelessness	U.S.	(Killworth et al., 1998b)
Heroin use	14 U.S. cities	(Kadushin et al., 2006)
Choking incidents in children	Italy	(Snidero et al., 2007, 2009, 2012)
Groups most at-risk for HIV/AIDS	Ukraine	(Paniotto et al., 2009)
Heavy drug users	Curitiba, Brazil	(Salganik et al., 2011a)
Groups most at-risk for HIV/AIDS	Kerman, Iran	(Shokoohi et al., 2012)
Men who have sex with men	Japan	(Ezoe et al., 2012)
Groups most at-risk for HIV/AIDS	Almaty, Kazakhstan	(Scutelniciuc, 2012a)
Groups most at-risk for HIV/AIDS	Moldova	(Scutelniciuc, 2012b)
Groups most at-risk for HIV/AIDS	Thailand	(Aramrattan and Kanato, 2012)
Groups most at-risk for HIV/AIDS	Rwanda	(Rwanda Biomedical Center, 2012)
Groups most at-risk for HIV/AIDS	Chongqing, China	(Guo et al., 2013)
Groups most at-risk for HIV/AIDS	Tabriz, Iran	(Khounigh et al., 2014)
Men who have sex with men	Taiyuan, China	(Jing et al., 2014)
Drug and alcohol users	Kerman, Iran	(Sheikhzadeh et al., 2014)

Table 3.1: Network scale-up studies that have been completed.

quantities: the total number of out-reports from the frame population ($y_{F,H}$) and the average number of in-reports in the hidden population ($\bar{v}_{H,F}$). We now show how to estimate $y_{F,H}$ with a probability sample from the frame population and $\bar{v}_{H,F}$ with a relative probability sample from the hidden population.

The total number of out-reports ($y_{F,H}$) can be estimated from respondents' reported number of connections to the hidden population,

$$\widehat{y}_{F,H} = \sum_{i \in s} \frac{y_{i,H}}{\pi_i}, \tag{3.1}$$

where s denotes the sample, $y_{i,H}$ denotes the reported number of connections between i and H, and π_i is i's probability of inclusion from a conventional probabilistic sampling design from the frame population. Because $\widehat{y}_{F,H}$ is a standard Horvitz-Thompson estimator, it is consistent and unbiased as long as all members of F have a positive probability of inclusion under the sampling design (Sarndal et al., 2003). This result is stated formally in Appendix B.2 (Result B.2.1).

Estimating the average number of in-reports for the hidden population ($\bar{v}_{H,F}$) is more complicated. First, it will usually be impossible to obtain a conventional probability sample from the hidden population. Fortunately, as we show below, estimating $\bar{v}_{H,F}$ only requires a relative probability sampling design in which hidden population members have a nonzero probability of inclusion and respondents' probabilities of inclusion are known up to a constant of proportionality, $c\pi_i$ (see Appendix B.3.1). Under certain conditions, two widely used sampling methods designed for hard-to-reach populations offer this possibility: respondent-driven sampling (Heckathorn, 1997) and time-location sampling (Karon and Wejnert, 2012). Of course, any problems with these sampling methods could introduce problems into estimates.

A second problem arises because we do not expect respondents to be able to easily and accurately answer direct questions about their visibility ($v_{i,F}$). That is, we do not expect respondents to be able to answer questions such as "How many people on the sampling frame would include you when reporting a count of the number of drug injectors that they know?" Instead, we propose asking hidden population members a series of questions about their connections to certain groups and their visibility to those groups. For example, each sampled hidden population respondent could be asked "How many widowers do you know?" and then "How many of these widowers are aware that you inject drugs?" This question pattern can be repeated for many groups (e.g., widowers, doctors, etc.). We call data with this structure *enriched aggregate relational data* to emphasize its similarity to the aggregate relational data that is familiar to scale-up researchers. An interviewing procedure called the *game of contacts* enables researchers to collect enriched aggregated relational data, even in realistic field settings (Salganik et al., 2011b; Maghsoudi et al., 2014).

Given a relative probability sampling design and enriched aggregate relational data, we can now formalize our proposed estimator for $\bar{v}_{H,F}$. Let $A_1, A_2, \ldots, A_J$, be the set of groups about which we collect enriched aggregate relational data (e.g., wid-

owers, doctors, etc). Here, to keep the notation simple, we assume that these groups are all contained in the frame population, so that $A_j \subset F$ for all j; in Appendix B.3.4 we extend the results to groups that do not meet this criterion. Let $\mathcal{A}$ be the concatenation of these groups, which we call the *probe alters*. For example, if A_1 is widowers and A_2 is doctors, then the probe alters $\mathcal{A}$ is the collection of all widowers and all doctors, with doctors who are widowers included twice. Also, let $\tilde{v}_{i,A_j}$ be respondent i's report about her visibility to people in A_j and let v_{i,A_j} be respondents i's actual visibility to people in A_j (i.e., the number of times that this respondent would be reported about if everyone in A_j was asked about their connections to the hidden population).

The estimator for $\bar{v}_{H,F}$ is:

$$\widehat{\bar{v}}_{H,F} = \frac{N_F}{N_{\mathcal{A}}} \frac{\sum_{i \in s_H} \sum_j \widetilde{v}_{i,A_j}/(c\pi_i)}{\sum_{i \in s_H} 1/(c\pi_i)}, \tag{3.2}$$

where $N_{\mathcal{A}}$ is the number of probe alters, c is the constant of proportionality from the relative probability sample, and s_H is a relative probability sample of the hidden population. Equation 3.2 is a standard weighted sample mean (Sarndal et al., 2003, Sec. 5.7) multiplied by a constant, $N_F/N_{\mathcal{A}}$. Result B.3.2 shows that, in order for this estimator to be consistent and essentially unbiased,[1] two main conditions must be satisfied.

The first condition required for the estimator in Equation 3.2 is the *probe alter condition*, which describes the required relationship between the visibility of the hidden population to the probe alters and the visibility of the hidden population to the

[1]We use the term essentially unbiased because Equation 3.2 is not, strictly speaking, unbiased; however, the large literature on ratio estimation confirms that these estimators are, for all practical purposes, unbiased (e.g. Sarndal et al., 2003, chap. 5). For a more precise discussion of this issue, see Appendix B.5.

frame population:

$$\frac{v_{H,\mathcal{A}}}{N_{\mathcal{A}}} = \frac{v_{H,F}}{N_F}, \tag{3.3}$$

where $v_{H,\mathcal{A}}$ is the total visibility of the hidden population to the probe alters, $v_{H,F}$ is the total visibility of the hidden population to the frame population, $N_{\mathcal{A}}$ is the number of probe alters, and N_F is the number of people in the frame population. In words, Equation 3.3 says that the rate at which the hidden population is visible to the probe alters must be the same as the rate at which the hidden population is visible to the frame population. For example, in a study to estimate the number of drug injectors in a city, drug treatment counselors would be a poor choice for membership in the probe alters because drug injectors are probably more visible to drug treatment counselors than to typical members of the frame population. On the other hand, postal workers would probably be a reasonable choice for membership in the probe alters because drug injectors are probably about as visible to postal workers as they are to typical members of the frame population. Additional results about the probe alter condition are presented in Appendix B.3: (i) Result B.3.3 presents three other algebraically equivalent formulations of probe alter condition, some of which offer additional intuition; (ii) Result B.3.4 provides a method to empirically test the probe alter condition; and (iii) Result B.3.5 quantifies the bias in estimates when the probe alter condition is not satisfied.

The second condition required for the estimator in Equation 3.2 is *accurate aggregate reports about visibility*:

$$\tilde{v}_{H,\mathcal{A}} = v_{H,\mathcal{A}}, \tag{3.4}$$

where $\tilde{v}_{H,\mathcal{A}}$ is the total reported visibility of members of the hidden population to the probe alters ($\sum_{i \in H} \sum_{j \in J} \tilde{v}_{i,A_j}$) and $v_{H,\mathcal{A}}$ is the total actual visibility of members of

the hidden population to the probe alters ($\sum_{i \in H} \sum_{j \in J} v_{i,A_j}$). In words, Equation 3.4 says that hidden population members must be correct in their reports about their visibility to probe alters in aggregate, but Equation 3.4 does not require the stronger condition that each individual report be accurate. In practice, there are two main ways that there might not be accurate aggregate reports about visibility. First, hidden population members might not be accurate in their assessments of what others know about them. For example, research on the "illusion of transparency" suggests that people tend to over-estimate how much others know about them (Gilovich et al., 1998). Second, although we propose asking hidden population members what other people know about them (e.g., "How many of these widowers know that you are a drug injector?") what actually matters for the estimator is what other people would report about them (e.g., "How many of these widowers would include you when reporting a count of the number of drug injectors that they know?"). In cases where the hidden population is extremely stigmatized, some respondents to the scale-up survey might conceal the fact that they are connected to people whom they know to be in the hidden population, and if this were to occur, it would lead to a difference between the information that we collect ($\tilde{v}_{i,\mathcal{A}}$) and the information that we want ($v_{i,\mathcal{A}}$). Unfortunately, there is currently no empirical evidence about the possible magnitude of these two problems in the context of scale-up studies. However, Result B.3.4 quantifies the bias introduced into estimates if the assumption of accurate aggregate reports about visibility is not correct.

To recap, using two different data collection procedures—one with the frame population and one with the hidden population—we can estimate the two components of the expression for N_H given in Equation 2.10. Further, we can combine these components to form the *generalized scale-up estimator*:

$$\widehat{N}_H = \frac{\widehat{y}_{F,H}}{\widehat{\widehat{v}}_{H,F}}. \tag{3.5}$$

Result B.3.9 demonstrates that $\widehat{N}_H$ will be consistent and essentially unbiased if the estimator for the numerator is consistent and unbiased, if the estimator for the denominator is consistent and essentially unbiased, and if there are no false positive reports.

One attractive feature of the generalized scale-up estimator (Equation 3.5) is that it is a simple combination of standard survey estimators. This structure enables us to derive very general sensitivity results about the impact of violations of various assumptions, either individually or jointly, on the estimated size of the hidden population. These sensitivity results are summarized in Appendix B.4.

3.2 Comparison between the generalized and basic scale-up estimators

Having introduced the generalized scale-up estimator, we now compare it to the basic scale-up estimator, which was used in most of the studies listed in Table 3.1. While the generalized scale-up estimator is derived from an identity relating in-reports and out-reports, the basic scale-up estimator was derived from a social network model. In order to review this model, we need to define another quantity: we call $d_{i,U}$ person i's degree, the number of undirected network connections she has to everyone in U.

The basic scale-up model assumes that each person's connections are formed independently, that reporting is perfect, and that visibility is perfect (Killworth et al., 1998b). Together, these three assumptions lead to the probabilistic model:

$$y_{i,A_j} = d_{i,A_j} \sim \text{Binomial}\left(d_{i,U}, \frac{N_{A_j}}{N}\right), \tag{3.6}$$

for all i in U and for any group A_j. In words, this model suggests that the number of connections from a person i to members of a group A_j is the result of a series of

$d_{i,U}$ independent random draws, where the probability of each edge being connected to A_j is $\frac{N_{A_j}}{N}$.

The basic scale-up model leads to what we call a basic scale-up estimator:

$$\widehat{N}_H = \frac{\sum_{i \in s} y_{i,H}}{\sum_{i \in s} \widehat{d}_{i,U}} \times N, \tag{3.7}$$

where $\widehat{d}_{i,U}$ is the estimated degree of respondent i from the known population method (Killworth et al., 1998a). Killworth et al. (1998b) showed that Equation 3.7 is the maximum-likelihood estimator for N_H under the basic scale-up model, conditional on the additional assumption that $d_{i,U}$ is known for each $i \in s$.

In order to investigate when the basic and generalized scale-up estimators might be the same, we will compare the population quantities that basic and generalized scale-up aim to estimate. By comparing these two estimands, we will develop a framework for understanding the basic scale-up assumptions in the context of design-based estimation. In a census of the population where we observe degrees, we can write the basic scale-up estimator in Equation 3.7 as

$$\widehat{N}_H = \frac{y_{F,H}}{d_{F,U}} \times N = \frac{y_{F,H}}{\bar{d}_{U,F}}, \tag{3.8}$$

where $d_{F,U} = \sum_{i \in F} d_{i,U}$ and $\bar{d}_{U,F} = d_{U,F}/N = d_{F,U}/N$.

In Section 3.1, we showed that the generalized scale-up estimand is

$$\widehat{N}_H = \frac{y_{F,H}}{\bar{v}_{H,F}}. \tag{3.9}$$

By comparing Equations 3.8 and 3.9, we conclude that these two estimands will be the same when $\bar{v}_{H,F} = \bar{d}_{U,F}$. This comparison reveals that the basic scale-up estimand is equal to the hidden population size only in the special case when $\bar{v}_{H,F} = \bar{d}_{U,F}$, a result that follows from the identity developed in Section 3.1.

In order to help researchers reason about situations where $\bar{d}_{U,F}$ may or may not be equal to $\bar{v}_{H,F}$, we propose a decomposition that separates the difference between the two estimands into three measurable and substantively meaningful *adjustment factors*:

$$N_H = \underbrace{\left(\frac{y_{F,H}}{\bar{d}_{U,F}}\right)}_{\substack{\text{basic}\\ \text{scale-up}}} \times \underbrace{\frac{1}{\bar{d}_{F,F}/\bar{d}_{U,F}}}_{\substack{\text{frame ratio}\\ \phi_F}} \times \underbrace{\frac{1}{\bar{d}_{H,F}/\bar{d}_{F,F}}}_{\substack{\text{degree ratio}\\ \delta_F}} \times \underbrace{\frac{1}{\bar{v}_{H,F}/\bar{d}_{H,F}}}_{\substack{\text{true positive rate}\\ \tau_F}} = \underbrace{\left(\frac{y_{F,H}}{\bar{v}_{H,F}}\right)}_{\substack{\text{generalized}\\ \text{scale-up}}}. \tag{3.10}$$

$$\underbrace{\qquad\qquad\qquad\qquad\qquad\qquad\qquad\qquad\qquad\qquad}_{\text{adjustment factors}}$$

We now describe each of the three adjustment factors in turn. First, we define the frame ratio, ϕ_F, to be

$$\phi_F = \frac{\text{avg \# connections from a member of } F \text{ to the rest of } F}{\text{avg \# connections from a member of } U \text{ to } F} = \frac{\bar{d}_{F,F}}{\bar{d}_{U,F}}. \tag{3.11}$$

ϕ_F can range from zero to infinity, and in most practical situations we expect ϕ_F will be greater than one. Result B.2.6 shows that we can make consistent and essentially unbiased estimates of ϕ_F from a sample of F.[2]

Next, we define the degree ratio δ_F to be

$$\delta_F = \frac{\text{avg \# connections from a member of } H \text{ to } F}{\text{avg \# connections from a member of } F \text{ to the rest of } F} = \frac{\bar{d}_{H,F}}{\bar{d}_{F,F}}. \tag{3.12}$$

δ_F ranges from zero to infinity, and it is less than one when the hidden population members have, on average, fewer connections to the frame population than frame population members. Result B.3.7 shows that we can make consistent and essentially unbiased estimates of δ_F from samples of F and H.

[2]Note that, since $\bar{d}_{U,F} = (N_F/N)\ \bar{d}_{F,U}$, an equivalent expression for the frame ratio is

$$\phi_F = \frac{\bar{d}_{F,F}}{\bar{d}_{F,U}\ (N_F/N)} = \frac{\bar{d}_{F,F}}{\bar{d}_{F,U}}\ \frac{N}{N_F}.$$

Finally, we define the true positive rate, τ_F, to be

$$\tau_F = \frac{\#\text{ in-reports to } H \text{ from } F}{\#\text{ edges connecting } H \text{ and } F} = \frac{v_{H,F}}{d_{H,F}} = \frac{\bar{v}_{H,F}}{\bar{d}_{H,F}}. \tag{3.13}$$

τ_F relates network degree to network reports.[3] τ_F ranges from 0, if none of the edges are correctly reported, to 1 if all of the edges are reported. Substantively, the more stigmatized the hidden population, the closer we would expect τ_F to be to 0. Result B.3.8 shows that we can make consistent and essentially unbiased estimates of τ_F from a sample of H.

The decomposition in Equation 3.10 can be used to derive an expression for the bias in the basic scale-up estimator when we have a census and degrees are known:

$$\text{bias}(\widehat{N}_H^{\text{basic}}) \equiv \widehat{N}_H^{\text{basic}} - N_H \tag{3.14}$$

$$= \widehat{N}_H^{\text{basic}} \ \left[1 - (\phi_F \ \delta_F \ \tau_F)\right]. \tag{3.15}$$

Equation 3.15 reveals that if the product of the three adjustment factors ($\phi_F \ \delta_F \ \tau_F$) is 1, then the basic scale-up estimand is equal to the size of the hidden population. In all other cases it will not be.

The comparison between the basic and generalized scale-up estimators leads to two main conclusions. First, our comparison shows that while the generalized scale-up estimand is always equal to the hidden population size, the basic scale-up estimand is only equal to the hidden population size when additional conditions hold, a result that we further confirm and illustrate with a simulation study in Appendix B.7. Second, as Equation 3.10 shows, if the adjustment factors ϕ_F, δ_F, and τ_F are known, then

[3]Note that the fact that in-reports must equal out-reports means that τ_F can also be defined

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

Here we have written $y^+_{F,H}$ to mean the true positive reports among the $y_{F,H}$; see Appendix B.1 for a detailed explanation.

they can be used to improve basic scale-up estimates. Note that our comparison here has focused on what conditions lead to estimands—the population quantities being estimated by basic and generalized scale-up—that are equal to the hidden population size. The decision to use either a basic or a generalized scale-up estimator should also take into consideration the different assumptions needed to use each estimator with sampled data. We review these assumptions in Section 3.4 below.

3.3 Variance estimation

In addition to point estimates, the focus of the chapter thus far, researchers must also put accurate confidence intervals around their estimates. Unfortunately, the variance estimator currently used with the basic scale-up estimator has several problems and cannot be easily extended to accommodate the new generalized scale-up estimator (Equation 3.5). Therefore, we propose a new variance estimation approach based on the rescaled bootstrap (Rao et al., 1992; Rust and Rao, 1996). This rescaled bootstrap approach—which researchers have previously applied to other estimators—both improves variance estimation for the basic scale-up estimator and enables variance estimation for the generalized scale-up estimator.

The variance estimator currently used in scale-up studies was proposed in Killworth et al. (1998b):

$$\widehat{se}(\widehat{N}_H) = \sqrt{\frac{N \cdot \widehat{N}_H}{\sum_{i \in s} \widehat{d}_{i,U}}}. \tag{3.16}$$

This estimator was derived from the basic scale-up model (Equation 3.6), and so it suffers from the limitations of that model. In particular, it has three main problems, all of which lead it to produce intervals that tend to be too small to include the true value as often as they should. First, the variance estimator in Equation 3.16 does not include any information about the procedure used to sample respondents, which can lead to problems when complex sampling designs, such as stratified, multi-stage

designs, are used. Second, it implicitly assumes that the researchers have learned about $\sum_{i \in s} d_{i,U}$ independent alters, which is not true if there are barrier effects (i.e., non-random social mixing). Finally, like virtually all variance estimators, it only provides a measure of uncertainty introduced by sampling but not other possible sources of measurement error. To address the first two problems (but not the third), in Appendix B.6 we propose a rescaled bootstrap variance estimation procedure (Rao et al., 1992; Rust and Rao, 1996) that has strong theoretical foundations; does not depend on the basic scale-up model; can handle both simple and complex sample designs; and can be used for both the basic scale-up estimator and the generalized scale-up estimator.

Unfortunately, we cannot assess the performance of the rescaled bootstrap procedure when used with the generalized scale-up estimator because the generalized scale-up estimator has not yet been used for populations of known size. However, we can compare the coverage properties of several variance estimators when used in conjunction with the basic scale-up estimator using the internal consistency procedure proposed in Killworth et al. (1998a). In particular, for three real scale-up datasets—one collected via simple random sampling (McCarty et al., 2001) and two collected via complex sample designs (Salganik et al., 2011a; Rwanda Biomedical Center, 2012)—for each group of known size, we produced a size estimate using the basic scale-up estimator (Equation 3.7), and we produced confidence intervals using (i) the current variance estimation procedure (Equation 3.16); (ii) the simple bootstrap, which does not account for complex sample designs; and (iii) the rescaled bootstrap, which does account for complex sample designs.[4] This empirical evaluation (Figure 3.1) produced three main results. First, as expected, we found that the current variance estimation procedure performed poorly: purported 95% confidence intervals had empirical

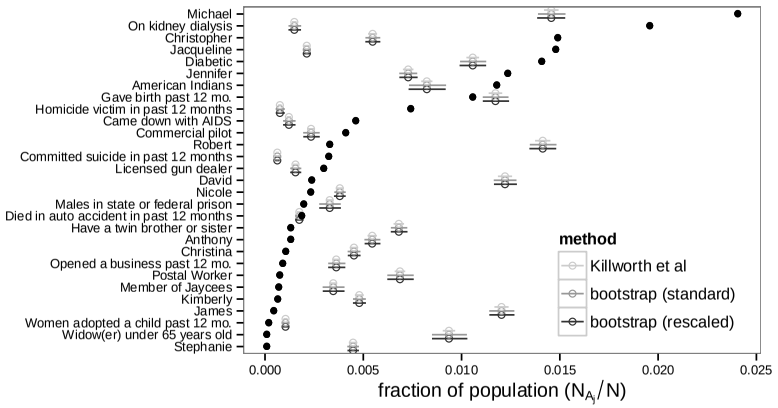
[4]Computer code to perform these calculations was written in R (R Core Team, 2014) and used the following packages: networkreporting (Feehan and Salganik, 2014); ggplot2 (Wickham, 2009); devtools (Wickham and Chang, 2013); plyr (Wickham, 2011); functional (Danenberg, 2013); and stringr (Wickham, 2012).

coverage rates of about 5%. Second, also consistent with expectation, we found that the rescaled bootstrap procedure produced wider intervals than both the current procedure and the simple bootstrap, especially in the case of complex sample designs. Third, and somewhat surprisingly, the rescaled bootstrap procedure did not work well in an absolute sense: purported 95% confidence intervals had empirical coverage rates of about 10%, only slightly better than the current procedure. We do not know the exact reason for the poor coverage rate of the rescaled bootstrap, but we suspect that it is caused by biases in the basic scale-up estimator (see Equation 3.10) and non-sampling errors not captured in the resampling procedure.

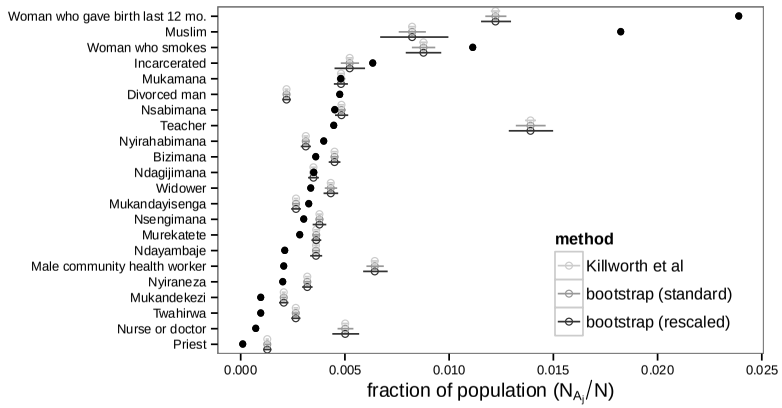
These empirical results, and the theoretical arguments in Appendix B.6, lead us to three conclusions. First, confidence intervals from the rescaled bootstrap procedure are marginally preferable to intervals from the current scale-up variance estimation procedure. Second, researchers should expect that the confidence intervals from the rescaled bootstrap procedure will understate the true uncertainty in scale-up estimates until there is a substantially better understanding of non-sampling errors. Third, variance estimation for the scale-up method is an important area for further research.

3.4 Recommendations for practice

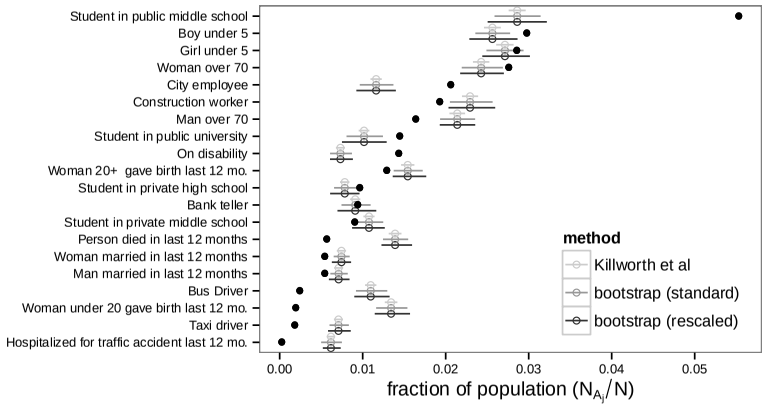
Our analytic results in Sections 3.1 and 3.2 lead to us to recommend a major departure from current scale-up practice. In addition to collecting a sample from the frame population, we recommend that researchers also collect a sample from the hidden population so that they can use the generalized scale-up estimator. We think that the generalized scale-up estimator is worth the extra effort, because it avoids the need to make several assumptions about the adjustment factors described in Section 3.2.



(a) United States (simple random sample)



(b) Rwanda (stratified, multi-stage)



(c) Curitiba, Brazil (multi-stage)

Figure 3.1: Assessing variance estimation procedures using scale-up studies in the United States (McCarty et al., 2001), Rwanda (Rwanda Biomedical Center, 2012), and Curitiba, Brazil (Salganik et al., 2011a). The true size of each group is shown with a black dot. Estimates made use the basic scale-up estimator are shown with circles. The rescaled bootstrap confidence intervals include the true group size for 3.4%, 13.6%, and 15.0% of the groups in the US, Rwanda, and Curitiba, respectively. The standard bootstrap confidence intervals include the true group size for 3.4%, 9.1%, and 10.0% of the groups. The currently used procedure (Equation 3.16), contains the true group size for 0.0%, 9.1%, and 5.0% of the groups.

However, we also acknowledge that it might not be possible to collect a sample from the hidden population in all situations, and that some researchers may prefer making assumptions about adjustment factors to collecting information from the hidden population. Fortunately, researchers can use our results to make this decision for themselves. Table 3.2 summarizes the conditions needed to produce consistent and essentially unbiased estimates using the basic and generalized scale-up estimators, and Appendix B.4 shows how the size estimates from each approach will be impacted when these conditions are not met.

3.4.1 Estimation with samples from F and H

We recommend that researchers who have samples from F and H use a generalized scale-up estimator to produce estimates of N_H (see Section 3.1):

$$\widehat{N}_H = \frac{\widehat{y}_{F,H}}{\widehat{\widehat{v}}_{H,F}}. \tag{3.17}$$

For researchers using the generalized scale-up estimator we have three specific recommendations. First, we recommend that the probe alters be designed so that the rate at which the hidden population is visible to the probe alters is the same as the rate at which the hidden population is visible to the frame population (see Result B.3.2 for a more formal statement, and see Section B.3.5 for more advice about choosing probe alters). Second, because the generalized scale-up estimator depends on assumptions (e.g., assumptions about the probe alters and reporting behavior), we recommend that researchers assess the sensitivity of their estimates to any assumptions that may be particularly problematic in their setting. Fortunately, the generalized scale-up estimator is a simple combination of relatively simple estimators so it is possible to assess sensitivity analytically; see Appendix B.1 and Appendix B.4 for more information. Finally, we recommend that researchers produce confidence intervals around their

estimate using the rescaled bootstrap procedure presented in Section 3.3, keeping in mind that this will likely produce intervals that are too small.

We also have three additional recommendations that will facilitate the cumulation of knowledge about the scale-up method. First, although the generalized scale-up estimator does not require aggregate relational data from the frame population about groups of known size, we recommend that researchers collect this data so that the basic and generalized estimators can be compared. Second, we recommend that researchers publish estimates of δ_F and τ_F, although these quantities play no role in the generalized scale-up estimator (Fig. 3.2). As a body of evidence about these adjustment factors accumulates (e.g., Salganik et al. (2011a); Maghsoudi et al. (2014)), studies that are not able to collect a sample from the hidden population will have an empirical foundation for adjusting basic scale-up estimates, either by borrowing values directly from the literature, or by using published values as the basis for priors in a Bayesian model. Finally, we recommend that researchers try to design their data collection so that other size estimation methods can also be used and compared to generalized scale-up estimates. For example, if respondent-driven sampling is used to collect information from H, then researchers could try to also collect the information needed to estimate hidden population size from H directly using the approaches described in Berchenko et al. (2013),Crawford et al. (2015), Handcock et al. (2015), Handcock et al. (2014), and UNAIDS (2010).

3.4.2 Estimation with only a sample from F

If researchers cannot collect a sample from the hidden population, we have four recommendations. First, we recommend two simple changes to the basic scale-up estimator that remove the need to adjust for the frame ratio, ϕ_F. Recall, that the basic scale-up

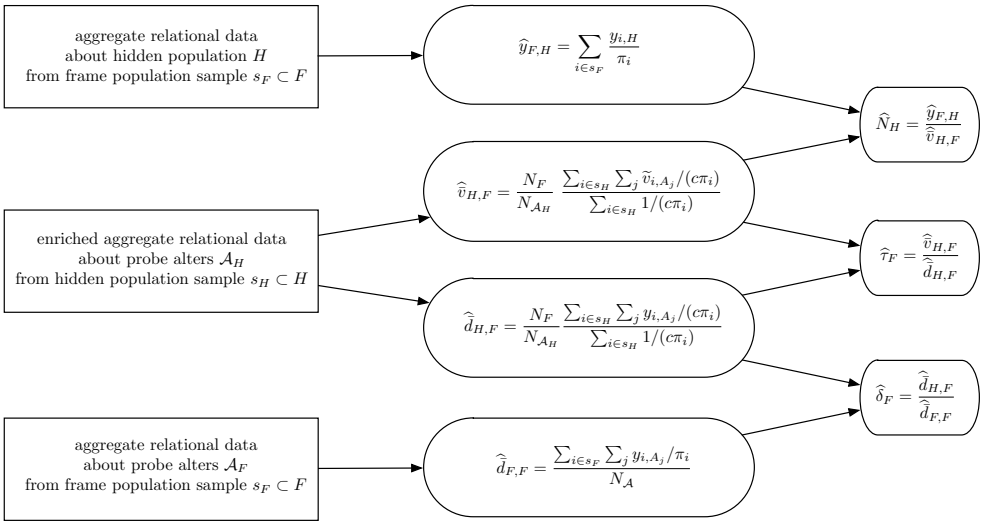


Figure 3.2: Recommended schematic of inputs and outputs for a study using the generalized scale-up estimator. We recommend that researchers produce size estimates using the generalized scale-up estimator, and that researchers produce estimates of the adjustment factors δ_F and τ_F in order to aid other researchers.

estimator that has been used in previous studies is (see Section 3.2):

$$\widehat{N}_H = \frac{\widehat{y}_{F,H}}{\widehat{d}_{F,U}} \times N = \frac{\widehat{y}_{F,H}}{\widehat{d}_{F,U}/N}. \tag{3.18}$$

Instead of Equation 3.18, we suggest a new estimator that more directly deals with the fact that researchers sample from the frame population F (typically adults), and not from the entire population U (adults and children):

$$\widehat{N}_H = \frac{\widehat{y}_{F,H}}{\widehat{d}_{F,F}} \times N_F = \frac{\widehat{y}_{F,H}}{\widehat{d}_{F,F}/N_F}. \tag{3.19}$$

There are two differences between our proposed basic scale-up estimator (Equation 3.19) and the one used in previous studies (Equation 3.18). First, we recommend that researchers estimate $\widehat{d}_{F,F}$ (i.e., the total number of connections between adults and adults) rather than $\widehat{d}_{F,U}$ (i.e., the total number of connections between adults and everyone). In order to do so, researchers should design the probe alters for the frame

population so that they have similar personal networks to the frame population; in Appendix B.2.4 we define this requirement formally, and in Section B.2.4 we provide guidance for choosing the probe alters. Second, we recommend that researchers use N_F rather than N. These two simple changes remove the need to adjust for the frame ratio ϕ_F, and thereby eliminate an assumption about an unmeasured quantity. An adjusted version of this proposed basic scale-up estimator would then be:

$$\widehat{N}_H = \underbrace{\frac{\widehat{y}_{F,H}}{(\widehat{d}_{F,F}/N_F)}}_{\substack{\text{proposed basic} \\ \text{scale-up}}} \times \underbrace{\frac{1}{\widehat{\delta}_F} \times \frac{1}{\widehat{\tau}_F}}_{\substack{\text{adjustment} \\ \text{factors}}} \tag{3.20}$$

Our second recommendation for researchers using a basic scale-up estimator is to be explicit about the values that they assume for δ_F and τ_F. Further, like researchers using the generalized scale-up estimator, we recommend that researchers using a basic scale-up estimator assess the sensitivity of their estimates to any assumptions that may be particularly problematic in their setting. These sensitivity assessments can be done analytically using the ideas developed in Appendix B.1 and Appendix B.4. Finally, we recommend that researchers produce confidence intervals around their estimate using the rescaled bootstrap procedure presented in Section 3.3, while explicitly accounting for the fact that there is uncertainty around assumed adjustment factors, and bearing in mind that this will likely produce intervals that are too small.

3.5 Conclusion and next steps

In this chapter, we developed the generalized network scale-up estimator. This new estimator improves upon earlier scale-up estimators in several ways: it enables researchers to use the scale-up method in populations with non-random social mixing and imperfect awareness about membership in the hidden population, and it accommodates data collection with complex sample designs and incomplete sampling

frames. We also compared the generalized and basic scale-up estimators, leading us to introduce a framework that makes the design-based assumptions of the basic scale-up estimator precise. Researchers who use either the basic or generalized scale-up estimator can use our results to assess the sensitivity of their size estimates to the various assumptions of each estimator.

The approach that we followed to derive the generalized scale-up estimator has three elements, and these elements may prove useful in other problems related to sampling in networks. First, we distinguished between the network of reports and the network of relationships. Second, using the network of reports, we derived a simple identity that permitted us to develop a design-based estimator free of any assumptions about the structure of the network of relationships. Third, we combined data from different types of samples. Together, these three elements may help researchers in other situations derive relatively simple, design-based estimators that are an important complement to complex, model-based techniques.

Although the generalized scale-up estimator has many attractive features, it also requires that researchers obtain two different samples, one from the frame population and one from the hidden population. In cases where studies of the hidden population are already planned (e.g., the behavioral surveillance studies of the groups most at-risk for HIV/AIDS), the necessary information for the generalized scale-up estimator could be collected at little additional cost by appending a modest number of questions to existing questionnaires. In cases where these studies are not already planned, researchers can collect their own data from the hidden population, or they can borrow estimated adjustment factors from other published studies.

The generalized scale-up estimator, like all estimators, depends on a number of assumptions. We think two of them will be most problematic in practice. First, the estimator depends on the assumption that there are no false positive reports, which is unlikely to be true in all situations. Although we have derived an estimator that

works even in the presence of false positive reports (Appendix B.1), we were not able to design a practical data collection procedure that would allow us to estimate one of the terms it requires. Second, the generalized scale-up estimator depends on the assumption that hidden population members have accurate aggregate awareness about visibility (Equation 3.4). That is, researchers have to assume that hidden population respondents can accurately report whether or not their alters would report them, and we expect this assumption will be difficult to check in most situations. The results in Table B.2 show how violations of these assumptions—no false positive reports and accurate aggregate awareness about visibility—impact the resulting size estimates.

Our results and their limitations highlight several directions for further work, and we think that two are most urgent. First, as discussed in Chapter 2, researchers need practical techniques for estimating the rate of false positive reporting. These estimates, combined with the estimator in Appendix B.1, would permit the relaxation of one of the most important remaining assumptions made by all scale-up studies to date. Second, the proposed bootstrap variance estimator improves on the current variance estimation procedure, but still performs poorly in an absolute sense. We believe that more research on non-sampling error could lead to more realistic confidence intervals. We hope that the framework developed in this chapter will provide a basis for this and other future research.

Quantity	Conditions required	Result
reported connections to H ($\widehat{y}_{F,H}$)	1. probability sample from F 2. no false positive reports about connections to H ($\eta_F = 1$)	Result B.2.1
average personal network size of F ($\widehat{\bar{d}}_{F,F}$)	1. probability sample from F 2. accurate groups of known size total $N_{\mathcal{A}}$ 3. probe alter condition ($\bar{d}_{\mathcal{A},F} = \bar{d}_{F,F}$) 4. accurate reporting condition ($y_{F,\mathcal{A}} = d_{F,\mathcal{A}}$)	Result B.2.3
average visibility of H ($\widehat{\bar{v}}_{H,F}$)	1. relative probability sample from H 2. accurate groups of known size total $N_{\mathcal{A}_H}$ 3. probe alter condition ($\frac{v_{\mathcal{A}\cap F,H}}{N_{\mathcal{A}\cap F}} = \frac{v_{H,F}}{N_F}$) 4. accurate aggregate reports about visibility ($\tilde{v}_{H,\mathcal{A}_H\cap F} = v_{H,\mathcal{A}_H\cap F}$)	Result B.3.2
generalized scale-up ($\widehat{N}_H = \frac{\widehat{y}_{F,H}}{\widehat{\bar{v}}_{H,F}}$)	1. conditions needed for $\widehat{y}_{F,H}$ 2. conditions needed for $\widehat{\bar{v}}_{H,F}$	Result B.3.9
basic scale-up ($\widehat{N}_H = \frac{\widehat{y}_{F,H}}{\widehat{\bar{d}}_{F,F}}$)	1. conditions needed for $\widehat{y}_{F,H}$ 2. condition needed for $\widehat{\bar{d}}_{F,F}$ 3. personal networks of H are the same average size as personal networks of F ($\delta_F = 1$) 4. no false negative reports about connections to H ($\tau_F = 1$)	Sections 3.1-3.2

Table 3.2: Summary of the conditions needed for the basic and generalized network scale-up estimators, and their components, to produce estimates that are consistent and essentially unbiased. This table uses the version of the basic scale-up estimator we recommend in Section 3.4.2. For the basic scale-up estimator, if the adjustment factors τ_F and δ_F exactly cancel, so that $\frac{1}{\delta_F} = \tau_F$, then they do not both need to be equal to 1.

Chapter 4

Quantity vs quality: A survey experiment to improve the network scale-up method

People everywhere are embedded in a wide range of different personal networks. These networks arise from all of the possible ways two people could be connected to one another: through relationships (such as friendship and kinship); through behaviors (such as sharing a meal or attending an event); through group affiliation (such as being in a professional organization or book club); and so forth. Researchers can use the network reporting framework (Chapter 2) as a basis for developing techniques to ask survey respondents about *any* of these networks, as long as network connections are symmetric (i.e., whenever i is connnected to j, then j is also connected to i).

Which of the many possible networks can be expected to provide researchers with the most useful information? The first goal of Chapter 4 is to explore this important question. We conjecture that researchers can trade off the *quantity* and the *quality* of the information obtained from survey respondents by varying the *strength* of the network tie that respondents report about. We provide a first test of this conjecture

using original data collected as part of a survey experiment that we designed and conducted in Rwanda. By randomly assigning respondents to one of two possible tie definitions, our experiment shows that a particular, relatively strong network tie produces more accurate estimates than a particular, relatively weak network tie. Thus, there may indeed be a trade-off between the quantity and the quality of survey reports. However, our results do not provide a completely general description of the relationship between tie strength and the accuracy of network reporting estimates. Therefore, our study also provides a methodological template that researchers can use to conduct similar experiments in the future. As evidence about the relationship between tie strength and reporting accuracy accumulates, our conjecture can be confirmed or replaced with a better one.

The second goal of Chapter 4 is to addresses an important substantive problem: estimating the size of populations most at-risk for HIV/AIDS. We use our survey to estimate the number of injecting drug users, men who have sex with men, female sex workers, and clients of female sex works in Rwanda. This empirical study builds directly upon the methods we developed in Chapters 2 and 3. We use the basic scale-up estimator to produce size estimates, and then use the sensitivity framework introduced in Chapter 3 to illustrate how the estimates are sensitive to the basic scale-up assumptions.

4.1 Introduction

Many important problems in science and policy require estimates of the sizes of hidden populations. For example, in order to respond to the human immunodeficiency virus (HIV) epidemic, scientists and policymakers need information about the sizes of key populations at risk for the disease: men who have sex with men, female sex workers, male clients of female sex workers, and people who inject drugs (UNAIDS, 2010).

These size estimates are critical to designing HIV services, evaluating the outcomes of HIV-related interventions, and predicting the future course of the HIV epidemic (UNAIDS, 2010; UNAIDS and WHO, 2011; RBC/IHDPC et al., 2012).

Unfortunately, traditional survey techniques are not well-suited to making accurate size estimates for hidden populations (UNAIDS and WHO, 2011; Bernard et al., 2010). In Chapter 3, we discussed a promising alternative called the network scale-up method, which is based on the idea that ordinary people have embedded within their personal networks information that can be used to estimate the size of hidden populations, if that information can be properly collected, aggregated, and adjusted (Bernard et al., 2010; Killworth et al., 1998b; Bernard et al., 1989). The network scale-up method has many advantages over other size estimation procedures, as has been described in detail elsewhere (Bernard et al., 2010; Salganik et al., 2011a): it can easily be standardized across time and location; it can be used to concurrently produce estimates for several different hidden populations; it can be partially self-validating because it can easily be applied to populations of known size; it can produce estimates at the national, regional or city level; it does not require respondents to report that they themselves are members of a stigmatized group; and it can be easily added to existing survey instruments. Because of its appeal, the network scale-up method has been used in studies around the world, and its use is accelerating (see Table 3.1).

The study described in this chapter makes three main contributions. First, we show that the type of network that respondents are asked to report about is a potentially crucial parameter that researchers are free to vary. Previous scale-up studies have asked respondents to report about their network of acquaintances, but there is no reason to think that this particular network will produce optimal, or even near-optimal estimates. In order to explore this important issue, we conducted a large, nationally representative survey experiment in Rwanda. By i) randomizing respondents to report about one of two different types of personal network and ii) estimating

quantities whose true size is known, we are able to show that, somewhat surprisingly, asking for information about fewer people can produce more accurate size estimates. The methodology that we have developed means that our study design can be replicated and extended in other settings, enabling cumulative improvement of the scale-up method over time.

Second, we use the results of our survey to produce substantively important hidden population size estimates for four key populations at increased risk for HIV in Rwanda: female sex workers, male clients of sex workers, men who have sex with men, and people who inject drugs. Sub-Saharan Africa is central to global efforts to combat HIV, and Rwanda is an important example of the challenge that many countries in the region face: in 2010, the national prevalence of HIV among adults was estimated to be 3%, yet little was known about the sizes of key populations at increased risk for HIV (UNAIDS, 2013). Our study demonstrates that the scale-up method is feasible in Rwanda, and likely feasible in developing countries all over the world.

Third, we build on the results of Chapters 2 and 3 to introduce a new framework for sensitivity analysis that enables researchers to quantitatively assess how basic scale-up estimates are impacted by many potential sources of bias. This framework for sensitivity analysis has been derived from first principles, meaning that it accounts for all of the sources of bias that have been previously discussed (but not resolved) in the scale-up literature; our framework also identifies new potential sources of bias not previously considered. This sensitivity framework can be used in future scale-up studies, whether or not they have an embedded experiment. We illustrate the framework by applying it to our estimates for the size of four key populations at increased risk for HIV in Rwanda.

4.2 Methods

4.2.1 The network scale-up method

Network scale-up estimates come from survey data collected from a representative sample of the general population (see Chapter 3 and Bernard et al. (2010)). Respondents are asked about their social connections to people in several hidden populations (e.g., "How many female sex workers do you know?") and their connections to people in groups of known size (e.g., "How many teachers do you know?"). The responses are combined to produce estimates of the sizes of the hidden population using the basic scale-up estimator:

$$\widehat{N}_H = N \frac{\sum_{i \in F} y_{i,H}/\pi_i}{\sum_{i \in F} \widehat{d}_{i,U}/\pi_i}, \tag{4.1}$$

where $\widehat{N}_H$ is the estimated size of the hidden population, N is the size of the total population, π_i is the probability of inclusion for the ith survey respondent, $y_{i,H}$ is the number of members of the hidden population respondent i reports being connected to (e.g., the number of sex workers i reports knowing), and $\widehat{d}_{i,U}$ is the estimated personal network size of respondent i (Section 3.2; Bernard et al., 2010; Killworth et al., 1998a; Feehan and Salganik, 2015). The personal network size estimates $\widehat{d}_{i,U}$ come from the known population estimator, which is based on the number of connections that respondents report to the groups of known size:

$$\widehat{d}_{i,U} = \frac{\sum_j y_{i,j}/\pi_i}{\sum_j N_j}, \tag{4.2}$$

where $y_{i,j}$ is the number of people respondent i reports knowing in population j and N_j is the total size of known population j (Appendix B.2.4; Killworth et al., 1998b). The estimators in Equations 4.1 and 4.2 will be consistent and unbiased under conditions described in Section 3.2, Appendix B.2.4, and Appendix B.3.7. Intuitively, the basic

scale-up estimator (Equation 4.1) is like a sample proportion, but rather than being taken over all respondents, it is taken over all the members of respondents' personal networks.

4.2.2 Tie definition

In order to use the scale-up method in practice, researchers need to define what it means to be socially connected to another person. These connections are called ties in the social networks literature (Wasserman and Faust, 1994), and can also be referred to as edges or links.

Almost all previous scale-up surveys have followed the tie definition used in the original scale-up study (Bernard et al., 1989): two people are considered connected if they both know each other by sight and by name and have been in contact in the past two years (see Table 4.1). However, there is no particular reason to believe that this widely used tie definition leads to the best possible estimates. In fact, we conjecture that choosing a tie definition from a continuum of possible definitions of varying tie strength (Granovetter, 1973) likely induces a trade-off between the quality and quantity of information collected from each respondent.

In order to understand this hypothesized trade-off between the quality and quantity of information collected from each respondent more carefully, it is useful to embed it within the total survey error framework. As with other survey-based methods, error in scale-up estimates can be decomposed into two broad categories: sampling error and non-sampling error (Groves and Lyberg, 2010). Sampling error arises from the fact researchers interview only a sample of people rather than an entire population, and non-sampling error arises from all the other sources of error such as incompleteness in the sampling frame, mistakes in data processing, and inaccuracies in responses (see Groves and Lyberg (2010) for a more detailed review). Previous network research leads us to predict that both sampling error and non-sampling error will depend on

Tie Definitions in Survey Experiment	
Acquaintance ($n = 2,236$)	Meal ($n = 2,433$)
• people of all ages who live in Rwanda	• people of all ages who live in Rwanda
• people the respondent knows, by sight AND name, and who also know the respondent by sight and name	• people the respondent knows, by sight AND name, and who also know the respondent by sight and name
• *people the respondent has had some contact with – either in person, over the phone, or on the computer in the previous 12 months*	• *people the respondent has shared a meal or drink with in the past 12 months, including family members, friends, co-workers, or neighbors, as well as meals or drinks taken at any location, such as at home, at work, or in a restaurant.*

Table 4.1: The two definitions of a personal network connection (also called a tie) that were used in the Rwanda survey. All of the conditions need to be satisfied in order for the respondent to consider someone a member of her network. We randomly assigned one of these two definitions to each household in our sample. The acquaintance definition is modeled after the one used in almost all previous scale-up studies (Bernard et al., 2010); however, in this version we asked about people the respondent has had contact with in the past 12 months, instead of 24 months.

which tie definition researchers choose for their survey. One set of findings suggests that non-sampling error will vary by tie strength because people have more accurate information about their strong ties than their weak ties (Marsden, 1990; Marin, 2004; Brewer, 2000; Brewer and Webster, 1999; Sudman, 1988). A second set of findings suggests that sampling error will vary by tie strength because people have more weak ties than strong ties (Goel et al., 2010; DiPrete et al., 2011). Therefore, unless these two forces completely balance, total survey error will be a function of tie strength (see Figure 4.1 for a possible example, and see Appendix C.2 for a formal model that leads to a trade-off between quantity and quality).

The possible relationship between tie strength and total error means that by using other tie definitions, researchers can potentially get more accurate estimates at no additional cost. We assess this possibility empirically by conducting a survey experiment to compare the accuracy of estimates made using two different tie definitions.

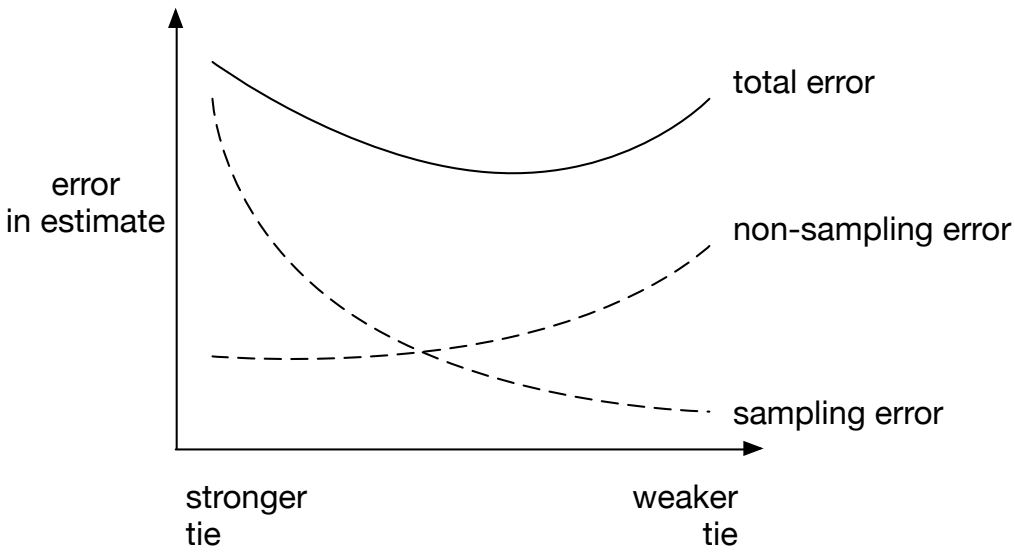


Figure 4.1: Hypothesized relationship between tie strength and error. Previous social network research suggests that non-sampling error and sampling error will both likely be functions of tie strength. This means that total error will also be a function of tie strength. The curves in this figure illustrate one possible relationship between tie strength and errors; other situations could lead to different trade-offs.

4.2.3 Survey experiment for evaluating tie definitions

In our survey experiment, each household was randomly assigned to one of two experimental arms. The questionnaire used in each arm was identical except for the tie definition that was used (Table 4.1). The first arm's tie definition, which we call the *acquaintance* definition, is modeled after the one used by all previous scale-up studies (Bernard et al., 2010). The second arm's tie definition, which we call the *meal* definition, has never been used before. We designed the meal definition, in consultation with Rwandan researchers and policymakers, so that i) it would represent a stronger tie than the acquaintance definition; and ii) it would be nested in the acquaintance definition, meaning that anyone who is connected to a respondent under the meal definition is also, by definition, connected to the respondent under the acquaintance definition (but not vice-versa). We predict that using the meal definition we will

learn about fewer people in each interview, but that the information we receive will produce better size estimates.

4.2.4 Internal consistency checks for evaluating tie definitions

The information collected about ties to groups of known size (Table 4.2) makes it possible to check the accuracy of the scale-up estimates for these groups of known size (Killworth et al., 1998a), a practice known in the scale-up literature as internal consistency checks (Bernard et al., 2010; Killworth et al., 1998a). In order to conduct an internal consistency check, for each group of known size we i) pretend its size is not known; ii) use the remainder of the groups of known size to estimate the respondents' personal network sizes (Equation 4.2); and iii) apply the scale-up method to estimate the size of the held-out group (Equation 4.1). These internal consistency checks provide us a natural method for deciding which of the tie definitions leads to more accurate estimates, at least for the groups of known size.

4.2.5 Linear blending for size estimates

Our survey produces two estimates for the size of each key population, one from each survey arm. However, policymakers typically require a single consensus estimate. Therefore, in Appendix C.4 we formally derive a linear blending technique to combine the results from both arms into a single consensus estimate, much like a meta-analysis (Hedges and Olkin, 1985). We consider all possible combinations of the estimates with the form,

$$\widehat{N}_H = w\widehat{N}_{\text{meal}} + (1-w)\widehat{N}_{\text{acq}}. \tag{4.3}$$

Group name	Size	Source
Priests	1,004	Catholic Church
Nurses or Doctors	7,807	Ministry of Health
Twahirwa	10,420	ID database
Mukandekezi	10,520	ID database
Nyiraneza	21,705	ID database
Male Community Health Worker	22,000	Ministry of Health
Ndayambaje	22,724	ID database
Murekatete	30,531	ID database
Nsengimana	32,528	ID database
Mukandayisenga	35,055	ID database
Widowers	36,147	RDHS (05, 07, 10)
Ndagijimana	37,375	ID database
Bizimana	38,497	ID database
Nyirahabimana	42,727	ID database
Teachers	47,745	Ministry of Educ.
Nsabimana	48,560	ID database
Divorced Men	50,698	RDHS (05, 07, 10)
Mukamana	51,449	ID database
Incarcerated people	68,000	ICRC 2010 report
Women who smoke	119,438	RDHS (05)
Muslim	195,449	RDHS (05, 07, 10)
Women who gave birth in the last 12 mo.	256,164	RDHS (10)

Table 4.2: The known populations used to estimate network sizes. RDHS denotes the Rwanda Demographic and Health Survey from the years indicated in parentheses; ID database denotes groups of names from the national identity card database; ICRC denotes the International Committee of the Red Cross.

Appendix C.4 proves that, under the assumption that the estimators from each experimental arm are unbiased and uncorrelated, the optimal linear blending weight is:

$$w = \frac{\sigma^2_{\text{acq}}}{\sigma^2_{\text{acq}} + \sigma^2_{\text{meal}}}, \tag{4.4}$$

where σ^2_{acq} is the sampling variance for the acquaintance definition estimate and σ^2_{meal} is the sampling variance for the meal definition estimate. In words, Equation 4.4 says that the larger the variance of the acquaintance definition's estimate relative to the variance of the meal definition's estimate, the more weight the meal definition

estimate gets in the blended estimate. In practice, we use Equation 4.4 by plugging in sample-based estimates of σ^2_{acq} and σ^2_{meal}.

4.2.6 Framework for sensitivity analysis

There is a large literature describing many potential sources of bias in the basic scale-up estimator (Bernard et al., 2010; Feehan and Salganik, 2015; Killworth et al., 2003; Kadushin et al., 2006; Maltiel et al., 2015; McCormick and Zheng, 2007; Killworth et al., 2006; Johnsen et al., 1995; Zheng et al., 2006). Therefore, following a long tradition in epidemiology (Greenland, 1996, 2003, 2005), we develop a framework for sensitivity analysis. Our framework enables us to precisely quantify the effect of a variety of biases on our estimates, and it allows other researchers to calculate estimates under different assumptions about possible biases.

Using results developed in Section 3.2 (Equation 3.10), we consider the multiplicative bias of the basic scale-up estimator:

$$N_H = \alpha \ \widehat{N}_H. \tag{4.5}$$

where α is an overall adjustment factor that captures all biases in the basic scale-up estimator. Section 3.2 and Appendix C.5 show that α can be written as the product of four quantities:

$$\alpha = \left(\frac{\eta_F}{\tau_F}\right)\left(\frac{1}{\phi_F \delta_F}\right) \tag{4.6}$$

Each of the four adjustment factors is defined precisely in Section 3.2 and Appendix B.1, and each adjustment factor could potentially be estimated empirically; together the four adjustment factors account for the non-sampling errors discussed in the scale-up literature. Previous studies that report estimates from the basic scale-up

estimators have been implicitly assuming that the product of these four quantities is 1.

The two reporting terms (η_F and τ_F) summarize the accuracy of respondents' reports. The precision, η_F, quantifies respondents' tendency to give false positive reports (e.g., if respondents report that some people they are connected to are sex workers when they are not). On the other hand, the true positive rate, τ_F, quantifies respondents' tendency to give true positive reports (e.g., if respondents are connected to sex workers and able to report this correctly). In general, we would expect the true positive rate to be less than 1 because members of hidden populations might attempt to keep this information secret from other people.

The structural terms (δ_F and ϕ_F) summarize the differences between the hidden population, the survey respondents, and the entire population. The degree ratio, δ_F, quantifies whether hidden population members tend to have smaller personal networks than the survey respondents (e.g. if the average sex worker has fewer connections to adults than the average adult). Finally, the frame ratio, ϕ_F quantifies whether survey respondents (e.g., adults) tend to have larger personal networks than the entire population (e.g. adults and children).

In addition to assessing the sensitivity of our estimates to specific biases, our sensitivity framework (Equation 4.6) ensures that our estimates can be improved over time as more is learned. If additional studies are conducted to estimate any of the quantities in Equation 4.6 (η_F, ϕ_F, δ_F, and τ_F) (Salganik et al., 2011a; Feehan and Salganik, 2015; Salganik et al., 2011b), these new estimates can be combined with the results from our study to produce improved hidden population size estimates that are less dependent on assumptions about unobserved quantities (see Appendix C.5).

4.2.7 Data collection and processing

To conduct our survey experiment, we collected original data using the same survey infrastructure as the 2010-2011 Rwanda Demographic and Health Survey (DHS) (RBC/IHDPC et al., 2012). The DHS program is one of the largest and most widely used sources of information about international health: since 1984, the DHS program has conducted over 260 surveys in more than 85 developing countries, and it has well established protocols for developing questionnaires, training interviewers, supervising interviews, processing data and overall supervision of household-based surveys (Fabic et al., 2012; Corsi et al., 2012). By using the DHS infrastructure, we ensure that our research design can be used in face-to-face surveys in developing countries around the world.

Our sample was drawn from the preparatory frame constructed for the 2012 Rwanda Census, which contained a complete list of 14,837 villages, which are the smallest administrative units in the country. We used a stratified, two-stage cluster design with these villages as the primary sampling units.

We conducted interviews with 4,669 respondents from 2,125 households in 130 villages (household response rate: 99%; individual response rate: 97%) (RBC/IHDPC et al., 2012). All household members over the age of 15 were interviewed in each selected household. Eight survey teams – each consisting of one supervisor, two male interviewers, and two female interviewers – conducted these interviews between June and August of 2011. Upon arriving in each sampled village, the survey team first updated the list of households. A number was assigned to each household in the updated listing, and the supervisor used a randomly pre-generated table to select households and assign them to tie definitions. Balance checks suggest that our randomization procedure was implemented according to the study design (Appendix C.1.2). All interviews were conducted in Kinyarwanda and were face-to-face in the respondent's household. All responses were recorded on paper forms and data

were entered twice using CSPro. Table 4.2 shows the definitions and data sources of the 22 known populations that we used to estimate the size of respondents' personal networks (Equation 4.2). The full details of the sampling plan and the survey instrument are described elsewhere (RBC/IHDPC et al., 2012).

When analyzing our data, consistent with common scale-up practice, we truncated extreme outliers by top coding all responses about connections to groups of known size and hidden populations at 30, affecting 0.2% of responses (Zheng et al., 2006). To produce estimates of the sampling uncertainty for all of the estimates we report, we the rescaled bootstrap technique of Rao et al. (1992) to account for the complex sample design (Appendix B.6; Feehan and Salganik, 2015). Unless otherwise noted, all estimates were produced using weights to account for the complex sample design. All of our calculations were done in R (36) using the following packages: networkreporting (Feehan and Salganik, 2014), plyr (Wickham, 2011), ggplot2 (Wickham, 2009), stringr (Wickham, 2012), stargazer (Hlavac, 2014), RItools (Bowers et al., 2010), and car (Fox and Weisberg, 2011). Our dataset and a copy of the survey instrument are freely available to download (RBC/IHDPC et al., 2012), and code to replicate our analyses will be made available online.

4.3 Results

For each of the 22 groups of known size, respondents to the meal definition reported fewer average connections, indicating that our survey experiment impacted responses (Figure C.1). Using these reports to estimate the average size of respondents' personal networks (Equation 4.2) shows that meal definition elicits information about 60% fewer people than acquaintance definition: the estimated mean network size for the meal definition is 108 (95% Confidence Interval (CI): 99, 118) while for the acquaintance definition it is 251 (95% CI: 237, 267) (Figure 4.2).

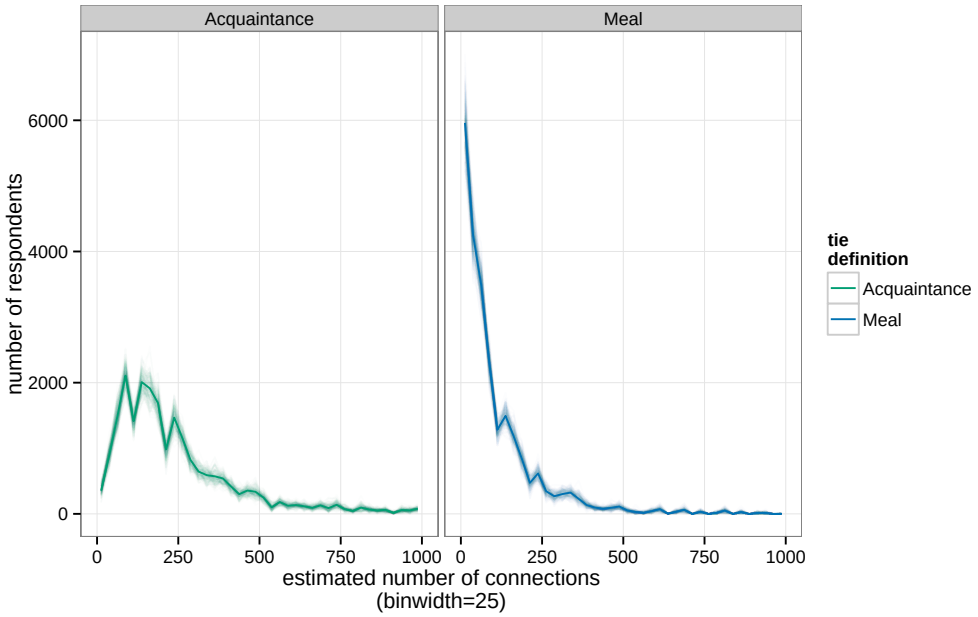


Figure 4.2: Estimates of the network size distributions produced from the acquaintance tie definition (on the left in green) and from the meal definition (on the right in blue). Results from each bootstrap resample are plotted with partial transparency, meaning that we are more confident about the distributions in regions that are more solidly colored. The estimated mean network size for the acquaintance tie definition is 251 (95% CI: 237, 267) while for the meal definition it is 108 (95% CI: 99, 118).

4.3.1 Internal consistency checks

The internal consistency checks, which compare the performance of each tie definition when estimating the size of known populations, show that in most cases, the estimates are close to the true values (Figure 4.3). However, there are a few exceptions: the most notable ones are teachers (whose size is over-estimated), Muslims (whose size is underestimated), and women who gave birth in the past 12 months (whose size is underestimated). On average, across the 22 groups of known size, the estimates from the meal definition had lower mean squared error than the estimates from the acquaintance definition (two-sided $P < 0.001$), a result that is robust to the specific error metric used (Figure C.4).

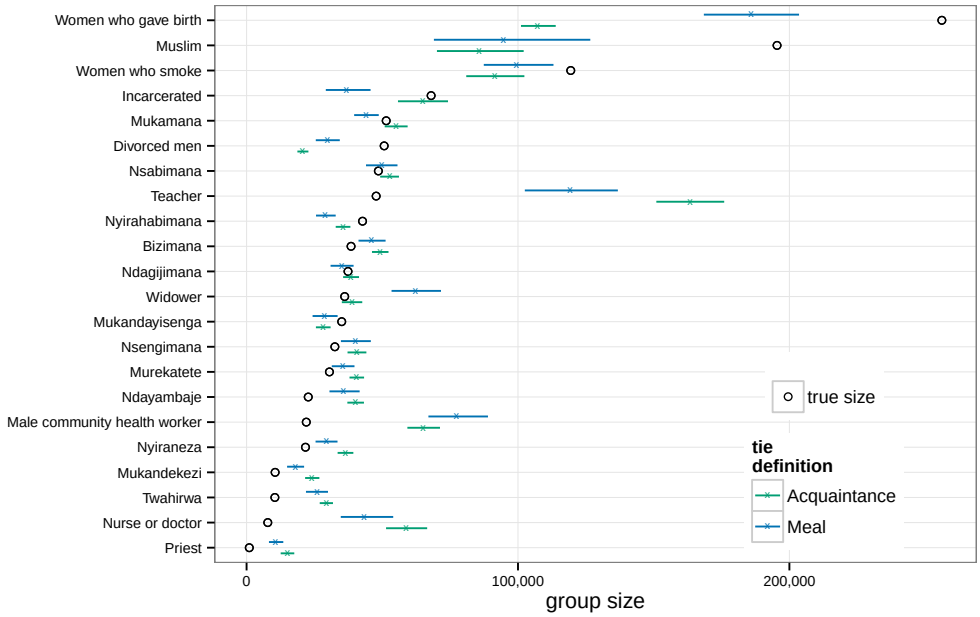


Figure 4.3: Internal consistency check for the 22 groups of known size (Table 4.2). For most of the groups of known size, the hold-out estimates from both tie definitions are close to the true value. However, the uncertainty intervals – which only capture sampling error – do not include the truth as often as would be expected, a pattern consistent with other scale-up studies (Chapter 3; Feehan and Salganik, 2015).

4.3.2 Size estimates for hidden populations and sensitivity analysis

Figure 4.4 shows the estimated sizes of four key populations at risk for HIV in Rwanda: male clients of sex workers, female sex workers, men who have sex with men, and people who inject drugs. As described in more detail in the discussion, these estimates are consistently higher than earlier estimates in Rwanda, but they are generally lower than international benchmarks.

In the absence of empirical evidence about the magnitude of possible biases in the basic scale-up estimator, Figure 4.4 assumes that the estimates from each tie definition are unbiased ($\alpha = 1$). In Figure 4.5, we relax this assumption using our framework for sensitivity analysis to show how the blended estimate depends on the assumed multiplicative bias (α) for each hidden population. For example, if we

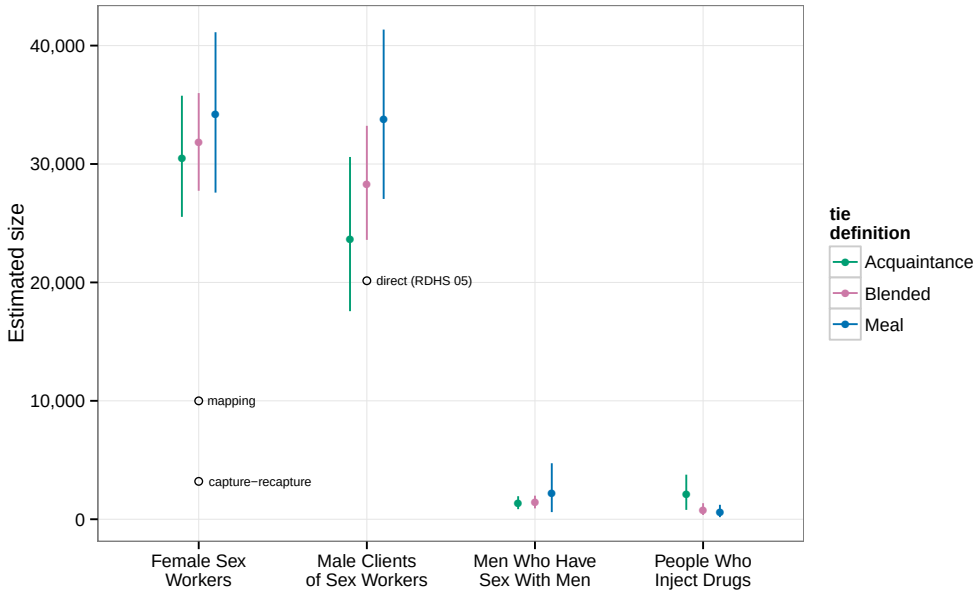


Figure 4.4: Estimated hidden population size using each tie definition, and the blended estimator. Estimates in green (on the left) are for the acquaintance definition, estimates in blue (on the right) are for the meal definition, and estimates in red (in the center) are both acquaintance and meal blended together (using Equations 4.3 and 4.4). Earlier estimates from Rwanda are also shown.

consider the estimate for the number of female sex workers, Figure 4.5 shows that if the meal estimator was unbiased ($\alpha = 1$), but the acquaintance estimator tended to be too small ($\alpha = 1.5$) then the blended size estimate should be about 40,000, which is higher than the estimate under the assumption that both arms are unbiased. In general, Figure 4.5 shows that higher values of α lead to higher estimates. However, the exact nature of the relationship between assumed values for α and size estimates depends on the sampling variance of the estimator from each arm. Appendix C.5 shows how researchers can use our sensitivity framework to combine our data with any set of assumptions about η_F, ϕ_F, δ_F, and τ_F to produce adjusted estimates.

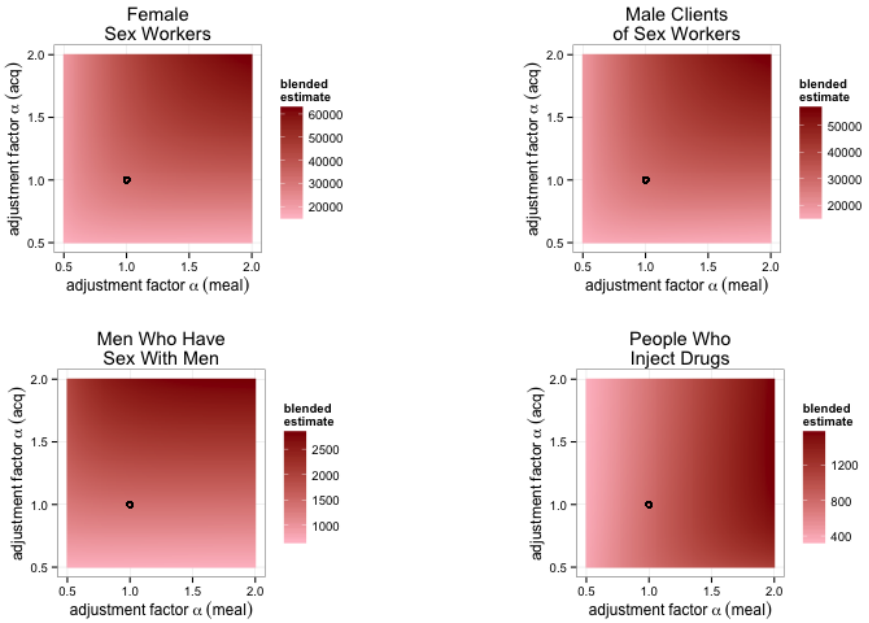


Figure 4.5: Sensitivity of blended estimates for each hidden population. Each panel shows the range of size estimates for one of four populations at-risk for HIV in Rwanda. Adjustment factor values (α, Equation 4.6) for the estimator from the meal definition (x axis) and for the acquaintance estimator (y axis) combine to produce a blended estimate. The circle shows the estimate in Figure 4.4, which assumes that both the meal and the acquaintance definitions produce unbiased estimates.

4.4 Discussion

Our survey experiment demonstrates that the tie definition that respondents are asked to report about is a potentially critical parameter that researchers can vary. There appears to be a trade-off between quantity and quality of network reports: for the groups of known size, the internal consistency checks reveal that the meal definition outperformed the acquaintance definition that has been used in all previous scale-up studies. These results show that learning about fewer people can, somewhat surprisingly, lead to more accurate estimates. Our study, therefore, makes the prediction that future scale-up studies will produce more accurate estimates using stronger tie definitions.

We also used our survey to estimate the sizes of four populations at-risk for HIV in Rwanda. Our blended scale-up estimates are consistently higher than earlier esti-

mates from Rwanda made using other methods (Figure 4.4). For female sex workers, our estimate is higher than two earlier estimates, one using mapping and one using capture-recapture (RBC/IHDPC et al., 2012). Our estimate for the number of male clients of sex workers is higher than an earlier direct estimate (RBC/IHDPC et al., 2012). There are no previous estimates for the number of people who inject drugs and men who have sex with men in Rwanda. On the other hand, our estimates are comparable to, or lower than, benchmark estimates from the Joint United Nations Programme on HIV/AIDS (UNAIDS) (UNAIDS, 2014), which were derived from published literature on population size estimates from around the world. Our estimated number of female sex workers is within the range given by the UNAIDS benchmarks. However, our estimates for the number of male clients of sex workers and the number of injecting drug users are both lower than the UNAIDS benchmarks. There are no UNAIDS benchmarks available for men who have sex with men in Africa. Additional details about these comparisons are presented in Appendix C.3.

Finally, we introduced a framework for sensitivity analysis that enables scale-up researchers to assess how several different potential sources of bias will impact size estimates. Our framework is derived from first principles, enabling it to account for all of the sources of potential bias, only some of which were previously discussed in the scale-up literature. We demonstrate the framework by assessing the sensitivity of our basic scale-up estimates for the size of four populations at-risk for HIV in Rwanda. Appendix C.5 has a step-by-step guide to using the framework for sensitivity analysis, which can be applied in other studies whether or not a tie definition experiment has been conducted.

4.4.1 Limitations

Our study has several important limitations. The fact that the meal definition worked better than the acquaintance definition on average across the 22 groups of known size

in Rwanda does not guarantee that it will work better for hidden populations in Rwanda or hidden populations in other countries.

Further, we do not claim that the meal definition is optimal. In fact, we hope future research explores a wider range of tie definitions with the goal of developing an empirical understanding of the general relationship between tie strength and total error. We recommend that future studies explore tie definitions that come from a parameterized family that is sufficiently flexible for use in many different countries. For example, the meal and acquaintance tie definitions are members of a two-parameter family, where one parameter is an interaction type (e.g., had some contact, shared a meal or drink) and one parameter is a time window (e.g., in the past year). By focusing on a parameterized family of tie definitions, it will be possible to compare and combine results from different studies, enabling knowledge about the method to build up as more scale-up studies are conducted.

Empirically, our blended estimates (Figure 4.4) depend on the assumption that the basic scale-up estimator was unbiased in this setting. However, our framework for sensitivity analysis shows that different plausible assumptions about the bias can lead to a wide range of possible estimates (Figure 4.5). Therefore, we recommend that future studies attempt to measure these possible biases directly (see e.g., Salganik et al. (2011b); Maghsoudi et al. (2014)) so that estimates are less dependent on assumptions.

4.4.2 Conclusion

Estimating the size of hidden populations such as sex workers and people who inject drugs is a critical problem in many settings. The network scale-up method is a promising approach, and in this paper we show that it is more flexible and potentially more accurate than had been previously realized. Further, the methodology that we developed and deployed—a survey experiment with blending (Figure 4.6)—provides

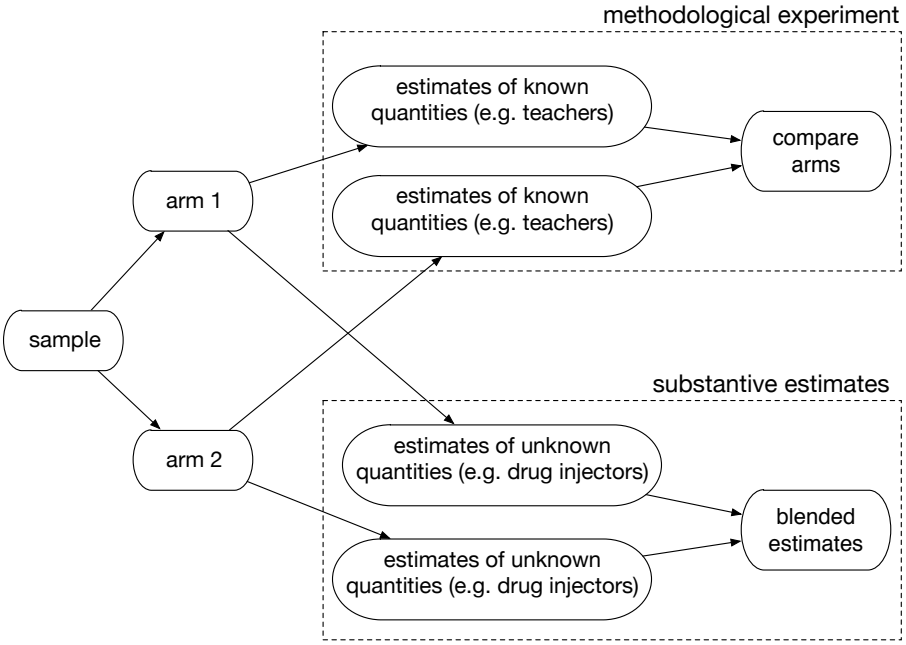


Figure 4.6: Survey experiment with blending. Respondents from the sample are randomly placed in one of two arms, and a different variation of the scale-up method (in our case, a different tie definition) is used in each arm. Researchers can learn about which experimental condition performs better by comparing estimates of the known quantities to their actual values. Researchers can produce substantive size estimates for hidden populations by blending the results from each arm together.

a general template that future scale-up researchers can use. If this design becomes standard, future studies will provide both estimates about specific hidden populations and more general insights that will lead to cumulative methodological improvement.

Chapter 5

The network survival estimator: evidence from Rwanda

5.1 Introduction and overview

Adult mortality plays a fundamental role in important questions about population health and well-being. For example, how healthy is the population of Rwanda compared to the population of the United States? Which country has a bigger gap between the health of the rich and the poor? Should public health officials in Kigali worry more about HIV or smoking? Counting adult deaths is a big part of the answer to these questions and to many others: researchers and policymakers need measurements of adult death rates to compute life expectancy; to understand health differences between the rich and the poor; to assess the health impact of major events like the spread of HIV and smoking; and to set budgets and make plans for programs like Social Security, health insurance, and pensions.

In the United States, information about adult deaths comes from the vital registration system, which ensures that a death certificate gets filled out every time someone dies. The information from these death certificates yields high quality *gold*

standard measurements of US death rates by age and sex. Rwanda, on the other hand, is a victim of the 'scandal of invisibility': it has no system that reliably produces death certificates, meaning that most adults just disappear without ever having their deaths formally recorded (Setel et al., 2007; Mikkelsen et al., 2015). Unfortunately, this scandal goes well beyond Rwanda: AbouZahr et al. (2015b) reports that an estimated two-thirds of worldwide deaths are never recorded in any way, and that about half of the World Health Organization member states "obtain either no data for mortality and cause of death, or obtain data of of such poor quality that they are of little value for public health policy and planning."

As a result of this lack of adequate mortality data, the figures for life expectancy at birth that are compiled by international agencies are essentially model-based extrapolations for almost all of Sub-Saharan Africa and much of Asia (Reniers et al., 2011). The long-term solution to the scandal of invisibility is to develop effective vital registration systems capable of recording deaths in every country around the world (AbouZahr et al., 2015b). However, progress on this front has been very slow: Mikkelsen et al. (2015) estimates that between 2000 and 2012, the number of deaths registered worldwide went from 36% to 38%. Researchers have worked on the problem of estimating adult death rates in countries that lack high-quality death registration systems for decades. In the meantime, critical questions still go unanswered because accurate and reliable strategies for directly estimating adult death rates are not available in many of the world's poorest countries.

Our study aims to help by introducing a new, social network-based method for estimating adult death rates. This new approach, the *network survival method*, was designed to overcome many of the measurement challenges that have hindered adult mortality estimation to date. The new method blends together insights from two separate literatures: demographic research on estimating death rates from sibling reports, and sociological and anthropological research on collecting social network information

from surveys. The result of this blending is that the network survival method enables researchers to estimate national-level adult death rates from information collected about deaths among survey respondents' personal network members.

Any approach to estimating adult death rates will have to make assumptions, and the network survival method is no exception. Fortunately, we derive the new estimator from first principles, which enables us to precisely articulate the conditions that must hold for it to produce consistent and essentially unbiased estimates. Moreover, our derivation reveals that the network survival approach can be used with a wide variety of different types of personal networks. This flexibility could permit future researchers to vary which personal network survey respondents are asked to report about, systematically investigating which types of social network contain the most accurate and useful information about adult mortality.

The remainder of this chapter begins by reviewing survey-based adult mortality estimation, paying particular attention to the current state of the art: the direct sibling survival estimator. Next, we introduce the network survival method and explain how the new estimator can be derived. We then turn to a discussion of the first empirical results from the new method. To test the network survival estimator, we designed and conducted a nationally-representative survey of 4,669 people in Rwanda. To illustrate the generality of the new method, we randomized survey respondents to report about one of two different personal networks; we then produce two different sets of estimated age-, sex-, and year-specific adult death rates using the network survival method. Next, we compare the estimates produced by the network survival method to other adult mortality estimates that have been produced for Rwanda. Finally, we conclude with a discussion of the results and an outline of the directions that future work could take. Several appendices contain additional information and technical details.

5.2 Background

5.2.1 The problem: estimating death rates

The death rate describes the number of deaths that occur in a group, relative to the group's exposure to the possibility of dying. Mathematically, for a demographic group α (for example, women aged 45-49 in 2011), the death rate can be written

$$M_\alpha = \frac{O_\alpha}{E_\alpha}, \tag{5.1}$$

where O_α is the number of deaths and E_α is the amount of exposure to demographic group α. Death rates are a type of *occurrence-exposure* rate, which is why we use O_α for deaths (occurrences), and E_α for exposure.

Adult death rates are difficult to estimate from a survey for two main reasons (Timaeus, 1991). First, surveys typically ask respondents to report about themselves; for example, on a poll conducted before an election, respondents might be asked which candidate they are planning to vote for. For deaths, this approach is not possible: people who have died cannot be interviewed. Second, adult deaths are quite rare. Rare events are difficult to estimate using standard survey techniques because they require very large samples to yield estimates that are precise enough to be useful (Kalton and Anderson, 1986). Any survey-based approach to estimating adult death rates will have to find a way to overcome those two primary obstacles.

If deaths rates are difficult to estimate from surveys, why focus on survey-based approaches at all? In countries that lack good vital registration data, fertility rates and child mortality were once as poorly understood as adult mortality is now. But today, even the world's poorest countries have high-quality estimates of fertility and child mortality rates. This change largely resulted from the development of new methods that allowed researchers to use household surveys to collect information

from women about their births and about their children's deaths (Hill and Choi, 2004; Timaeus, 1991). These regular, survey-based measurement efforts began at a large scale with the World Fertility Survey Program, and survive today in the massive, internationally coordinated Demographic and Health Survey program (Hill et al., 2007; Corsi et al., 2012; Fabic et al., 2012). This success in using the science of sampling and survey research to measure births and child deaths makes a powerful case for developing survey-based approaches to estimating adult mortality: in many settings that lack death registration data, high-quality household surveys are already being regularly conducted in order to provide up-to-date estimates of fertility and child mortality. If researchers can develop and validate methods for estimating adult death rates from these surveys, then it may be possible to provide up-to-date estimates of adult mortality as well.

Survey-based approaches are not the only way to try and estimate adult death rates; there is a large literature on alternative techniques, based on prospective or cohort designs, incomplete sources of death certificates, one or many censuses, and historical records. In our overview here, we focus on survey-based techniques, since they are most relevant to our new estimator. In practice, Hill et al. (2007) observes that, because data on adult mortality are so rare, these different mortality estimation approaches tend to complement, rather than compete with one another. Other authors have provided more complete overviews of mortality estimation (see, for example, Hill et al., 1983; Timaeus, 1991; Hill, 2000, 2003; Gakidou et al., 2004; Hill et al., 2005; Bradshaw and Timaeus, 2006; Hill et al., 2007; Reniers et al., 2011; Moultrie et al., 2013; AbouZahr et al., 2015a).

5.2.2 Sibling survival

The state of the art in survey-based mortality estimation is the direct sibling survival method (Rutenberg and Sullivan, 1991), which is the most recent of several sibling

survival methods that researchers have studied and improved over many years. Appendix D.5.1 provides a detailed review; here, we focus on the current state of sibling survival research.

Direct sibling survival requires collecting *sibling histories*: detailed, individual-level information about survey respondents' siblings. Respondents are first asked to enumerate their siblings and then to provide information about each one, including the sibling's birthday, survival status, and date of death (where appropriate). Collecting this detailed information makes the survey relatively complex and time-consuming, when compared to the straightforward, aggregate reports required by the indirect sibling survival methods (see Appendix D.5.1). But, the more detailed information collected using the direct sibling survival method also allows researchers to estimate adult and maternal mortality using many fewer assumptions than the indirect approaches (Rutenberg and Sullivan, 1991; Stanton et al., 2000).

The direct sibling survival method seems like a promising way to overcome the two fundamental challenges in estimating death rates from surveys: since respondents report about their siblings, it is possible to learn about people who have died; and, since respondents typically have multiple siblings, each interview produces information about more than one person, increasing the effective size of the sample. Moreover, as a part of the Demographic and Health Survey (DHS) program, sibling histories have been collected in over 150 surveys from dozens of countries across the developing world[1] (Corsi et al., 2012; Fabic et al., 2012). Nonetheless, relatively few researchers have made use of this vast quantity of data to study adult mortality (Reniers et al., 2011; Gakidou et al., 2004). For example, despite the fact that very little is known about national-level adult mortality in Sub-Saharan Africa (Setel et al., 2007), only a small (but accelerating) number of studies have tried to use the sibling histories to construct estimates of recent trends in adult mortality there (Timaeus and Jasseh,

[1] `http://www.dhsprogram.com/What-We-Do/Survey-Search.cfm` (accessed April 14, 2015)

2004; Obermeyer et al., 2010; Rajaratnam et al., 2010; Reniers et al., 2011; Wang et al., 2013; Masquelier et al., 2014).

There are two reasons why the DHS sibling histories may have been relatively under-used. First, surveys with typical DHS sample sizes of 5,000 to 30,000 respondents (Corsi et al., 2012) cannot be used to produce timely direct estimates of age- and sex-specific death rates because the sampling variation from the direct sibling survival estimator is too large (Stanton et al., 2000; Timaeus and Jasseh, 2004; Hill et al., 2006). Instead, researchers have had to resort to a combination of pooling data across countries and across time, smoothing regressions, and model life tables to estimate adult mortality from DHS sibling histories (Timaeus and Jasseh, 2004; Obermeyer et al., 2010; Rajaratnam et al., 2010; Reniers et al., 2011; Wang et al., 2013; Masquelier et al., 2014). This need to smooth the raw data requires researchers to make several difficult-to-verify assumptions, reducing the appeal of producing estimates based on sampled data (Masquelier, 2013).

Second, the DHS sibling histories may be relatively under-used because there is considerable methodological debate about how they should be analyzed. Researchers have long discussed several potential biases likely to arise from estimating death rates by asking respondents to report about their siblings. Several common concerns have emerged: (i) there is no way to learn about sibships (sets of people who are siblings) that have no survivors left to be sampled by the survey and, more generally, sibships with more survivors are more likely to be sampled by the survey, potentially biasing estimates if sibship size and mortality are correlated (Gakidou and King, 2006; Trussell and Rodriguez, 1990; Graham et al., 1989; Masquelier, 2013; Reniers et al., 2011; Gakidou et al., 2004); (ii) there are many ways that respondents' reports about their siblings may not be accurate; for example, respondents may omit some siblings from their survey reports, and if the tendency to omit a sibling is correlated with the chances that the sibling is alive, then this may introduce bias into the

resulting estimates (Helleringer et al., 2014a,b, 2013; Merdad et al., 2013; Masquelier and Dutreuilh, 2014); (iii) the respondent is, by definition, alive, making it unclear whether the respondent's experience should be included or omitted from the death rate estimates (Reniers et al., 2011; Masquelier, 2013).

Debate about these potential methodological problems has not been resolved. For example, Gakidou and King (2006) proposed a solution to address the potential correlation between sibship size and mortality, but it has proven to be controversial in practice (Masquelier, 2013). Subsequent studies have therefore been divided: one group has applied the Gakidou-King selection bias adjustments (Kassebaum et al., 2014; Wang et al., 2013; Rajaratnam et al., 2010) while another has not (Reniers et al., 2011; Moultrie et al., 2013; Masquelier et al., 2014).

To conclude, the direct sibling survival method is a promising approach to overcoming the two main challenges that must be faced to estimate death rates from a survey: it enables researchers to learn about people who died, and it enables researchers to learn about more than one person from each interview. Unfortunately, in practice, the direct sibling survival method has two big disadvantages: first, it cannot typically be used to produce direct estimates of death rates because the sampling variation of direct estimates is too large; and, second, the sibling survival method is clouded by several potential sources of bias. It is not clear precisely what effect these potential biases might have on sibling survival estimates, or how these potential biases might interact with one another.

5.3 The network survival method

The network survival method is a new approach to estimating adult death rates from a survey. The method takes advantage of the fact that survey respondents in any population are embedded in many different types of personal networks, and

that respondents have information about other the people they are connected to through these personal networks. The network survival method can be seen as a generalization of the direct sibling survival method; in fact, our new data collection procedure was designed to combine attractive features of the sibling histories and a survey-based method for studying personal networks called scale-up (Killworth et al., 1998b; Bernard et al., 2010; Feehan and Salganik, 2015).

Network survival estimates are based on collecting two types of information about survey respondents' personal networks: first, respondents are asked about their connections to people who died (e.g., "How many people do you know who died in the previous 12 months?"). Similar to a sibling history, respondents are asked to enumerate each person who died, and to provide additional information about each one (for example, the deceased's age and sex). Second, respondents are asked about their connections to several different groups whose total size is known (e.g., "How many policemen do you know?" where the number of policemen is available from administrative records or estimated from a survey). This information about connections to groups of known size is used to estimate the total size of respondents' personal networks. These counts of connections that respondents report having to groups of known size are called *aggregate relational data* (McCormick et al., 2012b). Appendix D.7 has an example of a network survival survey instrument.

Asking survey respondents to report about the members of their personal networks helps to get around both of the major difficulties in estimating death rates from a survey: since respondents report about others, it is possible to learn about people who have died, even though the people who died cannot be interviewed directly. And, since respondents are asked to report about all of the people in their personal networks, researchers get information about much more than just one person from each interview, increasing the effective sample size.

Network survival estimates can be computed in two steps. The first step separately estimates each of the death rate's two components: the number of deaths and the exposure. The second step combines these two components of the death rate using a standard ratio estimator (Sarndal et al., 2003, chap. 5). In the remainder of this section, we turn to a more detailed description of how the network survival method estimates each of the two components of a death rate. Our focus will be on describing the main ideas behind the new estimator; Appendixes D.1, D.2, and D.3 has proofs and further technical details, including a precise description of the conditions that must hold in order for network survival estimates to be consistent and essentially unbiased.

5.3.1 Estimating the amount of exposure, E_α

The first part of a death rate is the amount of exposure in demographic group α (E_α). This quantity can be estimated with the size of the mid-period population in demographic group α. In the Rwanda results, below, we estimate this quantity directly using our survey's sampling design. Our estimates of exposure in Rwanda therefore make the assumption that the frame population is *complete*; that is, we assume that everyone in demographic group α in the general population is also in the frame population (see Appendix D.2.3). If researchers are not willing to make the assumption that their frame population is complete, then an alternate estimator can be used (Appendix D.2.3, Result D.2.5). In other situations, it may also be possible to use other sources of estimates for the exposure, such as a recent population census.

5.3.2 Estimating the number of deaths, O_α

The second part of a death rate is the number of deaths in demographic group α (O_α). Estimating this quantity is complex, because it requires us to take survey respondents' *reports* about deaths, and convert them into an estimate for the *number*

of deaths in the population. To make this conversion from reports into a population-level estimate, we use the network reporting framework (Chapter 2), which allows researchers to create new estimators based on survey respondents' reports about their personal network connections.

Figure 5.1 illustrates the main idea behind the network reporting framework. The left-hand panel, Figure 5.1(a), depicts individuals in a population who have been asked to report which of their personal network members have died in the past 12 months (of course, only living people can be interviewed). Each directed arrow $i \to j$ indicates that i reports that j has died. The right-hand panel, Figure 5.1(b) presents the same information as the left-hand one, but rearranged so that the people who report are on the left hand side, and the people who could be reported about are on the right-hand side (note that living people can both report and be reported about, since it can happen that a living person is erroneously reported as dead). Each report $i \to j$ now goes from the left-hand side to the right-hand side. Using the framework introduced in Chapter 2, we can derive a reporting identity (see also Feehan and Salganik, 2015):

$$\text{number of deaths} = \frac{\text{total number of reports about deaths}}{\text{average reports per death}}. \tag{5.2}$$

The identity in Equation 5.2 reveals that we can estimate the number of deaths from survey respondents' reports by estimating (i) the total number of reports about deaths that would result from interviewing everyone, and (ii) and the average number of reports per death. Intuitively, if we simply added up the number of reports about deaths, then we would overcount the number of deaths because each death can be reported more than once; to adjust for this overcounting, we must divide by the average number of reports per death. Note that Figure 5.1 depicts detailed individual reports $i \to j$, but in practice reports are not typically collected at that level of detail

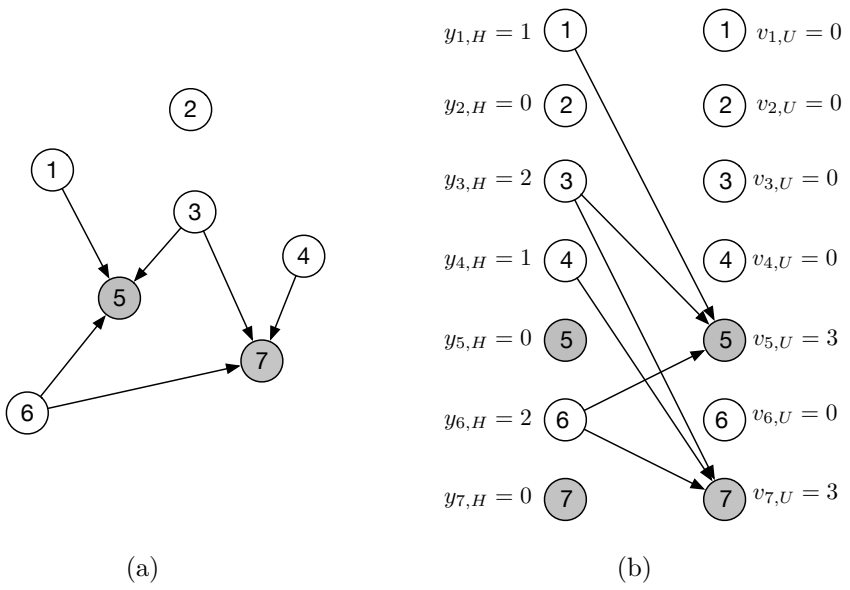


Figure 5.1: How the network survival estimator converts the number of reports about deaths into an estimate for the number of deaths. Panel (a) shows a population of 7 people, 2 of whom have died (shown in grey). A directed edge $i \to j$ indicates that i counts j as having died when answering the question "How many people do you know who have died in the past 12 months?" Panel (b) shows the same population, but redrawn so that each person now appears twice: as someone who reports, on the left, and as a someone who could be reported about, on the right. (Note that people who have died cannot report, since they cannot be interviewed.) This arrangement illustrates the fact that the total number deaths ($O_\alpha = 2$) must equal the total reported deaths ($y_{F,O_\alpha} = 6$) divided by the average number of reports per death ($\bar{v}_{O_\alpha,F} = 3$).

(i.e., we typically would know that person i reports one death, but not that the death was specifically person j). Fortunately, the identity in Equation 5.2 requires estimates of aggregate quantities, so this level of detail is not required. Our goal now is to provide an intuitive explanation for how the estimator works; Appendix D.2 has a more technical derivation of the new estimator and a full statement of all of the conditions that must hold for it to be consistent and essentially unbiased.

We need some notation in order to be able to describe several important quantities mathematically. Let F be the frame population, meaning the set of people who are eligible to be interviewed by our survey; in many cases, the frame population will be all adults over a particular age. Recall from Equation 5.1 that α is a particular demographic group (for example, women aged 45-49 in 2011), O_α is the number of

deaths in the demographic group, and E_α is the amount of exposure to the possibility of dying in the demographic group. Let y_{i,O_α} be the number of deaths in demographic group α reported by survey respondent i. Let $y_{F,O_\alpha} = \sum_{i \in F} y_{i,O_\alpha}$ be the number of deaths in demographic group α that would be reported if everyone in the frame population F was interviewed (ie, in a census). And, let $\bar{v}_{O_\alpha,F}$ be the *visibility* of deaths in demographic group α to the frame population F: the average number of times that each death would be reported if everyone in the frame population was interviewed.

Finally, we say that there are no *false positive* reports when $v_{i,F} = 0$ for all $i \notin O_\alpha$; in other words, if we assume there are no false positive reports, then we assume that respondents may omit some deaths, but they do not incorrectly report that someone died when in fact she did not. See Appendix D.2 for more information.

Using this notation, the identity in Equation 5.2 can be written

$$O_\alpha = \frac{y_{F,O_\alpha}}{v_{U,F}/O_\alpha}. \tag{5.3}$$

If we make the additional assumption that there are no false positive reports, then $v_{U,F} = v_{O_\alpha,F}$ and we can re-write Equation 5.3 as

$$O_\alpha = \frac{y_{F,O_\alpha}}{\bar{v}_{O_\alpha,F}}. \tag{5.4}$$

The network survival estimate for the number of deaths in demographic group α (O_α) is based on the identity in Equation 5.4. To estimate the numerator of Equation 5.4, y_{F,O_α}, we follow standard survey methodology and propose:

$$\widehat{y}_{F,O_\alpha} = \sum_{i \in s} y_{i,O_\alpha}/\pi_i, \tag{5.5}$$

where π_i is the probability that respondent i was included in our sample. π_i is typically known from the survey's sampling design. Equation 5.5 is a standard Horvitz-Thompson estimator, which is straightforward to compute using the data we collect about respondents' connections to people who have died, y_{i,O_α} (see, e.g., Sarndal et al., 2003, chap. 2).

The denominator of Equation 5.4 is the visibility of deaths, $\bar{v}_{O_\alpha,F}$. This quantity is more difficult to estimate. We can imagine many possible approaches, but the one adopted by the network survival estimator is very simple: we propose using the estimated average personal network size of *survey respondents* in demographic group α to estimate the visibility of deaths in demographic group α. (We will describe how to estimate personal network sizes below.) For example, our approach is to assume that the visibility of deaths among women aged 50-54 (i.e., the number of times each of these deaths could be reported) is the same as the personal network size of women in the frame population aged 50-54. This approach of using respondents' average personal network size to estimate the visibility of deaths will be exactly correct if (1) people who die in group α have personal networks that are the same size, on average, as people in the frame population in group α (the *decedent network assumption*); and, (2), respondents are perfectly aware of and report all of the deaths in their personal networks (the *accurate reporting assumption*). These are both strong assumptions. Fortunately, the network reporting framework permits us to be mathematically precise about what these assumptions say (Appendix D.2). This precision is useful for two reasons: first, it allows us to understand exactly how sensitive our estimates are to each assumption (Section 5.3.4 and Appendix D.3); and, second, it opens up the possibility that future studies could be designed to measure violations of these assumptions. These measurements could then be used to improve future estimates.

5.3.3 Estimating respondents' network sizes

We propose estimating the average size of the personal networks of respondents in demographic group α using an adapted version of the known population method of Killworth et al. (1998a). The known population method is based on asking respondents questions about their connections to groups of known size (e.g., "How many policemen do you know?"); intuitively, the more connections a respondent reports to policemen, the bigger we estimate her personal network to be. Respondents are typically asked about their connections to about 20 different groups of known size, and the results are combined using the known population estimator (Killworth et al., 1998a; Bernard et al., 2010; Feehan and Salganik, 2015).

The known population estimator was designed to estimate personal network sizes for all respondents, rather than for specific demographic groups. Therefore, in Appendix D.1, we derive an adapted estimator for the average network size of respondents in a particular demographic group α. The adapted known population estimator is:

$$\widehat{\bar{d}}_{F_\alpha,F} = \frac{\sum_{i\in s_\alpha}\sum_j y_{i,A_j}/\pi_i}{\sum_j N_{A_j}} \frac{N_F}{N_{F_\alpha}} = \frac{\widehat{y}_{F_\alpha,\mathcal{A}}}{N_{\mathcal{A}}} \frac{N_F}{N_{F_\alpha}}, \tag{5.6}$$

where $\bar{d}_{F_\alpha,F} = d_{F_\alpha,F}/N_{F_\alpha}$ is the average number of network connections between frame population members in demographic group α (F_α) and frame population members in any demographic group (F), per frame population member in demographic group α; N_F is the size of the frame population; N_{F_α} is the number of frame population members who are also in demographic group α; s_α is the subset of survey respondents in demographic group α; $j \in \{1,\ldots,J\}$ indexes the groups of known size; y_{i,A_j} is the number of connections that respondent i reports to group of known size A_j; $\widehat{y}_{F_\alpha,\mathcal{A}} = \sum_{i\in s_\alpha}\sum_j y_{i,A_j}/\pi_i$ is the estimated total number of connections from the frame population to all groups of known size; N_{A_j} is the size of the jth group of

known size; and $N_{\mathcal{A}} = \sum_j N_{A_j}$ is the total size of all of the groups of known size. Table 4.2 shows the groups of known size we asked respondents about.

Appendix D.1 has a detailed derivation of the adapted known population estimator, including the precise conditions that must hold for its estimates to be consistent and essentially unbiased. We will now illustrate the idea behind Equation 5.6 intuitively, leaving rigorous technical details to the appendix. For our illustration, we will assume that the frame population F is all adults, and that the demographic group α is women aged 45-55. The quantity we wish to estimate is $\bar{d}_{F_\alpha,F}$, the number of network connections from the average women aged 45-55 to all adults.

Researchers who use the known population method have typically asked respondents about their connections to approximately twenty groups of known size (Bernard et al., 2010); for example, in Rwanda, we asked about connections to 22 groups of known size (Table 4.2). However, in order to keep this example as simple as possible, we will only consider one group of known size: policemen. So, although the entire set of groups of known size $\mathcal{A}$ typically contains members of a variety of different groups, in this case it only consists of policemen.

In this example, the estimand for the right-hand side of Equation 5.6 can be written:

$$\frac{\widehat{y}_{F_\alpha,\mathcal{A}}}{N_{\mathcal{A}}} \frac{N_F}{N_{F_\alpha}} = \frac{\text{\# reported connections from women 45-55 to policemen}}{\text{total number of policemen}} \times \frac{\text{total number of adults}}{\text{number of women 45-55}}. \tag{5.7}$$

We will build up intuition about the estimand in Equation 5.7 in two steps. These two steps correspond to the two main conditions that must hold for the adapted known population method to provide consistent and unbiased estimates of $\bar{d}_{F_\alpha,F}$.

The first condition is that women aged 45-55 accurately report their connections to policemen. Each individual report does not have to be accurate, but it must be the case that reporting errors cancel out, so that the reports are correct in aggre-

gate. When this reporting condition holds, the aggregate reports are the same as the aggregate network connections: total reported connections to policemen equals total actual network connections to policemen ($y_{F_\alpha,\mathcal{A}} = d_{F_\alpha,\mathcal{A}}$). The estimand in Equation 5.7 then becomes

$$\frac{\substack{\text{\# connections} \\ \text{from women 45-55 to policemen}}}{\text{total number of policemen}} \times \frac{\text{total number of adults}}{\text{number of women 45-55}}. \tag{5.8}$$

The second condition is that policemen are typical of all adults, in terms of their connections to women aged 45-55. What does typical mean? It means that policemen have the same average number of connections to women aged 45-55 that the average adult does. This is called the *probe alter condition*, and in general it can be written $\bar{d}_{\mathcal{A},F_\alpha} = \bar{d}_{F,F_\alpha}$. (With many groups of known size, the probe alter condition says that all of the known groups taken together as a set must be typical of all adults, but not that any one group has to be typical; this is likely an advantage to having several different groups of known size).

In our example, with policemen as the only group of known size, the probe alter condition is:

$$\frac{\substack{\text{\# connections} \\ \text{from policemen to women 45-55}}}{\text{total number of policemen}} = \frac{\substack{\text{\# connections} \\ \text{from adults to women 45-55}}}{\text{total number of adults}}. \tag{5.9}$$

For example, if the average adult has 5 network connections to women aged 45-55, then the probe alter condition says that the average policeman should also have 5 network connections to women aged 45-55.

Finally, since the network connections are undirected, the total number of connections from women aged 45-55 to policemen is exactly the same as the total number of connections from policemen to women aged 45-55 (i.e., $d_{F_\alpha,\mathcal{A}} = d_{\mathcal{A},F_\alpha}$). Therefore, the numerator of the left-hand side of Equation 5.9 is equal to the number of connections from women 45-55 to policemen. We can now use the probe alter condition

(Equation 5.9) to substitute for the left-hand factor in 5.8 to obtain

$$\frac{\frac{\text{\# connections}}{\text{from adults to women 45-55}}}{\text{total number of adults}} \times \frac{\text{total number of adults}}{\text{number of women 45-55}} = \frac{\frac{\text{\# connections}}{\text{from women 45-55 to adults}}}{\text{number of women 45-55}}. \tag{5.10}$$

Again, the equality follows because the total number of connections from adults to women aged 45-55 is equal to the number of connections from women aged 45-55 to adults. This last expression is the quantity we wish to estimate: the average number of network connections from women 45-55 to adults.

5.3.4 Adjustment factors for $\widehat{O}_\alpha$

A major advantage of using the network reporting framework to derive the identity for the number of deaths (Equation 5.4) from first principles is that it enables researchers to understand precisely how violations of the main network survival assumptions—the decedent network condition and the accurate reporting condition—will affect estimated death rates. In Appendix D.3, we propose a decomposition that relates the true number of deaths (O_α) to the network survival estimand ($y_{F,O_\alpha}/\bar{d}_{F_\alpha,F}$) and three multiplicative adjustment factors ($\delta_{F,\alpha}$, $\eta_{F,\alpha}$, and $\tau_{F,\alpha}$):

$$O_\alpha = \underbrace{\left(\frac{y_{F,O_\alpha}}{\bar{d}_{F_\alpha,F}}\right)}_{\substack{\text{network survival} \\ \text{estimand for } O_\alpha}} \times \underbrace{\left(\frac{1}{\delta_{F,\alpha}}\right) \times \left(\frac{\eta_{F,\alpha}}{\tau_{F,\alpha}}\right)}_{\text{adjustment factors}}. \tag{5.11}$$

The first adjustment factor, the degree ratio ($\delta_{F,\alpha}$), is related to the structure of the underlying social network: it is exactly 1 when the decedent network assumption is satisfied, less than 1 if survey respondents have bigger personal networks than people who died, and greater than 1 otherwise. The other two adjustment factors, the true positive rate ($\tau_{F,\alpha}$) and the precision ($\eta_{F,\alpha}$) are related to the accuracy of reporting; when respondents' reports are perfectly accurate, then both $\tau_{F,\alpha}$ and $\eta_{F,\alpha}$

are 1. If there are false positive reports, then the precision will be less than 1; and, if respondents do not report all of the deaths that actually happen in their personal networks, then the true positive rate will be less than 1. Appendix D.3 has more information, including precise definitions of each adjustment factor.

5.3.5 Putting it all together to estimate death rates, $\widehat{M}_\alpha$

By combining our estimates of y_{F,O_α}, $\bar{v}_{O_\alpha,F}$, and E_α, we can estimate the death rate with

$$\widehat{M}_\alpha = \frac{\widehat{\bar{y}}_{F,O_\alpha}}{\widehat{\bar{v}}_{O_\alpha,F}} \frac{1}{\widehat{E}_\alpha} = \frac{\widehat{\bar{y}}_{F,O_\alpha}}{\widehat{\bar{d}}_{F_\alpha,F}} \frac{1}{\widehat{E}_\alpha}, \tag{5.12}$$

where the second step substitutes $\widehat{\bar{d}}_{F_\alpha,F}$ for $\widehat{\bar{v}}_{O_\alpha,F}$ as the network survival method does. Result D.2.5 in Appendix D.2.3 has the formal derivation, and Figure 5.2 illustrates how the different pieces of Equation 5.12 fit together conceptually.

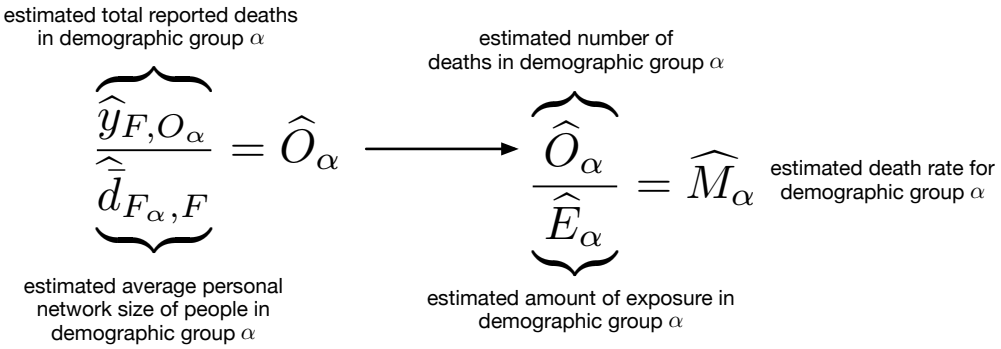


Figure 5.2: The components of network survival estimates for the death rate in a particular demographic group α (for example, α could be females aged 45-49 who died in the past year). The estimated total number of reported deaths is divided by the estimated average number of times each death gets reported (also called the visibility) to produce an estimate of the total number of deaths, $\widehat{O}_\alpha$. This estimated total number of deaths is then divided by an estimate for the amount of exposure, $\widehat{E}_\alpha$, to produce an estimated death rate, $\widehat{M}_\alpha$. The network survival method uses average estimated personal network size among survey respondents in demographic group α to estimate the visibility of deaths in demographic group α.

If the frame population is complete, meaning that everyone in demographic group α in the general population is also in the frame population, then Result D.2.6 in Appendix D.2.3 shows that we can use the simplified estimator

$$\widehat{M}_\alpha = \frac{\widehat{\bar{y}}_{F,O_\alpha}}{\widehat{\bar{d}}_{F_\alpha,F}} \frac{N_{\mathcal{A}}}{N_F}, \tag{5.13}$$

where $N_{\mathcal{A}}$ is the sum of the sizes of the groups of known size $N_{\mathcal{A}} = \sum_j N_{A_j}$. We use Equation 5.13 to produce estimated death rates from our Rwanda survey, below.

5.4 Results

We designed and conducted a household survey in Rwanda to test our new method. Our nationally-representative sample of 4,669 respondents enables us to produce network survival estimates of age-specific adult death rates in Rwanda over a 12-month period.

Deriving the network survival estimator from first principles reveals that it can be used to produce death rate estimates from information collected about a wide variety of different types of personal network. The challenge is to find the type of network that produces the most accurate information about death rates (Chapter 4; Feehan et al. (2015)). To demonstrate this generality, our study randomized respondents to report about one of two different types of personal network: half of our sample reported about their *acquaintance network*, while the other half of the sample reported about their *meal network* (Table 4.1). Each sampled household was randomly assigned to one of two possible definitions of a network tie, and all members of the household reported about the same tie definition.

Our survey used the same interviewers, data entry protocols, training techniques and sampling frame as the 2010 Rwanda DHS. Our sample was drawn using a stratified, two-stage cluster design, and interviews were conducted between June and Au-

gust of 2011. Following the guidelines of the DHS program (ICF International, 2012, sec. 1.13.7), we de-normalize the sampling weights by using the UN Population Division estimates for the size of Rwanda's population aged 15 and above in 2010 (United Nations, 2013). The full details of the sampling plan are described in RBC/IHDPC et al. (2012).

Note that Standard DHS are designed to have samples of 5,000 to 30,000 respondents (Corsi et al., 2012), and the 2010 Rwanda DHS has a sample size of about 13,500. Therefore, the estimated age-specific death rates from each tie definition are based on samples that are less than one fifth the size of the 2010 Rwanda DHS.

In total, respondents reported 3,853 deaths; of these, all but 8 had age and sex recorded. The 8 reported deaths that are missing age, sex, or both, are excluded from this analysis. Following conventional practice in the network scale-up literature, the aggregate relational data were topcoded at 30, which affected 0.2 percent of the responses. To estimate the sampling uncertainty for all quantities, we used the rescaled bootstrap (Rao and Wu, 1988a; Rao et al., 1992; Feehan and Salganik, 2015). All of our estimates were computed in R (R Core Team, 2014) using the following packages: networkreporting (Feehan and Salganik, 2014), plyr (Wickham, 2011), dplyr (Wickham and Francois, 2015), stringr (Wickham, 2012), ggplot2 (Wickham, 2009), devtools (Wickham and Chang, 2013), car (Fox and Weisberg, 2011), and gridExtra (Auguie, 2012). Our dataset is freely available from the DHS website[2].

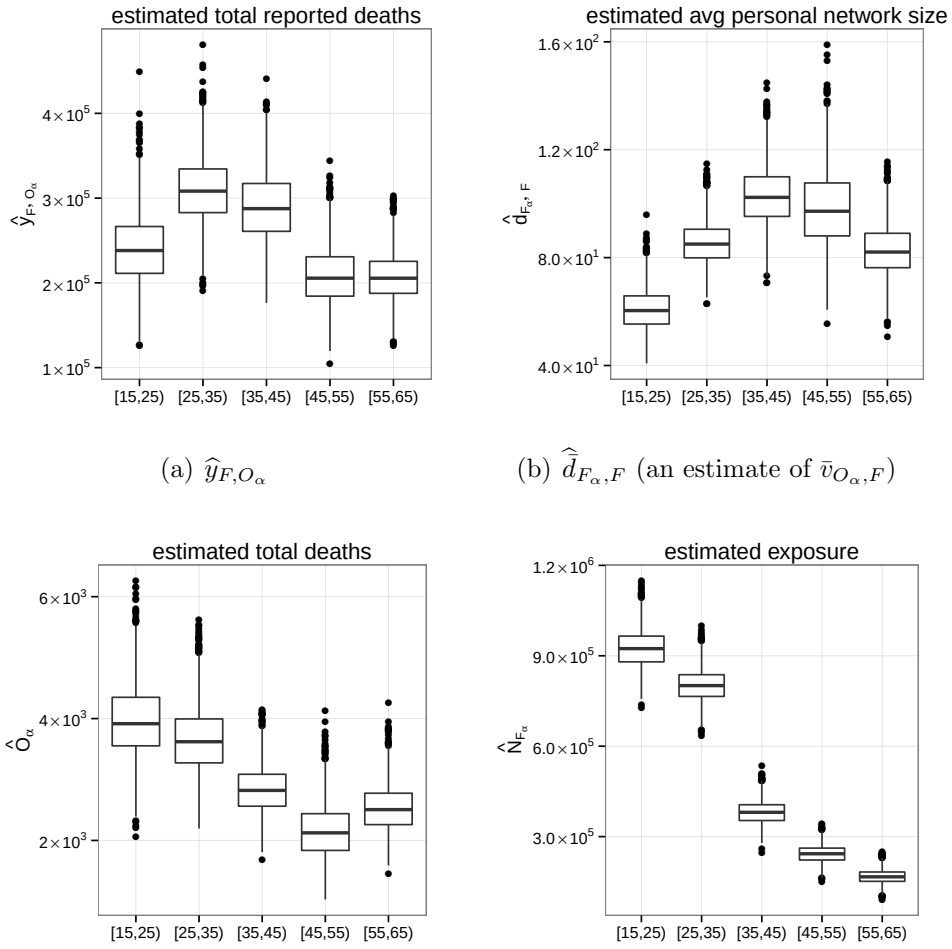
In order to illustrate the components of the network survival estimator, Figure 5.3 shows detailed results for one specific case: estimated Rwandan male death rates from reports about the meal tie definition. Each panel has a series of boxplots, and each boxplot shows the first quartile (top of rectangle), median (horizontal line), and third quartile (bottom of rectangle) of the estimated sampling distribution of the relevant quantity. For more information on interpreting boxplots, see McGill et al. (1978)

[2] `http://dhsprogram.com/what-we-do/survey/survey-display-422.cfm`

and Wickham (2009, `geom_boxplot`). Panel 5.3(a) shows, for each age group, the estimated total number of reports about deaths ($\widehat{y}_{F,O_\alpha}$, Equation 5.5). Since each death can be reported multiple times, this quantity on its own is not enough to estimate the total number of deaths in the population. Panel 5.3(b) shows, for each age group, the estimated size of respondents' personal networks, which is used as an estimate for the visibility of deaths ($\widehat{\bar{d}}_{F_\alpha,F}$, Equation 5.6). Dividing the total estimated reports about deaths (Panel 5.3(a)) by the estimated visibility of deaths (Panel 5.3(b)) produces the estimated total number of deaths by age group ($\widehat{O}_\alpha$) shown in Panel 5.3(c). Panel 5.3(d) shows the estimated number of people in each age group ($\widehat{N}_{F_\alpha}$), which is used as an estimate of exposure; this quantity comes from the sampling design.

Figure 5.4 shows the estimated age-specific death rates ($\widehat{M}_\alpha$), logged (right-hand panel) and not logged (left-hand panel); these estimated death rates come from dividing the estimated total number of deaths ($\widehat{O}_\alpha$) by the estimated amount of exposure ($\widehat{E}_\alpha$). Each line in Figure 5.4 shows the estimate from one bootstrap resample; taken together, the set of lines shown in each panel show the estimated sampling uncertainty of the death rates. Analogous figures illustrating the estimates for each component of the age-specific death rates for females and for the acquaintance tie definition can be found in Appendix D.4.

When interpreting and comparing multiple sets of age-specific death rates, it can be helpful to consider a summary measure. ${}_{45}q_{15}$ is a summary measure which is interpretable as the conditional probability of dying before age 60 among people who survive to age 15, and who then face the given age-specific death rates. For example a set of age specific death rates with ${}_{45}q_{15}$ of 0.2 implies that 20% of people who survive to age 15 and then face those age-specific death rates will die before age 60. This natural interpretation can help to provide intuition about how plausible a set of age-specific death rates is, and differences in ${}_{45}q_{15}$ can help to provide intuition



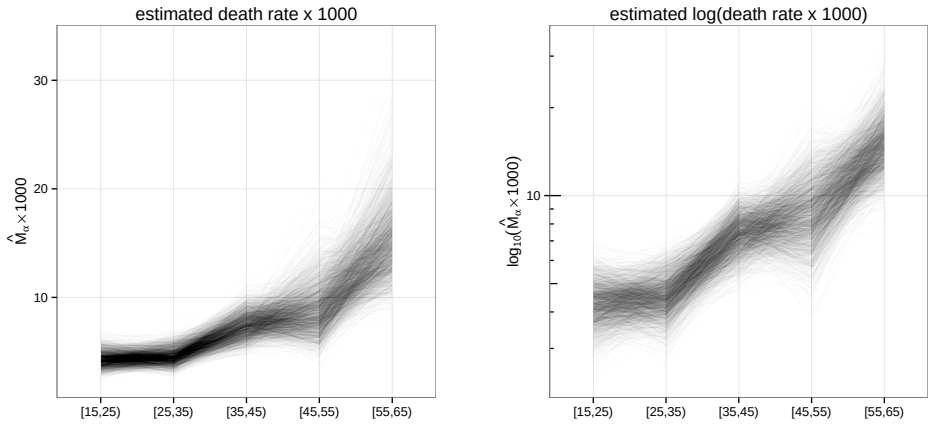
(a) $\widehat{y}_{F,O_\alpha}$

(b) $\widehat{\bar{d}}_{F_\alpha,F}$ (an estimate of $\bar{v}_{O_\alpha,F}$)

(c) $\widehat{O}_\alpha = \frac{\widehat{y}_{F,O_\alpha}}{\widehat{\bar{d}}_{F_\alpha,F}}$

(d) $\widehat{N}_{F_\alpha}$ (an estimate of E_α)

Figure 5.3: Estimating components of age-specific death rates for Rwandan Males for 12 months prior to our survey using responses from the meal tie definition. Figure 5.2 has a conceptual overview. The average personal network size of survey respondents ($\widehat{\bar{d}}_{F_\alpha,F}$; Panel 5.3(b)), is used as an estimate of the visibility of deaths ($\bar{v}_{O_\alpha,F}$; i.e., the number of times each death could be reported). The estimated number of deaths in the population ($\widehat{O}_\alpha$; Panel 5.3(c)) is obtained by dividing estimated total reports about deaths ($\widehat{y}_{F,O_\alpha}$; Panel 5.3(a)) by the estimated visibility of deaths ($\widehat{\bar{v}}_{O_\alpha,F}$; Panel 5.3(b)). The estimated size of the frame population ($\widehat{N}_{F_\alpha}$) is used as an estimate of the population exposure E_α. Estimated age-specific death rates ($\widehat{M}_\alpha$; Figure 5.4) are are obtained by dividing the estimated number of deaths ($\widehat{O}_\alpha$; Panel 5.3(c)) by the amount of exposure ($\widehat{E}_\alpha$; Panel 5.3(d)). Sampling uncertainty from each step is estimated using the rescaled bootstrap approach to account for the complex sample design (Rao and Wu, 1988a; Rao et al., 1992).



(a) $\widehat{M}_\alpha \times 1,000$ (b) $\log(\widehat{M}_\alpha \times 1,000)$

Figure 5.4: Estimated age-specific death rates for Rwandan Males for 12 months prior to our survey using responses from the meal tie definition. These estimated death rates are produced using the components illustrated in Figure 5.3. The left-hand panel shows the death rates multiplied by 1,000, and the right-hand panel shows the log of the death rates multiplied by 1,000. (Note that the scale for the y-axis is different for each panel.) Each line shows the result of one bootstrap resample; taken together, the lines show the estimated sampling uncertainty for each set of death rates.

about how similar or different two sets of age-specific death rates are. Appendix D.6 reviews the formal definition for ${}_{45}q_{15}$, and describes how we construct estimates of ${}_{45}q_{15}$ from survey data. Figure 5.5 shows the estimated ${}_{45}q_{15}$ from the meal and acquaintance tie definitions, for males and for females. The distributions in Figure 5.5 reflect the sampling uncertainty in estimated ${}_{45}q_{15}$ values, as computed using the rescaled bootstrap. Table 5.1 provides several summaries of these estimated sampling distributions.

The left and middle panels in Figure 5.6 compare the estimated age-specific death rates across the two tie definitions for males and for females. The solid blue and dashed green lines in Figure 5.5 show estimated ${}_{45}q_{15}$ from the age-specific death rates across the two tie definitions for males and for females. Four main conclusions emerge from comparing the adult mortality estimates across the network and acquaintance tie definitions for males and females. First, for all tie definitions and for both sexes, the absolute level of the estimated death rates is plausible: ${}_{45}q_{15}$ es-

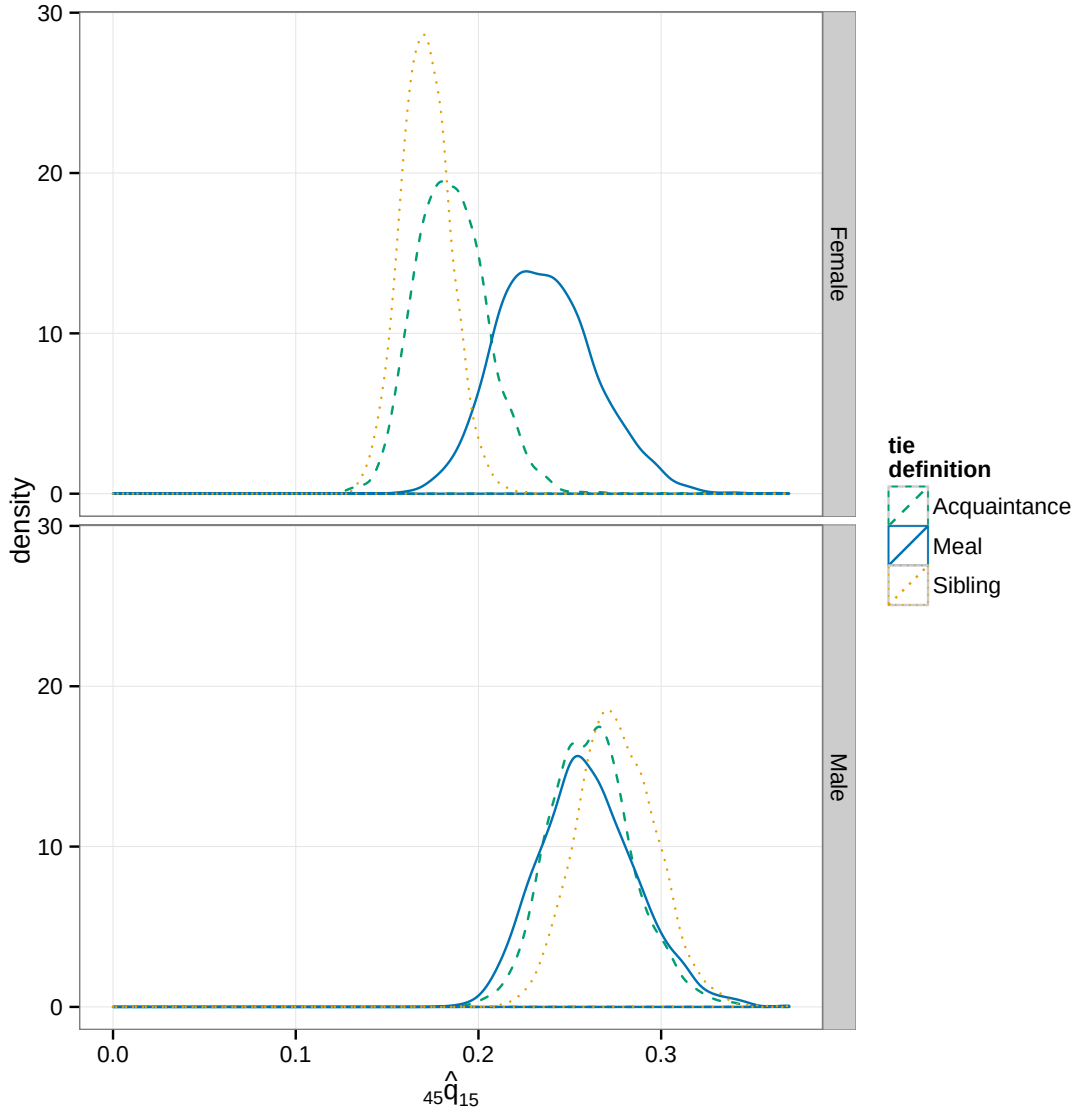


Figure 5.5: Comparison between estimated conditional probabilities of death at adult ages ($_{45}q_{15}$) computed from death rate estimates for the two types of personal network investigated with our survey (Acquaintance and Meal), and direct sibling survival. The distributions reflect sampling uncertainty, as estimated with the rescaled bootstrap.

Table 5.1: Estimated ${}_{45}q_{15}$ values, by tie definition and sex, along with summaries of the estimated sampling distribution for each estimator. The coefficient of variation is the estimated standard error divided by the estimated mean; it is a comparable index of sampling variation across the different estimates.

Tie definition	Sex	mean	median	std. error	coef. of variation	95% CI
Acquaintance	female	0.19	0.18	0.02	0.11	[0.15-0.23]
Meal	female	0.24	0.24	0.03	0.12	[0.19-0.30]
Sibling	female	0.17	0.17	0.01	0.08	[0.15-0.20]
Acquaintance	male	0.26	0.26	0.02	0.09	[0.22-0.31]
Meal	male	0.26	0.26	0.03	0.10	[0.21-0.32]
Sibling	male	0.28	0.27	0.02	0.08	[0.24-0.32]

timates suggest females have about a 24% (meal) and 19% (acquaintance) chance of dying at adult ages, while estimates suggest that males have about a 26% chance of dying at adult ages (both meal and acquaintance). Second, for all tie definitions and for both sexes, the estimated age-pattern of death rates is plausible: the estimated death rates generally increase with age (with the exception of young females for the meal definition). Barring special circumstances (like war), we expect death rates to increase with age among adults, with the possible exception of young female death rates in settings where maternal mortality is high. Third, the sampling uncertainty in the estimated age-, sex-, and year-specific death rates is modest, even though these estimates were produced from samples about half the size of the smallest DHS: Table 5.1 shows that the coefficient of variation in the estimator for ${}_{45}q_{15}$ from the acquaintance and meal tie definitions ranges from about 0.09 to 0.12. And, fourth, the meal and acquaintance tie definition estimates are more similar for males than for females. For females, the discrepancy between estimates from the different tie definitions is especially large for the youngest age group, ages 15-24, where estimated death rates from the meal definition are much higher than the estimated death rates from the acquaintance definition.

Figure 5.7 gives an example of using the decomposition proposed in Section 5.3.4 and Appendix D.3 to illustrate the robustness of the estimated age-specific death

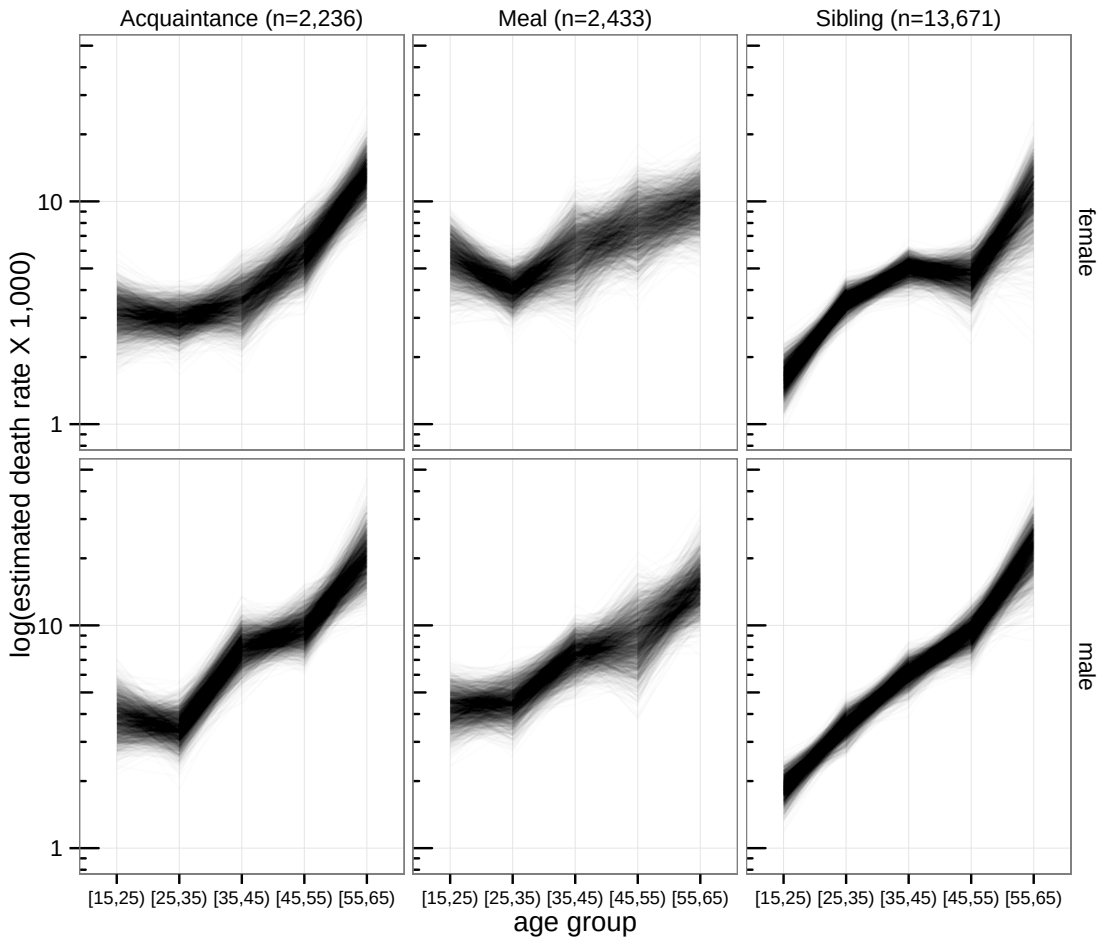


Figure 5.6: Comparison between network survival death rate estimates for the two types of personal network investigated with our survey (left-hand column and middle column), and direct sibling survival death rates estimates from the 2010 Rwanda Demographic and Health Survey (right-hand column). The top row has death rates estimated for females, while the bottom row has death rates estimated for males. The network survival estimates are based on reported deaths from the 12 months prior to the interview. The sibling method is not typically capable of producing estimates for only 12 months before the interview because the sampling variation that results is too large; instead, the sibling estimates are based on reported deaths in the 60 months prior to the interview.

rates for males to violations of the main assumptions required by the network survival method. The figure shows the extent to which death rates are affected by (i) violations of the decedent network condition (columns); and (ii) violations of the accurate reporting condition (rows). If future researchers collect the data needed to estimate any of the adjustment factors in our decomposition (Equation 5.11), then those estimated adjustment factors can be combined with the data we collected in Rwanda to produce adjusted death rates estimates. Note that, in Figure 5.7, we assume that the adjustment factors are the same for each age group; researchers who disagree with this assumption can use our decomposition to adjust each age group separately.

5.5 Comparison to other estimates

Ideally, we would evaluate the estimates produced by the network survival method by comparing them to a gold-standard set of age- and sex-specific death rates for Rwanda in 2011. Unfortunately, no such gold standard exists. However, we can compare the network survival estimates to other estimates that have been made for Rwanda at about the same time. These comparisons will allow us to understand the extent to which our network survival estimates agree or disagree with other mortality estimation approaches that have been applied in Rwanda, but we will not be able to determine which of the estimates we review is the most accurate.

5.5.1 Sibling survival from the 2010 RDHS

We compare the network survival estimates to direct sibling survival estimates produced from the 2010 Rwanda Demographic and Health Survey (NISR et al., 2012). The 2010 Rwanda DHS finished fieldwork in March 2011, right before our data collection started; it used a stratified, multi-stage design to select a sample of 12,540

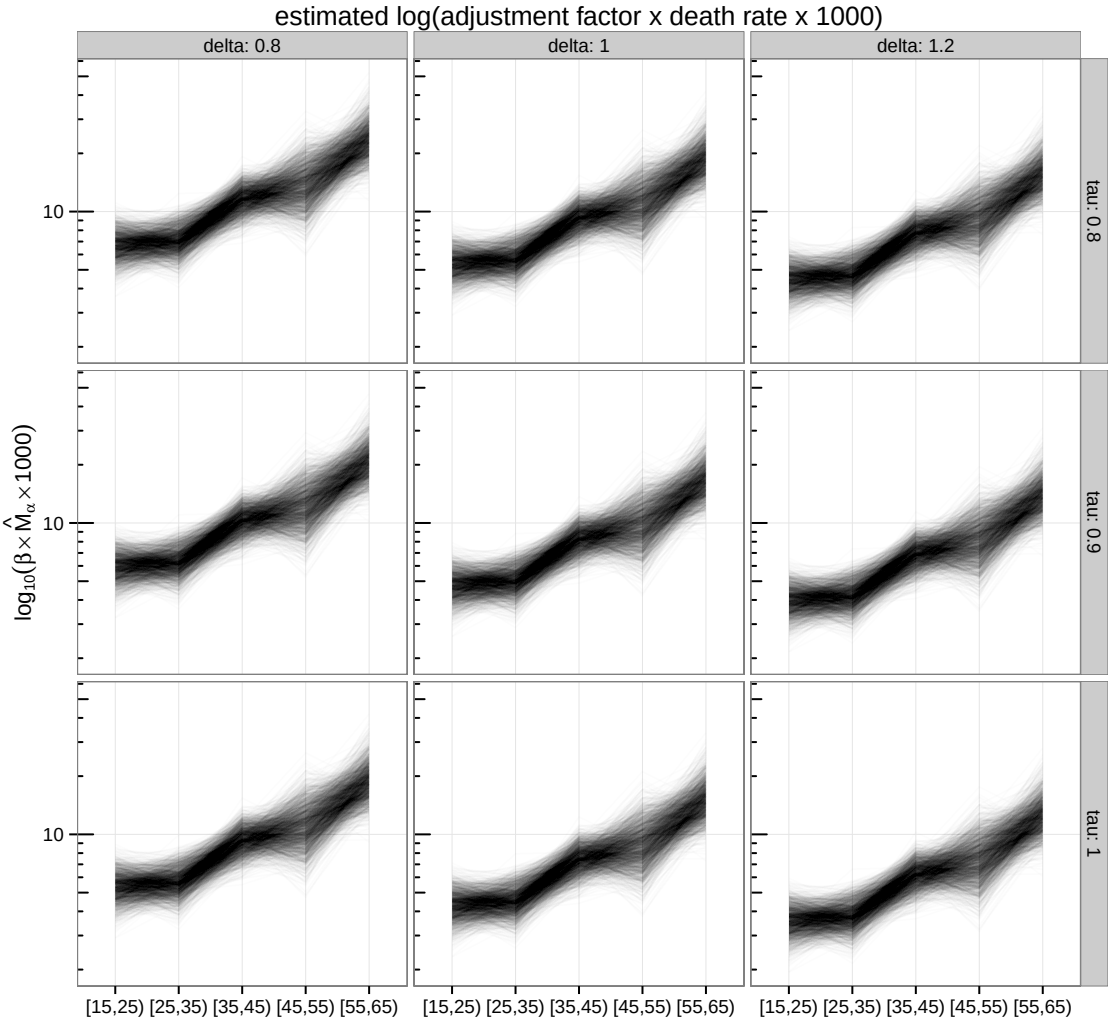


Figure 5.7: Estimated age-specific death rates for Rwandan males using the meal definition under several different possible violations of the decedent network condition and the accurate reporting condition. Panels on the bottom row have accurate reporting; panels in the middle row have $\tau_{F,\alpha} = 0.9$, meaning that respondents report 90% of the deaths they are connected to; and panels in the top row have $\tau_{F,\alpha} = 0.8$, meaning that respondents report 80% of the deaths they are connected to. Panels in the middle column have $\delta_{F,\alpha} = 1$, meaning that people who die have personal networks that are the same size, on average, as the frame population (i.e., the decedent network condition holds). Panels in the left column have $\delta_{F,\alpha} = 0.8$, meaning that people who die have personal networks that are only 80% as large as frame population members. Panels in the right column have $\delta_{F,\alpha} = 1.2$, meaning that people who die have personal networks that are 20% larger than frame population members. Both the decedent network condition and the accurate reporting condition hold when $\tau_{F,\alpha} = 1$ and $\delta_{F,\alpha} = 1$ for all α; therefore, the panel in the center of the bottom row is the same as Figure 5.4.

households. As is typical in a Demographic and Health Survey, the sibling history questions were only asked of the female respondents, and only women aged 15-49 were interviewed. Therefore, the sibling survival estimates below are based on the sibling histories of the 13,671 women who were interviewed in the sampled households. We estimate the sampling uncertainty in the age-specific death rate estimates from the sibling histories using the rescaled bootstrap, which accounts for the complex sample design (Rao and Wu, 1988a; Rao et al., 1992). Appendix D.5 has a detailed explanation of how we construct sibling survival estimates, and NISR et al. (2012) has a complete description of the 2010 Rwanda Demographic and Health Survey.

A major limitation of the direct sibling survival method is that it cannot generally be used to produce direct estimates of yearly age-specific death rates because samples are typically too small (Section 5.2.2). Even though the 2010 Rwanda DHS collected sibling histories from 13,671 women, we found that estimated death rates for the 12 months before the survey were too imprecise to usefully compare to network survival estimates (see Appendix D.5). Therefore, we follow the recommendations of the sibling survival literature and pool together information from reports about several years prior to the survey (Stanton et al., 2000; Timaeus and Jasseh, 2004). The sibling survival estimates are thus estimated average death rates over the 60 months before the survey, while the network survival estimates are estimated death rates for the 12 months prior to the survey.

Figure 5.8 shows that, in Rwanda, the network survival approach produced more reports about deaths from each interview than the DHS sibling histories did, even with the longer 60-month time period for reporting deaths with the sibling histories. Figure 5.8 shows that the acquaintance definition produced about 12 times more reports about deaths than the sibling histories, and the meal definition produced about 6 times more reports about deaths than the sibling histories.

Figure 5.6 compares the age-specific death rates produced from the network reporting estimates (left and middle columns) to the ones produced by the direct sibling survival method (right column). This comparison produces three main conclusions. First, within each sex, the level of estimated death rates across adult ages is similar for the sibling and the network survival estimates. Second, the sibling estimates for the youngest age group are lower than the network survival estimates for both sexes. Third, the differences in the age pattern suggested by each of three estimators is more pronounced for females than for males.

Taken together, these results show a reasonably high level of agreement between the death rates estimated from the 2010 Rwanda DHS sibling histories and both tie definitions of the network survival survey. However, the estimates are not identical, and the differences between them are especially pronounced for females and for young age groups. Without a gold standard, we cannot say which of the three is the most accurate.

5.5.2 Estimates from international organizations

We also compare the network survival estimates to estimated adult mortality levels published by three large organizations: the United Nations Population Division (UNPD) (United Nations, 2013)[3]; the World Health Organization (WHO, 2015)[4]; and, the Institute for Health Metrics and Evaluation (Wang et al., 2013)[5]. All three of these organizations regularly publish estimates of adult mortality levels across the world, by country and time period. Their estimates are derived from a complex combination of data sources, models, and expert judgment.

[3] UNPD estimates are taken from the 2012 revision of the World Population Prospects: `http://esa.un.org/wpp/Excel-Data/mortality.htm` (accessed April 22, 2015).

[4] WHO estimates are taken from the Global Health Observatory: `http://apps.who.int/gho/data/view.main.1360` (accessed April 22, 2015).

[5] IHME estimates are taken from the 2010 Global Burden of Disease study: `http://ghdx.healthdata.org/record/adult-mortality-estimates-country-1970-2010` (accessed April 22, 2015).

Estimates from the international organizations are frequently published and used as ${}_{45}q_{15}$, rather than detailed age-specific death rates. Therefore, our comparisons focus on ${}_{45}q_{15}$. Figure 5.9 compares estimated ${}_{45}q_{15}$ for males and females from six different sources, for the time period closest to 2011 (note that the time periods are not the same for all estimates). Figure 5.9 supports two main conclusions. First, for all sources, estimated ${}_{45}q_{15}$ is higher for males than for females; however, the acquaintance network, sibling, and IHME estimates suggest a bigger gap between female and male adult ${}_{45}q_{15}$ than the meal network, UNPD, and WHO estimates do. Second, similar to the results from estimated age-specific death rates, the survey-based estimates of ${}_{45}q_{15}$ are more consistent with one another for males than for females. Third, the sampling uncertainty around the three survey-based estimates of ${}_{45}q_{15}$ is comparable: the estimates from the meal network and the acquaintance network are about as precise as the estimates from the sibling histories, even though (i) the sibling estimates are based on more than five times as many interviews as either network survival estimate; and (ii) the sibling estimates are based on deaths reported over five times as long a time period as the network survival estimates. No sampling-based uncertainty estimates are available for the estimates from the international agencies.

5.6 Discussion

Understanding adult mortality is critical to a wide range of important research and policy questions, but estimating adult death rates remains difficult in countries that lack high-quality vital registration systems. In this study, we introduced the network survival method to help address this problem. The network survival method uses information about survey respondents' personal networks to estimate adult death rates. We derived the estimator from first principles, which permits us to understand exactly

what assumptions have to hold in order for the estimator to work well. Moreover, our derivation reveals that this approach is quite general: data can be collected about a wide variety of different types of personal networks in order to estimate adult death rates.

We designed and conducted a nationally-representative household survey to test the new method in Rwanda, a setting where improved methods for estimating adult death rates are needed. To test the generality of the network survival method, we randomized our sample so that half of the respondents reported about their acquaintance network, while the other half reported about their meal network. We were able to use the network survival method to estimate adult death rates by age and sex over a 12 month period, even though our sample was much smaller than a typical DHS. We also compared the network survival estimates to existing adult mortality estimates for Rwanda. Our results suggest that asking about survey respondents' personal networks can produce more information about deaths than the sibling history method. Substantively, the network survival estimates generally agreed with other adult mortality estimates available for Rwanda; however, across all of the sources of estimates, there is disagreement about the extent to which males face higher death rates than females at adult ages. Since there is no gold standard set of adult death rates available in Rwanda, we cannot determine which of the estimates is the most accurate.

The network survival approach helps to address the main concerns that currently cloud the sibling survival estimator. First, DHS samples are typically not large enough to allow researchers to directly estimate adult death rates from sibling histories (Hill et al., 2006). The data we collected about meal and acquaintance networks in Rwanda both provided much more information about deaths from each interview than the sibling histories did, permitting us to estimate adult death rates by age and sex for the 12 months prior to the survey without any need for pooling data across countries,

smoothing regressions or model life tables (Figure 5.8). Second, researchers have long debated possible structural biases in direct sibling survival estimates; for example, it is not possible to learn about sibships that have no survivors left to be interviewed. By deriving the network survival estimator from first principles, we show precisely what conditions must hold for the estimator to work well; there is no ambiguity about potential structural problems in network survival estimates. Further, by asking respondents to report about reasonably broad networks, we expect that researchers can minimize the risk of having all of the members of a network die out, leaving nobody who could be interviewed. Finally, many studies have examined the possible sources of reporting error when sibling histories are collected. Our derivation of the network survival estimator allows us to clearly identify the different possible sources of reporting error, and also how these different sources of reporting error potentially interact. This understanding enables researchers to understand exactly how robust estimated death rates are to violations of each assumption. This analytic understanding also opens up the possibility of developing data collection and analysis methods that measure and potentially adjust for assumptions that do not hold.

There are many potential directions for future work. First, our inability to determine which of the different sources of Rwandan adult mortality estimates was most accurate emphasizes the importance of conducting validation studies in settings where survey-based methods like network survival and sibling survival can be compared to gold standard adult death rates. Given the importance of adult mortality estimates and the scale of sibling history data collection, there has been surprisingly little work like this in the past (though the studies of Hill et al. (2006) and Helleringer et al. (2014b) are exceptions, and good foundations for future work to build upon). Second, the flexibility of the network survival approach is potentially a big advantage; future studies could develop a better understanding of what sort of personal network respondents should be asked to report about. This choice of tie definition potentially has

implications for effective sample size and for the size and nature of reporting errors and structural biases. Third, researchers could use the network reporting framework to better understand how to produce estimates from existing sibling history data. Previous work has not treated sibling histories as a type of network reporting, but we expect this approach will be productive in resolving some of the methodological debates surrounding the direct sibling survival method. Finally, although we focus on estimating timely, national-level adult death rates in this study, there is a serious demand for survey-based approaches to studying mortality in a wide range of other situations, including conflicts, natural disasters, famines, epidemic outbreaks, and other humanitarian crises (Checchi and Roberts, 2008; Epicentre, 2007). In many of these situations, understanding how to ask respondents to report about their personal networks may help researchers design surveys that will permit them to estimate quantities that have proven very difficult to estimate so far.

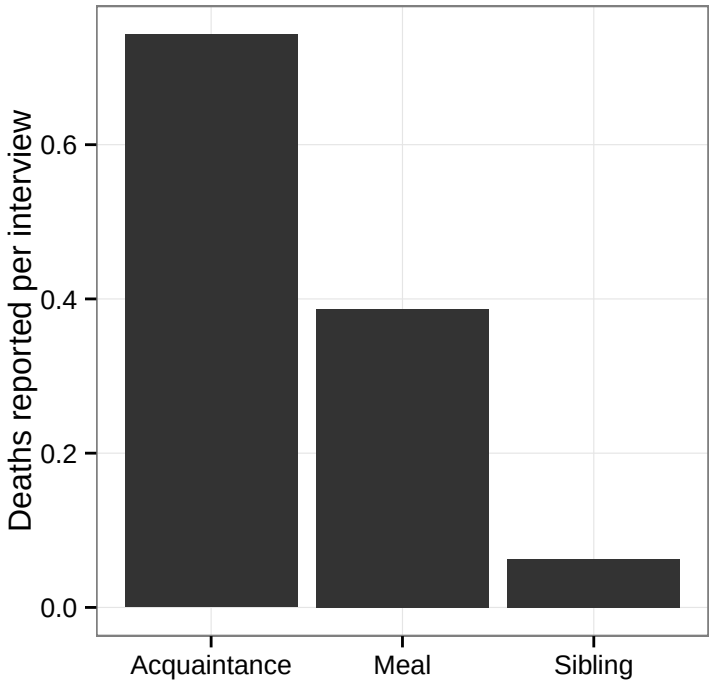


Figure 5.8: Average number of deaths reported from each interview in Rwanda using the acquaintance and meal tie definitions from the network survival study, and using the sibling history module of the 2010 Rwanda Demographic and Health Survey. Compared to the sibling reports, respondents reported about 12 times more deaths using the network survival tie definition and about 6 times more deaths using the meal tie definition, even though the sibling approach uses reported information about deaths in the 60 months prior to the survey, while the acquaintance and meal definitions use reported information about deaths in the 12 months prior to the survey.

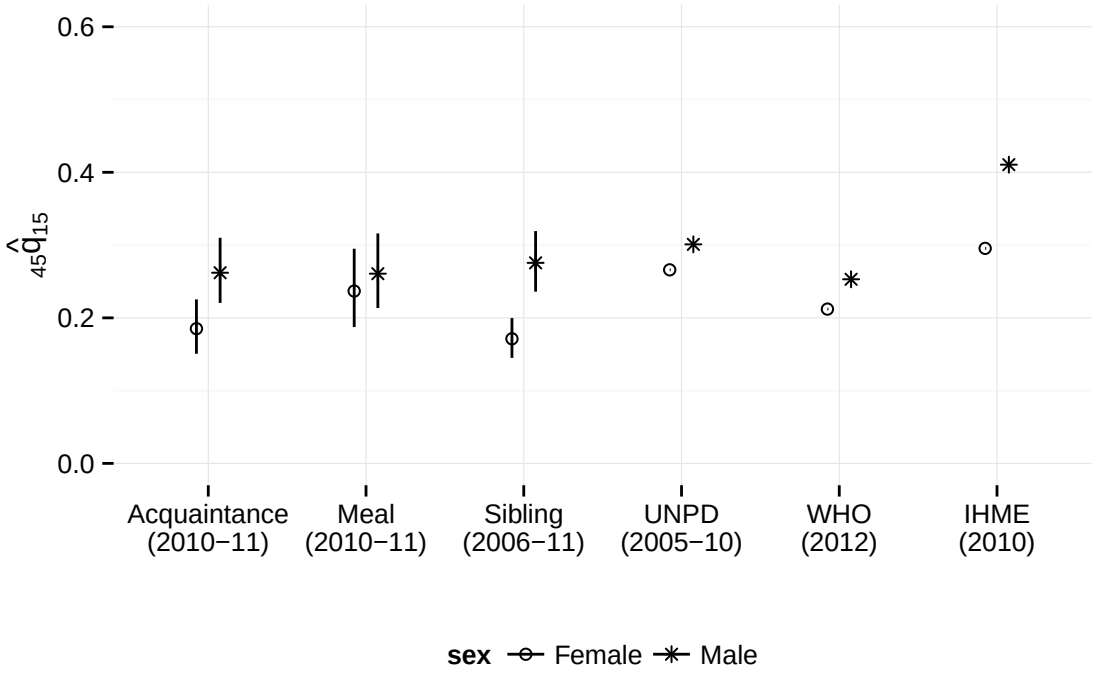


Figure 5.9: Estimated ${}_{45}q_{15}$ for Rwanda from six different sources: the acquaintance and meal tie definitions from our network survival survey; the sibling histories from the 2010 Rwanda Demographic and Health Survey; the United Nations Population Division (UNPD); the World Health Organization (WHO); and the Institute for Health Metrics and Evaluation (IHME). Error bars indicate 95% uncertainty intervals for the survey-based estimates, which were computed using the rescaled bootstrap. Note that the estimates are not for exactly the same time periods.

Chapter 6

Conclusion

Network reporting methods help researchers to generalize the traditional sample survey: instead of asking respondents to report about themselves, network reporting studies ask respondents to report about others. Since people everywhere are embedded in a wide range of different personal networks, this generalization is potentially a powerful way to to help researchers overcome many of the difficulties that arise when studying groups that are hidden or rare. This approach also offers the possibility of studying the structure of social networks themselves. However, there are considerable challenges that go hand in hand with asking respondents to report about others. These challenges can be considered the cost of the flexibility and the extra information that network reporting surveys provide.

The goal of this dissertation was to develop a general network reporting framework that brings together the common features shared by any study that asks respondents to report about others. Asking people to report about others is not a new approach: in fact, the idea is so natural that, one way or another, it has independently arisen in several different fields. In many cases, these methods were not conceptualized as network approaches, and in almost all of them, researchers have raised several of the concerns we have seen throughout this dissertation: possible errors in reporting; the

possible impact of network structure on resulting estimates; and problems of sampling and study design. What seemed to be missing was a synoptic way of understanding what network reporting studies have in common, and what distinguishes them from one another.

We introduced the core of the network reporting framework in Chapter 2. Starting from first principles, we built up a toolkit that we hope researchers will find helpful when reasoning about any network reporting problem. Throughout the dissertation, we showed that this toolkit can be used to better understand existing methods like network scale-up and sibling survival, and that it can be used to develop new estimators and data collection strategies, like network survival and generalized scale-up.

This dissertation has only begun to investigate the scientific possibilities introduced by network reporting methods, and our technical understanding of network reporting methods remains far from perfect. So there are many directions for future work. Throughout the dissertation, we offered specific recommendations for future work that was related to the topic of each individual chapter. Here we focus on a few broader themes that we hope researchers will pursue in the future.

First, we hope that future research will build up a deeper theoretical and empirical understanding of the relationship between different types of personal networks and the reports they produce. Do people tend to know more about their families or their friends? Does that vary, depending on what type of quantity respondents are being asked about? This issue may initially seem like a technical detail, but, like many of the questions surrounding network reporting, it actually gets to the heart of our understanding of social networks and how information diffuses through them. Both Chapter 4 and Chapter 5 show that there is good reason to expect that the network respondents are asked about matters. Chapter 4 provided a first effort towards building an understanding about the type of network and the resulting reports by presenting a conjecture about how the strength of networks ties may be related

to the accuracy of the information it produces. But we expect and hope that this conjecture will be refined and improved in the future.

Second, we hope that the network reporting framework is able to help researchers who are interested in designing new ways of collecting data. These new ways of collecting data could focus on (i) estimating the adjustment factors that relate network reports to the underlying social network (such as the transmission rate and the false positive rate); (ii) combining network reporting with exciting new sampling procedures and data collection modes, like mobile phones or the internet; and, (iii) developing new ways to estimate important quantitative features of social networks beyond the degree distribution. For example, social network research would benefit from a broader range of sample-based strategies for estimating how networks form, are maintained, and dissolve over time; how networks vary over physical space, including between countries; and about higher-order features of network structure like centrality and clustering.

Third, although this dissertation has adopted a design-based perspective, we hope that the some of the insights from this analysis will be helpful for model-based approaches as well. If more data about network reporting studies accumulate over time—particularly data about reporting errors and adjustment factors like the transmission rate and the false positive rate—then these model-based approaches will become particularly appealing, because they will offer researchers the ability to benefit from the insights that have been produced by previous studies.

Finally, our understanding of network reporting methods could benefit from deeper integration with insights from cognitive science and psychology. Experts in those disciplines work towards developing a rigorous understanding of the inner workings of human memory and perception. These phenomena are obviously related to network reports, which we expect to be governed in part by respondents' memories, and their perceptions of themselves and others. Our understanding of how to design and analyze

network reporting studies would surely benefit from incorporating more insights from these disciplines.

This dissertation is far from the last word on network reporting methods; there is much we have yet to learn. But there is also much to be gained: network reporting methods offer the prospect of opening up to quantitative inquiry a wide range of critical but neglected questions in science and public policy. We hope that, piece by piece, researchers will develop a rigorous mathematical and empirical understanding of how to ask survey respondents to report about others.

Appendix A

Notation

This appendix contains a table of notation, and a figure that illustrates several key concepts (Figure A.1).

Table A.1: Notation used in this dissertation.

Quantity	Explanation
U	the entire population
H	the hidden population (e.g. drug injectors)
F	the frame population (typically adults over a certain age)
N	size of the entire population, $\|U\|$ (i.e., everyone who could ever be interviewed or reported about)
N_F	size of the frame population, $\|F\|$ (i.e., everyone who could ever be interviewed)
N_H	size of the hidden population, $\|H\|$ (e.g. drug injectors, people who died, migrants, and so on)
$y_{i,H}$	out-reports from i about connections to H (i's answer to "How many drug injectors do you know?")

Table A.1: Notation used in this dissertation.

Quantity	Explanation
$y_{F,H}$	total out-reports from the frame population F about connections to H
$y^{+}_{F,H}$	true positive out-reports from the frame population F about connections to H (i.e., the sum of reported connections that actually lead to people in H)
$A_1, \ldots, A_J$	the J groups of known size (e.g., A_1 might be people named Amy, A_2 might be doctors, and so forth)
$\mathcal{A}$	multiset formed by concatening together members of all the groups of known size $A_1, \ldots, A_J$ (e.g. people named Amy and doctors, with doctors named Amy included twice)
$N_{A_1}, \ldots, N_{A_J}$	size of each of the J groups of known size (e.g., A_1 might be the number of people named Amy, A_2 might be the number doctors, and so forth)
$N_{\mathcal{A}}$	size of $\mathcal{A}$, $N_{\mathcal{A}} = \sum_j N_{A_j}$ (e.g. number of people named Amy plus the number of doctors, and so forth)
$d_{i,F}$	number of network connections from i to F (which is not necessarily the same as the number of reported connections from i to F)
$d_{A,B}$	number of network connections from group $A \subset U$ to group $B \subset U$ (note that A and B could be the same group, they could be entirely distinct groups, or they could overlap partially)

Table A.1: Notation used in this dissertation.

Quantity	Explanation
$\bar{d}_{A,B} = \frac{d_{A,B}}{N_A}$	average number of network connections from group A to group B, per member of group A (note that we always take averages with respect to the first subscript)
$\bar{v}_{H,F} = \frac{v_{H,F}}{N_H}$	average visibility (number of in-reports) per member of H
$\tilde{v}_{i,A_j}$	i's report about his visibility to group A_j (i's answer to "How many of the Amys you know are aware that you are a drug injector?")
s	a probability sample of people from the frame population F
π_i	the probability that $i \in F$ is included in the sample, which comes from the sampling design
s_H	a relative probability sample of people from the hidden population
$c\pi_i$	some constant c times the probability π_i that $i \in H$ is included in a relative probability sample of H
τ_F	the true positive rate for out-reports from F
δ_F	the degree ratio of hidden population members relative to frame population members
ϕ_F	the frame ratio
η_F	the false positive rate for out-reports from F
α	a demographic group (e.g., women aged 45-54 in 2010)
F_α	frame population members who are in demographic group α

Table A.1: Notation used in this dissertation.

Quantity	Explanation
N_{F_α}	the number of frame population members who are in demographic group α
N_α	the number of people in the entire population U who are also in demographic group α
O_α	the number of deaths in demographic group α (e.g. number of deaths among women aged 45-54 in 2010); we use O for deaths because death rates are occurrence-exposure rates, and deaths are occurrences in this context
E_α	person-years of exposure to the possibility of dying in demographic group α (e.g. years of life lived among women aged 45-54 in 2010)
$M_\alpha = \frac{O_\alpha}{E_\alpha}$	the death rate in demographic group α (e.g. the death rate for women aged 45-54 in 2010)
${}_{45}q_{15}$	conditional probability of death before age 60, among those who survive to age 15 and face a particular set of age-sepcific death rates
$\widehat{N}_H^{(A)}$	hidden population size estimate based on data from experimental arm A (e.g, the meal definition estimate for the number of female sex workers)
σ_A^2	the sampling variance of the hidden population size estimator from experimental arm A, $\widehat{N}_H^{(A)}$ (e.g., the sampling variance of the meal definition estimate for the number of female sex workers)

Table A.1: Notation used in this dissertation.

Quantity	Explanation
$w^\star = \frac{\sigma_A^2}{\sigma_A^2+\sigma_B^2}$	the optimal blending weight for the blended estimate $\widehat{N}_H = w\widehat{N}_H^{(A)} + (1-w)\widehat{N}_H^{(B)}$, under the assumption that the estimates from experimental arms A and B are unbiased and independent

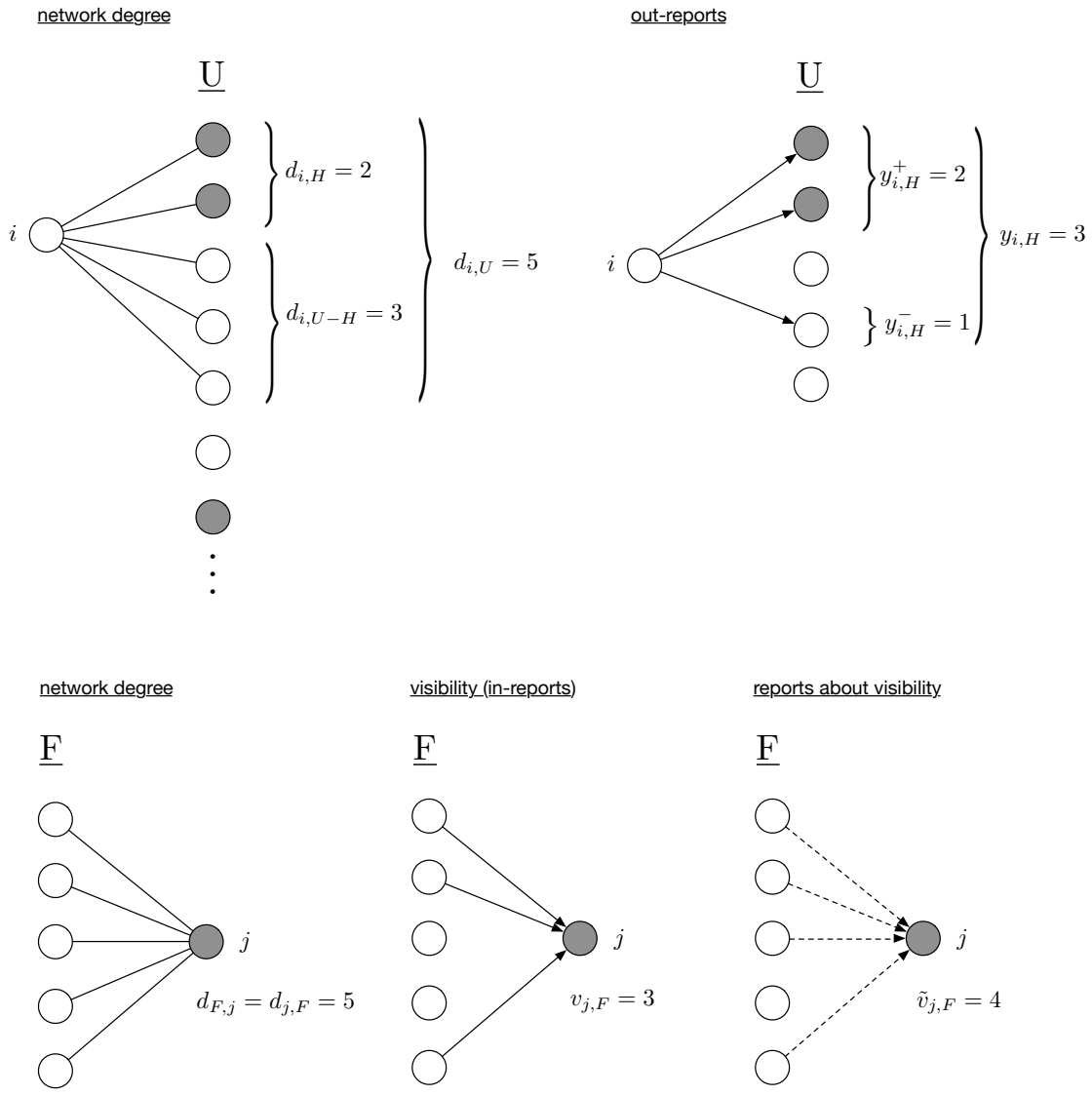


Figure A.1: Illustration of several important pieces of notation used throughout this thesis.

Appendix B

Network reporting appendices

B.1 Estimation with false positive reports

In the main text, we follow all previous scale-up studies to date in assuming that there are never any false positive reports. In this appendix, we generalize our analysis to the situation where false positive reports are possible.

In Chapter 2, we introduce the *precision* of out-reports, which we defined as

$$\eta_F = \frac{\text{\# reported edges from } F \text{ actually connected to } H}{\text{\# reported edges from } F \text{ to } H} = \frac{y^+_{F,H}}{y_{F,H}}. \tag{B.1}$$

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

The precision can be used to relate out-reports to hidden population size, since:

$$N_H = \frac{\eta_F \; y_{F,H}}{\bar{v}_{H,F}}. \tag{B.2}$$

In this appendix, we discuss forming estimators based on η_F. Equation 2.17 suggests the estimator:

$$\widehat{N}_H = \frac{\widehat{\eta}_F \ \widehat{y}_{F,H}}{\widehat{v}_{H,F}}. \tag{B.3}$$

If we could find a consistent and essentially unbiased estimator for η_F, then we could use Equation B.3 to form a consistent and essentially unbiased estimator for N_H, even in the presence of false positive reports.

Unfortunately, we are not aware of a practical strategy for estimating the precision of out-reports. The most direct approach would be to interview each alter that a respondent reports as being in the hidden population. In other words, if a respondent reports knowing 3 drug injectors, researchers could try to interview these three people and see if they are actually drug injectors. Killworth et al. (2006) attempted a version of this procedure, which they called an "alter-chasing" study, but they later abandoned it because of the numerous logistical challenges that arose; see also Laumann (1969) for a related attempt. A second possible approach would be to conduct a census of a networked population where respondents are asked about themselves and specific people to whom they are connected. For example, Goel et al. (2010) collected responses about the political attitudes of thousands of interconnected people on Facebook, including respondents' attitudes as well as their beliefs about specific alters' attitudes. For a subset of respondents, they could compare i's belief about j's attitude with j's report of her own attitude in order to measure the precision. Unfortunately, we think it would be difficult to include a sufficiently large number of members of a stigmatized hidden population in a study such as this.

B.2 Estimates with a sample from F

In this appendix, we present the full results for all of the estimators that require a sample from the frame population. First, we describe the general requirements that our sampling design for F must satisfy (Section B.2.1). Then we describe how to estimate the total number of out-reports, $y_{F,H}$ (Section B.2.2). Next we turn to some background material on multisets (Section B.2.3), which is needed for the following section on the known population method for estimating network degree (Section B.2.4). Finally, we present an estimator for the frame ratio, ϕ_F, which makes use of the known population method results (Section B.2.5).

B.2.1 Requirements for sampling designs from F

We follow Sarndal et al. (2003)'s definition of a probability sampling design, which we repeat here for convenience. Suppose that we have a set of possible samples $\{s_1, \ldots, s_j, \ldots, s_{\max}\}$, with each $s_j \subset F$. Furthermore, suppose $p(s_j)$ gives the probability of selection for each possible sample s_j. If we select a sample s_F at random using a process that will produce each possible sample s_j with probability $p(s_j)$, and if every element $i \in F$ has a nonzero probability of inclusion $\pi_i > 0$, then we will say that we have selected a *probability sample* and we call $p(\cdot)$ the *sampling design.*

B.2.2 Estimating the total number of out-reports, $y_{F,H}$

Estimating the total number of out-reports is a straightforward application of a standard survey estimator, as long as we have a probability sample from the frame.

Result B.2.1 *Suppose we have a sample s_F taken from the frame population using a probability sampling design with probabilities of inclusion given by π_i (Sec. B.2.1).*

Then the estimator given by

$$\widehat{y}_{F,H} = \sum_{i \in s_F} y_{i,H}/\pi_i \tag{B.4}$$

is consistent and unbiased for $y_{F,H}$.

Proof: This follows from the fact that Equation B.4 is a Horvitz-Thompson estimator (Sarndal et al., 2003, Section 2.8). ■

B.2.3 Reporting about multisets

Appendix B.2.4 and Appendix B.3 both describe strategies that involve asking respondents to answer questions about their network alters in specific groups. In this section, we develop the notation and some basic properties of responses generated this way; these properties will be then be used in the subsequent sections.

Suppose we have several groups $A_1, \ldots, A_J$ with $A_j \subset U$ for all j, and also a frame population F of potential interviewees. (Note that we do not require $A_j \subset F$.) Imagine concatenating all of the people in populations $A_1, \ldots, A_J$ together, repeating each individual once for each population she is in. The result, which we call the *probe alters*, $\mathcal{A}$, is a multiset. The size of $\mathcal{A}$ is $N_{\mathcal{A}} = \sum_j N_{A_j}$.

Let y_{i,A_j} be the number of members of group A_j that respondent i reports having among the members of her personal network. We also write $y_{i,\mathcal{A}} = \sum_j y_{i,A_j}$ for the sum of the responses for individual i across all of $A_1, \ldots, A_J$, and $y_{F,\mathcal{A}} = \sum_{i \in F} \sum_j y_{i,A_j}$ to denote the total number of reports from F to $\mathcal{A}$. Similarly, we write $d_{i,\mathcal{A}} = \sum_j d_{i,A_j}$ for the sum of the network connections from individual i to each $A_1, \ldots, A_J$, and $d_{F,\mathcal{A}} = \sum_{i \in F} \sum_j d_{i,A_j}$ for the total of the individual $d_{i,\mathcal{A}}$ taken over all i. As always, we will write averages with respect to the first subscript so that, for example, $\bar{d}_{\mathcal{A},F} = d_{\mathcal{A},F}/N_{\mathcal{A}}$.

We now derive a property of estimation under multisets that will be useful later on. Roughly, this property says that we can estimate the total number of reports from the frame population to the entire multiset of probe alters using only a sample from the frame population with known probabilities of inclusion (Section B.2.1). While this property might seem intuitive, we state it formally for two reasons. First, by stating it explicitly, we make it clear that this property is very general: it does not require any assumptions about the contact pattern between the frame population and probe alters, nor does it require any assumptions about the probe alters. Second, it will turn out to be useful in several later proofs, and so we state it for compactness.

Property B.2.2 *Suppose we have a sample s_F from F taken using a probability sampling design with probabilities of inclusion π_i (Section B.2.1). Then*

$$\widehat{y}_{F,\mathcal{A}} = \sum_{i \in s_F} y_{i,\mathcal{A}}/\pi_i \tag{B.5}$$

is a consistent and unbiased estimator for $y_{F,\mathcal{A}}$.

Proof: If we define $a_i = \sum_j y_{i,A_j}$, the sum of the responses to each A_j for individual i, then we can write our estimator as

$$\widehat{y}_{F,\mathcal{A}} = \sum_{i \in s_F} a_i/\pi_i. \tag{B.6}$$

This is a Horvitz-Thompson esimator (see, e.g., Sarndal et al., 2003, chap. 2); it is unbiased and consistent for the total $\sum_{i \in F} a_i = y_{F,\mathcal{A}}$. ■

B.2.4 Network degree and the known population method for estimating $\bar{d}_{F,F}$, $\bar{d}_{F,U}$, and $\bar{d}_{U,F}$

In order to conduct a scale-up study, we need a definition of the network that we will ask respondents to tell us about; that is, we need to define what it will mean for

two members of the population to be connected by an edge. To date, most scale-up studies have used slight variations of the same definition: the respondent is told that she should consider someone a member of her network if she "knows" the person, where to know someone means (i) you know her and she knows you; (ii) you have been in contact in the past 2 years; and, (iii), if needed, you could get in touch with her (Bernard et al., 2010). Of course, many other definitions are possible, and an investigation of this issue is a matter for future study (Chapter 4). The only restriction on the tie definition we impose here is that it be reciprocal; that is, the definition must imply that if the respondent is connected to someone, then that person is also connected to the respondent.

For a particular definition of a network tie an individual i's degree, $d_{i,U}$ must be conceptually well-defined, but it may not be very easy to directly observe. For the basic scale-up estimator, the most commonly used technique for estimating respondents' network sizes is called the known population method (Killworth et al., 1998a; Bernard et al., 2010).[1] The known population method is based on the idea that we can estimate a respondent's network size by asking how many connections she has to a number of different groups whose sizes are known. The more connections a respondent reports to these groups, the larger we estimate her network to be. Current standard practice is to ask a respondent about her connections to approximately 20 groups of known size in order to estimate her degree (Bernard et al., 2010); it turns out that the exact number of groups used has no impact on the bias of the estimates, as we show in Results B.2.3 and B.2.4.

The known population estimator was originally introduced to estimate the personal network size of each respondent individually (Killworth et al., 1998a), but in

[1]There are other techniques for estimating personal network size, including the summation method (McCarty et al., 2001; Bernard et al., 2010), which could be used in conjunction with many of our results. We focus on the known population method here because it is relatively easy to work with from a statistical perspective, and also because there is some evidence that it works better in practice (Salganik et al., 2011a; Rwanda Biomedical Center, 2012)

Sections 3.2 and 3.4.2 we showed that for the scale-up method the quantity of interest is actually the average number of connections from a member of the frame population F to the rest of the frame population F ($\bar{d}_{F,F}$).[2] This is fortunate, because it is easier to estimate an average degree over all respondents than it is to estimate the individual degree for each respondent.

Guidance for choosing the probe alters, $\mathcal{A}$

Result B.2.3, below, shows that the known population estimator will produce consistent and unbiased estimates of average network degree if (i) $y_{F,\mathcal{A}} = d_{F,\mathcal{A}}$ (*reporting condition*); and (ii) $\bar{d}_{\mathcal{A},F} = \bar{d}_{F,F}$ (*probe alter condition*). Stating these conditions precisely enables us to provide guidance about how the groups of known size $(A_1, A_2, \ldots A_J)$ should be selected such that the probe alters $\mathcal{A}$ will enable consistent and unbiased estimates.

First, the reporting condition ($y_{F,\mathcal{A}} = d_{F,\mathcal{A}}$) in Result B.2.3 shows that researchers should select probe alters such that reporting will be accurate in aggregate. One way to make the reporting condition more likely to hold is to select groups that are unlikely to suffer from transmission error (Shelley et al., 1995, 2006; Killworth et al., 2006; Salganik et al., 2011b; Maltiel et al., 2015). Another way to make the reporting condition more likely to hold is to avoid selecting groups that may lead to recall error (Killworth et al., 2003; Zheng et al., 2006; McCormick and Zheng, 2007; McCormick et al., 2010; Maltiel et al., 2015). That is, previous work suggests that respondents seem to under-report the number of connections they have to large groups, although the precise mechanism behind this pattern is unclear (Killworth et al., 2003). Researchers who have data that may include recall error can consider some of the empirically-calibrated adjustments that have been used in earlier stud-

[2] Although we have framed our discussion here in terms of $\bar{d}_{F,F}$, the same ideas apply to $\bar{d}_{U,F}$ and $\bar{d}_{F,U}$.

ies (Zheng et al., 2006; McCormick and Zheng, 2007; McCormick et al., 2010; Maltiel et al., 2015).

Second, the probe alter condition ($\bar{d}_{\mathcal{A},F} = \bar{d}_{F,F}$) in Result B.2.3 shows that researchers should select groups to be typical of F in terms of their connections to F. In most applied situations, we expect that F will consist of adults, so that researchers should choose groups of known size that are composed of adults, or that are typical of adults in terms of their connections to adults. Further, when trying to choose groups that satisfy the probe alter condition, it is useful to understand how connections from the individual known populations to the frame ($\bar{d}_{A_1,F}, \ldots, \bar{d}_{A_J,F}$) aggregate up into connections from the probe alters to the frame ($\bar{d}_{\mathcal{A},F}$). Basic algebraic manipulation shows that the probe alter condition can be written as:

$$\frac{\sum_j \bar{d}_{A_j,F}\ N_{A_j}}{\sum_j N_{A_j}} = \bar{d}_{F,F}. \tag{B.7}$$

Equation B.7 reveals that the probe alter condition requires that $\bar{d}_{F,F}$ is equal to a weighted average of the average number of connections between each individual known population A_j and the frame population F ($\bar{d}_{A_j,F}$). The weights are given by the size of each known population, N_{A_j}. The simplest way that this could be satisfied is if $\bar{d}_{A_j,F} = \bar{d}_{F,F}$ for every known population A_j. If this is not true, then the probe alter condition can still hold as long as groups for which $\bar{d}_{A_j,F}$ is too high are offset by other groups for which $\bar{d}_{A_{j'},F}$ is too low.

In practice it may be difficult to determine if the reporting condition and probe alter condition will be satisfied. Therefore, we recommend that researchers assess the sensitivity of their size estimates using the procedures described in Appendix B.4. Further, we note that in many realistic situations, N_{A_j} might not be known exactly. Fortunately, researchers only need to know $\sum_j N_{A_j}$, and they can assess the sensitivity

of their estimates to errors in the size of known populations using the procedures described in Appendix B.4.

The known population estimators

Given that background about selecting the probe alters, we present the formal results for the known population estimators for $\bar{d}_{F,F}$, $\bar{d}_{U,F}$, and $\bar{d}_{F,U}$.

Result B.2.3 *Suppose we have a sample s_F taken from the frame population using a probability sampling design with probabilities of inclusion given by π_i (see Section B.2.1). Suppose also that we have a set of known populations, $\mathcal{A}$. Then the known population estimator given by*

$$\widehat{\bar{d}}_{F,F} = \frac{\sum_{i \in s_F} \sum_j y_{i,A_j}/\pi_i}{N_{\mathcal{A}}} \tag{B.8}$$

is consistent and unbiased for $\bar{d}_{F,F}$ if

$$y_{F,\mathcal{A}} = d_{F,\mathcal{A}}, \qquad \textit{(reporting condition)} \tag{B.9}$$

and if

$$\bar{d}_{\mathcal{A},F} = \bar{d}_{F,F}. \qquad \textit{(probe alter condition)} \tag{B.10}$$

Proof: By Property D.1.1, we know that our estimator is unbiased and consistent for $y_{F,\mathcal{A}}/N_{\mathcal{A}}$. By the reporting condition in Equation B.9, this means it is unbiased and consistent for $d_{F,\mathcal{A}}/N_{\mathcal{A}}$. Then, by the probe alter condition in Equation B.10, it is also unbiased and consistent for $\bar{d}_{F,F}$. ■

Result B.2.4 *Suppose we have a sample s_F taken from the frame population using a probability sampling design with probabilities of inclusion given by π_i (see Sec-*

tion B.3.1). Suppose also that we have a set of known populations, $\mathcal{A}$. Then the known population estimator given by

$$\widehat{\bar{d}}_{U,F} = \frac{\sum_{i \in s_F} \sum_j y_{i,A_j}/\pi_i}{N_{\mathcal{A}}} \tag{B.11}$$

is consistent and unbiased for $\bar{d}_{U,F}$ if

$$y_{F,\mathcal{A}} = d_{F,\mathcal{A}}, \qquad \textit{(reporting condition)} \tag{B.12}$$

and if

$$\bar{d}_{\mathcal{A},F} = \bar{d}_{U,F}. \qquad \textit{(probe alter condition)} \tag{B.13}$$

Proof: By Property D.1.1, we know that our estimator is unbiased and consistent for $y_{F,\mathcal{A}}/N_{\mathcal{A}}$. By the reporting condition in Equation B.12, this means it is unbiased and consistent for $d_{F,\mathcal{A}}/N_{\mathcal{A}}$. Then, by the probe alter condition in Equation B.13, it is also unbiased and consistent for $\bar{d}_{U,F}$. ■

Since $\bar{d}_{F,U} = \frac{N}{N_F}\bar{d}_{U,F}$, as a direct consequence of Result B.2.4 we have the following corollary.

Corollary B.2.5 *If the conditions described in Result B.2.4 hold,*

$$\widehat{\bar{d}}_{F,U} = \widehat{\bar{d}}_{U,F}\ \frac{N}{N_F} \tag{B.14}$$

is consistent and unbiased for $\bar{d}_{F,U}$.

B.2.5 Estimating the frame ratio, ϕ_F

Given our estimator of $\bar{d}_{F,F}$ (Result B.2.3) and our estimator of $\bar{d}_{U,F}$ (Result B.2.4), we can estimate the frame ratio, ϕ_F.

Result B.2.6 *The estimator*

$$\widehat{\phi}_F = \frac{\widehat{\bar{d}}_{F,F}}{\widehat{\bar{d}}_{U,F}} \tag{B.15}$$

is consistent and essentially unbiased for ϕ_F *if* $\widehat{\bar{d}}_{F,F}$ *is consistent and essentially unbiased for* $\bar{d}_{F,F}$ *and* $\widehat{\bar{d}}_{U,F}$ *is consistent and essentially unbiased for* $\bar{d}_{U,F}$.

Proof: This follows from the properties of a ratio estimator (Sarndal et al., 2003, chap. 5). ■

More concretely, combining the estimator for $\bar{d}_{F,F}$ (Result B.2.3) and the estimator for $\bar{d}_{U,F}$ (Result B.2.4), and assuming that we have known populations $\mathcal{A}_{F_1}$ for $\bar{d}_{F,F}$, and $\mathcal{A}_{F_2}$ for $\bar{d}_{U,F}$, we obtain

$$\widehat{\phi}_F = \frac{N_{\mathcal{A}_{F_2}}}{N_{\mathcal{A}_{F_1}}} \frac{\sum_{i \in s_F} \sum_{A_j \in \mathcal{A}_{F_1}} y_{i,A_j}/\pi_i}{\sum_{i \in s_F} \sum_{A_k \in \mathcal{A}_{F_2}} y_{i,A_k}/\pi_i}. \tag{B.16}$$

In our discussion of $\widehat{\bar{d}}_{F,F}$ (Result B.2.3) and $\widehat{\bar{d}}_{U,F}$ (Result B.2.4), we concluded that we want the known populations $\mathcal{A}_{F_1}$ used for $\widehat{\bar{d}}_{F,F}$ to be typical of members of F in their connections to F. An analogous argument shows that we want the known populations $\mathcal{A}_{F_2}$ used for $\widehat{\bar{d}}_{U,F}$ to be typical of members of U in their connections to F. In general, we expect that it will not be appealing to assume that F and U are similar to each other in terms of their connections to F meaning that, unfortunately, it will not make sense to use the same set of known populations for $\widehat{\bar{d}}_{F,F}$ and $\widehat{\bar{d}}_{U,F}$. If researchers wish to estimate ϕ_F directly, one approach would be to choose $\mathcal{A}_{F_2}$ to be typical of U in such a way that some of the individual known populations are more typical of F, while others more typical of $U - F$. The multiset formed from only the ones that are more typical of F could then be our choice for $\mathcal{A}_{F_1}$. In this case, researchers would also want $\frac{N_{\mathcal{A}_{F_1}}}{N_{\mathcal{A}_{F_2}}} \approx \frac{N_F}{N}$. This complication is one of the reasons we

recommend in Section 3.4 that future scale-up studies estimate $\bar{d}_{F,F}$ directly, thus avoiding the need to estimate ϕ_F entirely.

B.3 Estimates with samples from F and H

In this appendix, we present the full results for all of the estimators that require a sample from the hidden population. First, we define the general requirements that our sampling design for H must satisfy (Section B.3.1). Then we describe a flexible data collection procedure called the game of contacts (Section B.3.2). Next, we introduce some background material on estimation using questions about multisets (Section B.3.3) and present an estimator for $\bar{v}_{H,F}$, the average number of in-reports among the members of the hidden population (Section B.3.4). Then, we present estimators for the two adjustment factors introduced in Section 3.2: the degree ratio, δ_F, and the true positive rate, τ_F (Section B.3.6). Finally, we present formal results for four different estimators for N_H (Section B.3.7).

B.3.1 Requirements for sampling designs from H

For the results that involve a sample from the hidden population s_H, we do not need a probability sample (Appendix B.2); instead, we need a weaker type of design. We require that every element $i \in H$ have a nonzero probability of selection $\pi_i > 0$, and that we can determine the probability of selection up to a constant factor c; that is, we only need to know $c\pi_i$. We are not aware of any existing name for this situation, so we will call it a *relative probability sample*. Because of the challenges involved in sampling hard-to-reach populations, the two most likely sampling designs for s_H will probably be time-location sampling (Karon and Wejnert, 2012) and respondent-driven sampling (Heckathorn, 1997). A relative probability sample allows us to use weighted sample means to estimate averages, but not totals. See Sarndal et al. (2003,

Section 5.7) for more details on weighted sample means, also sometimes called Hajek estimators, which is what we use to estimate averages from a sample of hidden population members.

B.3.2 Data collection

In order to make estimates about the hidden population's visibility to the frame population, researchers will need to collect what we call *enriched aggregate relational data* from each respondent, and a procedure called the *game of contacts* has produced promising results from a study of heavy drug users in Brazil (Salganik et al., 2011b). In the main text, we assumed that the groups in the probe alters $A_1, \ldots, A_J$ were all contained in the frame population ($A_j \subset F$ for all j). However, the estimators in this Appendix are more general because they allow for the possibility that some of the groups $A_1, \ldots A_J$ may not be contained entirely in F. For example, if the frame population is adults, then this flexibility enables researchers to use groups based on names, such as Michael, even though not all people named Michael are adults.

In order to allow for this flexibility, we need to introduce some new notation: let $A_1 \cap F, A_2 \cap F, \ldots, A_J \cap F$ be the intersection of these groups and the frame population, and let $\mathcal{A} \cap F$ be the concatenation of these intersected groups. For example, if the frame population is adults, A_1 is people named Michael, and A_2 is doctors, then $A_1 \cap F$ is adults named Michael, $A_2 \cap F$ is adult doctors, and $\mathcal{A} \cap F$ is the collection of all adult Michaels and all adult doctors, with adult doctors named Michael included twice. (In the special case discussed in the main text, $A_1 \cap F, \ldots A_J \cap F = A_1, \ldots, A_J$.)

The data collection begins with a relative probability sample (Section B.3.1) from the hidden population. For a set of groups, $A_1, A_2, \ldots A_J$, each respondent in the hidden population is asked, "How many people do you know in group A_j?" We call the response y_{i,A_j}. Next for each of the y_{i,A_j} alters, the respondent picks up a token and places it on a game board like the one in Figure B.1. From the location of the

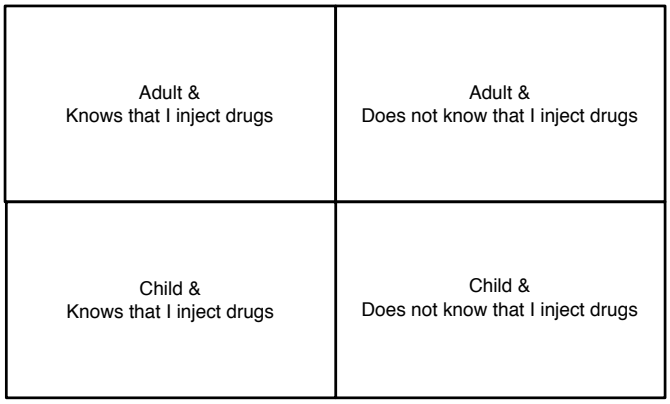


Figure B.1: Example of a game board that could be used in the game of contacts interviewing procedure if the hidden population was people who inject drugs and the frame was made up of adults. This board is a variation of the board used in Salganik et al. (2011b).

tokens on the board, the researcher can record whether each alter is in the frame population (or not) and whether the alter is aware that the respondent is in the hidden population (or not). This process is then repeated until the respondent has been asked about all groups.

If all of the probe alters are in the frame population, then the process is much easier for respondents and the game board can be modified to collect alternative information. If all of the probe alters are not in the frame population, then it is important for the researcher to define the frame population as clearly as possible. If the respondents are not able to correctly indicate whether the alters are in the frame population or not, it could lead to biased estimates of $\bar{v}_{H,F}$. For more on the operational implementation of this procedure, see Salganik et al. (2011b).

B.3.3 Estimation using aggregated relational data from the hidden population

In this section, we follow Section B.2.3 and present another useful property about estimates made using aggregate relational data from the hidden population. Roughly, this property says that we can estimate the average number of reports from the entire

	aware	not aware	total
frame population	$\widetilde{v}_{i,A_j \cap F}$	$\widetilde{h}_{i,A_j \cap F}$	$y_{i,A_j \cap F}$
not frame population	$\widetilde{v}_{i,A_j \cap (U-F)}$	$\widetilde{h}_{i,A_j \cap (U-F)}$	$y_{i,A_j \cap (U-F)}$
total	$\widetilde{v}_{i,A_j}$	$\widetilde{h}_{i,A_j}$	y_{i,A_j}

Table B.1: Responses collected during the game of contacts for each respondent i and each group A_j. We use $\widetilde{\ }$ to indicate reported values. For example, $\widetilde{v}_{i,A_j}$ is the respondent's reported visibility to people in A_j and v_{i,A_j} is respondent's actual visiblility to people in A_j. Also, using this notational convention, it is the case that $y_{i,A_j} = \widetilde{d}_{i,A_j}$, but we have written y_{i,A_j} in order to be consistent with the rest of the paper.

hidden population to the probe alters using only a relative probability sample from the hidden population (Section B.3.1). Similar to Property D.1.1, the result we present below does not require any assumptions about the contact pattern between the hidden population and the probe alters, nor about the probe alters themselves.

Property B.3.1 *Suppose we have a sample s_H from H taken using a relative probability design, allowing us to compute the relative probabilities of inclusion $c\pi_i$ for all sampled elements (Sec. B.3.1). Then*

$$\widehat{\bar{y}}_{H,\mathcal{A}} = \frac{\sum_{i \in s_H} y_{i,\mathcal{A}}/(c\pi_i)}{\sum_{i \in s_H} 1/(c\pi_i)} \tag{B.17}$$

is a consistent and essentially unbiased estimator for $\bar{y}_{H,\mathcal{A}} = y_{H,\mathcal{A}}/N_H$.

Proof: Note that the c in the relative probabilities of inclusion $c\pi_i$ cancel, so that

$$\widehat{\bar{y}}_{H,\mathcal{A}} = \frac{\sum_{i \in s_H} y_{i,\mathcal{A}}/(\pi_i)}{\sum_{i \in s_H} 1/(\pi_i)}. \tag{B.18}$$

If we define $a_i = \sum_j y_{i,A_j}$, the sum of the responses to each A_j for individual i, then we can write our estimator as

$$\widehat{\bar{y}}_{H,\mathcal{A}} = \frac{\sum_{i \in s_H} a_i/\pi_i}{\sum_{i \in s_H} 1/\pi_i}. \tag{B.19}$$

Now we have a standard weighted mean estimator (e.g. Sarndal et al., 2003, chap. 5); it is consistent and essentially unbiased for the average $\frac{1}{N_H}\sum_{i\in H} a_i = y_{H,\mathcal{A}}/N_H$. ■

B.3.4 Estimating the average visibility, $\bar{v}_{H,F}$

Given the data collection procedure described in Sec. B.3.2, we can estimate the average visibility ($\bar{v}_{H,F}$) as long as two conditions are satisfied: one about reporting and one about the visibility of the hidden population to the probe alters.

Result B.3.2 *Assume that we have a sample s_H taken from the hidden population using a relative probability design with relative probabilities of inclusion $c\pi_i$ for all sampled elements (Sec. B.3.1). Then*

$$\widehat{\widetilde{v}}_{H,F} = \frac{N_F}{N_{\mathcal{A}\cap F}} \frac{\sum_{i\in s_H}\sum_j \widetilde{v}_{i,A_j\cap F}/(c\pi_i)}{\sum_{i\in s_H} 1/(c\pi_i)} \tag{B.20}$$

is consistent and essentially unbiased for $\bar{v}_{H,F}$ if

$$\widetilde{v}_{H,\mathcal{A}\cap F} = v_{H,\mathcal{A}\cap F}, \qquad \textit{(reporting condition)} \tag{B.21}$$

and

$$\frac{v_{H,\mathcal{A}\cap F}}{N_{\mathcal{A}\cap F}} = \frac{v_{H,F}}{N_F}. \qquad \textit{(probe alter condition)} \tag{B.22}$$

Proof: Property B.3.1 holds for estimating $\overline{\widetilde{v}}_{F,\mathcal{A}\cap F}$ from $\widetilde{v}_{i,\mathcal{A}\cap F}$, just as it holds for estimating $\bar{y}_{H,\mathcal{A}\cap F}$ from $y_{i,\mathcal{A}\cap F}$. Applying Property B.3.1 here, we conclude that the estimator is consistent and essentially unbiased for

$$\frac{N_F}{N_{\mathcal{A}\cap F}}\overline{\widetilde{v}}_{H,\mathcal{A}\cap F} = \frac{N_F}{N_{\mathcal{A}\cap F}}\frac{\widetilde{v}_{H,\mathcal{A}\cap F}}{N_H}. \tag{B.23}$$

Next, by applying the reporting condition in Equation B.21 we can conclude that

$$\frac{N_F}{N_{\mathcal{A}\cap F}}\frac{\widetilde{v}_{H,\mathcal{A}\cap F}}{N_H} = \frac{N_F}{N_{\mathcal{A}\cap F}}\frac{v_{H,\mathcal{A}\cap F}}{N_H}. \tag{B.24}$$

Finally, by applying the probe alter condition in Equation B.22 and rearranging terms, we conclude that

$$\frac{N_F}{N_{\mathcal{A}\cap F}}\frac{v_{H,\mathcal{A}\cap F}}{N_H} = \frac{N_F}{N_H}\frac{v_{H,F}}{N_F} \tag{B.25}$$

$$= \bar{v}_{H,F} \tag{B.26}$$

■

Note that Result B.3.2 requires us to know the size of the probe alters in the frame population, $N_{\mathcal{A}\cap F}$. In some cases, this may not be readily available, but it may be reasonable to assume that

$$N_{\mathcal{A}\cap F} = \frac{N_F}{N}\ N_{\mathcal{A}}. \tag{B.27}$$

Furthermore, if $\mathcal{A}$ is chosen so that all of its members are in F, then $N_{\mathcal{A}\cap F} = N_{\mathcal{A}}$ and $v_{i,A_j\cap F} = v_{i,A_j}$. In this situation, we do not need to specifically ask respondents about connections to $\mathcal{A}\cap F$; we can just ask about connections to $\mathcal{A}$.

The reporting condition required for Result B.21 states that the hidden population's total reported visibility from the probe alters on the frame must be correct. This might not be the case, if for example, respondents systematically over-estimate how much others know about them (see e.g., Gilovich et al. (1998)). The required condition for the probe alters is slightly more complex. It needs to be the case that the rate at which the hidden population is visible to the probe alters is the same as the rate at which the hidden population is visible to the frame population. There are several equivalent ways of stating this condition, as we show in a moment. First,

we need to define two new quantities: the individual-level true positive rate and the average of the individual-level true positive rates.

Definition 1 *We define the individual-level true positive rate for respondent $i \in F$ to be*

$$\tau_i = \frac{v_{H,i}}{d_{i,H}}, \tag{B.28}$$

where $v_{H,i} = \sum_{j \in H} v_{j,i}$.

Definition 2 *We define the average of the individual true positive rates over a set F of respondents as*

$$\overline{\tau}_F = \frac{1}{N_F} \sum_{i \in F} \tau_i. \tag{B.29}$$

In general, $\overline{\tau}_F \neq \tau_F$. To see this, note that while $\overline{\tau}_F$ is the average of the individual-level true positive rates with each individual weighted equally, τ_F can be written as the weighted average of the individual true positive rates, with the weights given by each individual's degree. We can see the exact relationship between the two by expressing τ_F in terms of the τ_i:

$$\tau_F = \frac{\sum_{i \in F} \tau_i \; d_{i,H}}{\sum_{i \in F} d_{i,H}}, \tag{B.30}$$

since multiplying each τ_i by $d_{i,H}$ and summing is the same as summing the $v_{H,i}$.

Result B.3.3 *The following conditions are all equivalent.*

(i) $\frac{v_{H,\mathcal{A} \cap F}}{N_{\mathcal{A} \cap F}} = \frac{v_{H,F}}{N_F}$

(ii) $\tau_{\mathcal{A} \cap F} \; \bar{d}_{\mathcal{A} \cap F,H} = \tau_F \; \bar{d}_{F,H}$

(iii) $\overline{\tau}_{\mathcal{A} \cap F} \; \bar{d}_{\mathcal{A} \cap F,H} + cov_{\mathcal{A} \cap F}(\tau_i, d_{i,H}) = \overline{\tau}_F \; \bar{d}_{F,H} + cov_F(\tau_i, d_{i,H})$

(iv) $\bar{y}^{+}_{F,H} = \frac{\sum_j \bar{y}^{+}_{A_j \cap F,H} \; N_{A_j \cap F}}{\sum_j N_{A_j \cap F}}$,

where cov_F *is the finite-population covariance taken over the set* F.[3]

Proof: First, we show that

$$\tau_{\mathcal{A} \cap F} \; \bar{d}_{\mathcal{A} \cap F,H} = \tau_F \; \bar{d}_{F,H} \Longleftrightarrow \frac{v_{H,\mathcal{A} \cap F}}{N_{\mathcal{A} \cap F}} = \frac{v_{H,F}}{N_F}. \tag{B.31}$$

By definition, $\tau_F \; \bar{d}_{F,H} = (v_{H,F}/d_{F,H}) \times (d_{F,H}/N_F) = v_{H,F}/N_F$. The same argument demonstrates that $\tau_{\mathcal{A} \cap F} \; \bar{d}_{\mathcal{A} \cap F,H} = v_{H,\mathcal{A} \cap F}/N_{\mathcal{A}}$. We conclude that $(i) \Longleftrightarrow (ii)$.

Next, we show that (ii) is equivalent to (iii). We can use the relationship between τ_F and the τ_i, Equation B.30, to deduce that

$$\tau_F \; d_{F,H} = \sum_{i \in F} \tau_i \; d_{i,H} = N_F \; [\overline{\tau}_F \; \bar{d}_{F,H} + \mathrm{cov}_F(\tau_i, d_{i,H})]. \tag{B.32}$$

Dividing the left-most and right-most sides by N_F, we conclude that

$$\tau_F \; \bar{d}_{F,H} = \overline{\tau}_F \; \bar{d}_{F,H} + \mathrm{cov}_F(\tau_i, d_{i,H}). \tag{B.33}$$

The same argument shows that

$$\bar{d}_{\mathcal{A} \cap F,H} \; \tau_{\mathcal{A} \cap F} = \overline{\tau}_{\mathcal{A} \cap F} \; \bar{d}_{\mathcal{A} \cap F,H} + \mathrm{cov}_{\mathcal{A} \cap F}(\tau_i, d_{i,H}). \tag{B.34}$$

So we conclude that $(ii) \Longleftrightarrow (iii)$.

Finally, we show that (iv) is equivalent to (i). In Chapter 2, we showed that $\bar{y}^{+}_{F,H} = v_{H,F}/N_H$, which is the right-hand side of the identity in (i) (Equation 2.14).

[3] We define the finite-population covariance to have a denominator of N_F; this differs from some other authors, who define the finite-population covariance to have $N_F - 1$ in the denominator.

Similarly, starting with the left-hand side of the identity in (i), we have

$$\frac{v_{H,\mathcal{A}\cap F}}{N_{\mathcal{A}\cap F}} = \frac{\sum_j v_{H,A_j\cap F}}{\sum_j N_{A_j\cap F}} = \frac{\sum_j y^+_{A_j\cap F,H}}{\sum_j N_{A_j\cap F}} = \frac{\sum_j \bar{y}^+_{A_j\cap F,H}\ N_{A_j\cap F}}{\sum_j N_{A_j\cap F}}. \tag{B.35}$$

So we conclude that $(i) \iff (iv)$.

Since $(i) \iff (ii)$ and $(ii) \iff (iii)$, it follows that $(i) \iff (iii)$. Furthermore, since $(i) \iff (iv)$, it follows that (iv) is equivalent to (ii) and (iii). ■

Result B.3.3 shows that the probe alter condition can be expressed in many equivalent ways. One of these alternate expressions is especially useful because it leads to an empirical check of the probe alter condition that future scale-up studies can implement. This empirical check is a direct consequence of Result B.3.4, below. Intuitively, Result B.3.4 and the empirical check are a consequence of the identity in Equation 2.13, which says that in-reports from the perspective of H are also out-reports from the perspective of F.

Result B.3.4 *Suppose that the precision of out-reports from the frame population is the same as the precision of the out-reports from $\mathcal{A} \cap F$:*

$$\frac{y^+_{F,H}}{y_{F,H}} = \frac{y^+_{\mathcal{A}\cap F,H}}{y_{\mathcal{A}\cap F,H}} \tag{B.36}$$

Then the probe alter condition (B.22) is satisfied if and only if

$$\bar{y}_{F,H} = \bar{y}_{\mathcal{A}\cap F,H}. \tag{B.37}$$

Proof: First, note that, by Result B.3.3, the probe alter condition is equivalent to

$$\bar{y}^+_{F,H} = \frac{\sum_j \bar{y}^+_{A_j\cap F,H}\ N_{A_j\cap F}}{\sum_j N_{A_j\cap F}}. \tag{B.38}$$

Since $\bar{y}^{+}_{A_j \cap F,H} = y^{+}_{A_j \cap F,H}/N_{A_j \cap F}$ for all j, the right-hand side of Equation B.38 is equal to $\bar{y}^{+}_{\mathcal{A} \cap F,H}$, meaning that the probe alter condition is also equivalent to

$$\bar{y}^{+}_{F,H} = \bar{y}^{+}_{\mathcal{A} \cap F,H}. \tag{B.39}$$

Second, note that the assumption in Equation B.36 can be re-written as

$$\frac{\bar{y}^{+}_{F,H}}{\bar{y}_{F,H}} = \frac{\bar{y}^{+}_{\mathcal{A} \cap F,H}}{\bar{y}_{\mathcal{A} \cap F,H}}, \tag{B.40}$$

by multiplying the left-hand side by $\frac{N_F}{N_F}$ and the right-hand side by $\frac{N_{\mathcal{A} \cap F}}{N_{\mathcal{A} \cap F}}$. So we are left with the task of showing that if Equation B.40 is true, then Equation B.39 is satisfied if and only if Equation B.37 is satisfied. But this is the case, since Equation B.39 equates the numerators of the two fractions in Equation B.40 and Equation B.37 equates the denominators of the two fractions in Equation B.40. Two fractions that are equal will have equal numerators if and only if they have equal denominators. (Formally, if $a/b = c/d$ then $a = c$ if and only if $b = d$.) ■

The implication of Result B.3.4 is that if (i) researchers design the probe alters so that the frame population sample s_F can be used to estimate $\bar{y}_{\mathcal{A} \cap F,H}$; and (ii) researchers assume that the precision of out-reports from the frame population is the same as the precision of out-reports from $\mathcal{A} \cap F$, then they can evaluate how well the probe alter condition is satisfied empirically by comparing $\widehat{\bar{y}}_{F,H}$ and $\widehat{\bar{y}}_{\mathcal{A} \cap F,H}$.

Finally, we can foresee three practical problems that might arise when researchers try to estimate $\bar{v}_{H,F}$. First, researchers might not be able to choose the probe alters to satisfy the probe alter condition (Equation B.22) because of limited information about the true visibility of the hidden population with respect to different social groups. A second problem might arise if researchers are not able to choose the probe alters to satisfy the reporting condition (Equation B.21) because of limited information about the hidden population's awareness about visibility. Finally, a third problem might

arise due to errors in administrative records which would cause researchers to have incorrect information about the size of the multiset of probe alters on the frame ($N_{\mathcal{A}\cap F}$). Fortunately, as we show in Result B.3.5, it is possible to quantify the effect of these problems on the resulting estimates. In some cases they can cancel out, but in other cases they magnify each other.

Result B.3.5 *Suppose that $\widehat{N}_{\mathcal{A}\cap F}$, the researcher's estimate of $N_{\mathcal{A}\cap F}$, is incorrect, so that $\widehat{N}_{\mathcal{A}\cap F} = c_1 \cdot N_{\mathcal{A}\cap F}$. Suppose also that the reporting condition (Equation B.21) of Result B.3.2 is incorrect, so that $\tilde{v}_{H,\mathcal{A}\cap F} = c_2 \cdot v_{H,\mathcal{A}\cap F}$. Finally, suppose that the probe alter condition is incorrect, so that $\frac{v_{H,A\cap F}}{N_{\mathcal{A}\cap F}} = c_3 \cdot \frac{v_{H,F}}{N_F}$. Call the estimator under these imperfect conditions $\widehat{\bar{v}}^{\star}_{H,F}$. Then $\widehat{\bar{v}}^{\star}_{H,F}$ is consistent and essentially unbiased for $\frac{c_3\ c_2}{c_1}\bar{v}_{H,F}$ instead of $\bar{v}_{H,F}$.*

Proof: Under the assumptions listed above, we can write the new estimator as

$$\widehat{\bar{v}}^{\star}_{F,H} = \frac{1}{c_1}\ \frac{N_F}{N_{\mathcal{A}\cap F}}\ \frac{\sum_{i\in s_H}\sum_j \widetilde{v}_{i,A_j\cap F}/(c\pi_i)}{\sum_{i\in s_H} 1/(c\pi_i)}. \tag{B.41}$$

We follow the same steps as the proof of Result B.3.2, but each time we use one of our assumptions, the associated error is carried with it. So our estimator $\widehat{\bar{v}}^{\star}_{F,H}$ is consistent and essentially unbiased for

$$\frac{1}{c_1}\ \frac{N_F}{N_{\mathcal{A}\cap F}}\ \frac{\tilde{v}_{H,\mathcal{A}\cap F}}{N_H} = \frac{c_2}{c_1}\ \frac{N_F}{N_{\mathcal{A}\cap F}}\ \frac{v_{H,\mathcal{A}\cap F}}{N_H} = \frac{c_3\ c_2}{c_1}\ \frac{N_F}{N_{\mathcal{A}\cap F}}\ \frac{v_{H,F}}{N_H}. \tag{B.42}$$

In words, the estimand is now incorrect by $\frac{c_3\ c_2}{c_1}$. Since $\widehat{\bar{v}}_{F,H}$ is consistent and essentialy unbiased for $\bar{v}_{F,H}$, we conclude that $\widehat{\bar{v}}^{\star}_{F,H}$ is consistent and essentially unbiased for $\frac{c_3\ c_2}{c_1}\bar{v}_{F,H}$. Note that if the assumptions needed for Result B.3.2 hold, then $c_1 = 1$, $c_2 = 1$, and $c_3 = 1$, giving us our original result. ■

We present similar sensitivity results for all of our estimators in Appendix B.4.

B.3.5 Guidance for choosing the probe alters for the game of contacts, $\mathcal{A}$

Turning the results in Appendix B.3 into easy to follow steps for selecting the probe alters for the game of contacts is an open and important research problem. Here, we briefly offer three recommendations for selecting the probe alters for the game of contacts. We realize that these recommendations may be difficult to follow exactly in practice. Therefore, we also discuss the sensitivity of the estimators to errors in the construction of the probe alters. Finally, we discuss one type of data that should be collected from the frame population in order to help the researchers evaluate their choice of probe alters for the game of contacts.

First, we recommend that probe alters for the game of contacts be in the frame population. For example, if the frame population is adults, we recommend that all members of the probe alters be adults. This choice will simplify the data collection task in the game of contacts, and for all the advice listed below, we assume that it has been followed. If it is not possible, researchers can still use the more general procedures developed in this Appendix.

Second, we recommend that the probe alters be selected such that the probe alter condition in Result B.3.2 is satisfied. That is, the probe alters as a whole should be typical of the frame population in the following way: it should be the case that the rate at which the hidden population is visible to the probe alters is the same as the rate at which the hidden population is visible to the frame population ($\frac{v_{H,\mathcal{A}}}{N_{\mathcal{A}}} = \frac{v_{H,F}}{N_F}$). For example, in a study to estimate the number of drug injectors in a city, drug treatment counselors would be a poor choice for membership in the probe alters because drug injectors are probably more visible to drug treatment counselors than to typical members of the frame population. On the other hand, postal workers would probably be a reasonable choice for membership in the probe alters because

drug injectors are probably about as visible to postal workers as they are to typical members of the frame population.

Third, we recommend that the probe alters be selected so that the reporting condition in Result B.3.2 is satisfied ($\widetilde{v}_{H,\mathcal{A}} = v_{H,\mathcal{A}}$). One way to help ensure that this condition holds is to avoid selecting large groups that may cause recall error (Killworth et al., 2003; Zheng et al., 2006; McCormick and Zheng, 2007; McCormick et al., 2010; Maltiel et al., 2015).

In practice it might be difficult to meet each of these three conditions exactly, therefore we recommend a sensitivity analysis using the results in Appendix B.4.

Finally, the choice of probe alters for the game of contacts also has two implications for the design of the survey of the frame population. First, if researchers wish to estimate the degree ratio, δ_F, then they should design the probe alters $\mathcal{A}$ so that they can be asked of both members of the hidden population sample and members of the frame population sample (see Result B.3.7). Second, if researchers wish to test the probe alter condition using the approach in Result B.3.4, then additional information needs to be collected from each member of the frame population sample. For example, if one group in the probe alters for the game of contacts is postal workers, then members of the frame population sample should be asked if they are postal workers.

B.3.6 Term-by-term: δ_F and τ_F

In this section we describe how to estimate two adjustment factors: the degree ratio,

$$\delta_F = \frac{\bar{d}_{H,F}}{\bar{d}_{F,F}} \tag{B.43}$$

and the true positive rate,

$$\tau_F = \frac{\bar{v}_{H,F}}{\bar{d}_{H,F}}. \tag{B.44}$$

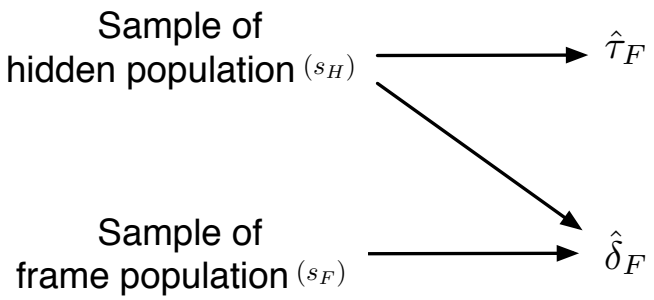


Figure B.2: We estimate the true positive rate $\widehat{\tau}_F$ using data from the survey of the hidden population, and we estimate the degree ratio $\widehat{\delta}_F$ using the sample of the hidden population and the sample of the frame population.

Estimating the degree ratio requires information from the survey of the hidden population and the survey of the frame population, while estimating the true positive rate only requires information from the survey of the hidden population (Fig. B.2).

As Equations B.43 and B.44 make clear, both adjustment factors involve $\bar{d}_{H,F}$ so we first present an estimator for that quantity.

Result B.3.6 *Suppose we have a sample s_H taken from the hidden population using a relative probability sampling design with relative probabilities of inclusion denoted $c\pi_i$ (Sec B.3.1). Then the estimator given by*

$$\widehat{\bar{d}}_{H,F} = \frac{N_F}{N_{\mathcal{A}\cap F}} \frac{\sum_{i\in s_H} \sum_j y_{i,(A_j\cap F)}/(c\pi_i)}{\sum_{i\in s_H} 1/(c\pi_i)} \tag{B.45}$$

is consistent and essentially unbiased for $\bar{d}_{H,F}$ if:

$$y_{H,\mathcal{A}\cap F} = d_{H,\mathcal{A}\cap F}, \qquad \textit{(reporting condition)} \tag{B.46}$$

and

$$\bar{d}_{\mathcal{A}\cap F,H} = \bar{d}_{F,H}. \qquad \textit{(probe alter condition)} \tag{B.47}$$

Proof: From Property B.3.1, we can see that our estimator is consistent and essentially unbiased for

$$\frac{N_F}{N_{\mathcal{A}\cap F}}\frac{y_{H,\mathcal{A}\cap F}}{N_H} = \frac{N_F}{N_H}\frac{y_{H,\mathcal{A}\cap F}}{N_{\mathcal{A}\cap F}}. \tag{B.48}$$

Under the reporting condition (Equation B.46) this becomes

$$\frac{N_F}{N_H}\frac{y_{H,\mathcal{A}\cap F}}{N_{\mathcal{A}\cap F}} = \frac{N_F}{N_H}\frac{d_{H,\mathcal{A}\cap F}}{N_{\mathcal{A}\cap F}} \tag{B.49}$$

Finally, applying the probe alter condition in Equation B.47, we have

$$\frac{N_F}{N_H}\frac{d_{H,\mathcal{A}\cap F}}{N_{\mathcal{A}\cap F}} = \frac{N_F}{N_H}\frac{d_{F,H}}{N_F} \tag{B.50}$$

$$= \bar{d}_{H,F}. \tag{B.51}$$

■

Result B.3.6 requires that reports are, in total, correct (Equation B.46). Like Result B.3.2, Result B.3.6 also requires us to know the size of the probe alters on the frame, $N_{\mathcal{A}\cap F}$. In some cases, this may not be readily available, but it may be reasonable to assume that

$$N_{\mathcal{A}\cap F} = \frac{N_F}{N}\ N_{\mathcal{A}}. \tag{B.52}$$

Furthermore, if $\mathcal{A}$ is chosen so that all of its members are in F, then $N_{\mathcal{A}\cap F} = N_{\mathcal{A}}$ and $y_{i,A_j\cap F} = y_{i,A_j}$. In this situation, we do not need to specifically ask respondents about connections to $\mathcal{A}\cap F$; we can just ask about connections to $\mathcal{A}$. Result B.3.6 also requires a specific rate of connectivity between the probe alters and the hidden population (Equation B.47). We discussed some of the consequences of these assump-

tion in the main text, where we reviewed the implications of our results for practice (Section 3.4).

Estimating the degree ratio, δ_F

We can combine our estimator for $\bar{d}_{H,F}$ (Result B.3.6) and our estimator for $\bar{d}_{F,F}$ (Result B.2.3), to estimate the degree ratio, δ_F.

Result B.3.7 *The estimator*

$$\widehat{\delta}_F = \frac{\widehat{\bar{d}}_{H,F}}{\widehat{\bar{d}}_{F,F}} \tag{B.53}$$

is consistent and essentially unbiased for δ_F if $\widehat{\bar{d}}_{H,F}$ is consistent and essentially unbiased for $\bar{d}_{H,F}$ and $\widehat{\bar{d}}_{F,F}$ is consistent and essentially unbiased for $\bar{d}_{F,F}$.

Proof: This follows from the properties of a compound ratio estimator (Appendix B.5). ■

More concretely, combing the estimators in Result B.3.6 and Result B.2.3, results in an estimator for $\widehat{\delta}_F$ with the following form:

$$\widehat{\delta}_F = \frac{\frac{N_F}{N_{\mathcal{A}_H \cap F}} \frac{\sum_{i \in s_H} \sum_{A_j \in \mathcal{A}_H} y_{i,(A_j \cap F)}/(c\pi_i^H)}{\sum_{i \in s_H} 1/(c\pi_i^H)}}{\frac{1}{N_{\mathcal{A}_F}} \sum_{i \in s_F} \sum_{A_k \in \mathcal{A}_F} y_{i,A_k}/\pi_i^F}. \tag{B.54}$$

If the probe alters for the frame population and the hidden population are the same, so that $\mathcal{A}_H = \mathcal{A}_F = \mathcal{A}$, and if the probe alters are randomly distributed in the frame population in the sense that

$$N_{\mathcal{A} \cap F} = N_{\mathcal{A}} \, \frac{N_F}{N}, \tag{B.55}$$

then we can reduce the constants in front of Equation B.54 to

$$\frac{\frac{N_F}{N_{\mathcal{A}\cap F}}}{\frac{1}{N_{\mathcal{A}}}} = \frac{\frac{N}{N_{\mathcal{A}}}}{\frac{1}{N_{\mathcal{A}}}} = N. \tag{B.56}$$

In other words, when the probe alters for the frame and hidden population are the same, and when the probe alters are randomly distributed in the frame population, all of the factors involving the size of $\mathcal{A}$ drop out. This fact allows researchers to use groups defined by first names (e.g., people named Michael) in the probe alters $\mathcal{A}$, even if the size of these groups is not known, as long as it is reasonable to assume that $\mathcal{A}$ satisfies Equation B.55 (c.f., Salganik et al. (2011a)).

Estimating the true positive rate, τ_F

We can combine our estimator for $\bar{v}_{H,F}$ (Result B.3.2) and our estimator for $\bar{d}_{H,F}$ (Result B.3.6) to estimate the true positive rate τ_F.

Result B.3.8 *The estimator*

$$\widehat{\tau}_F = \frac{\widehat{\bar{v}}_{H,F}}{\widehat{\bar{d}}_{H,F}} \tag{B.57}$$

is consistent and essentially unbiased for τ_F if $\widehat{\bar{v}}_{H,F}$ is a consistent and essentially unbiased estimator of $\bar{v}_{H,F}$ and if $\widehat{\bar{d}}_{H,F}$ is a consistent and essentially unbiased estimator of $\bar{d}_{H,F}$.

Proof: This follows directly from the properties of a compound ratio estimator (Appendix B.5). ■

More concretely, combing the estimator in Result B.3.2 and Result B.3.6 yields an estimator for $\widehat{\tau}_F$ with the following form:

$$\widehat{\tau}_F = \frac{\sum_{i\in s_H} \widetilde{v}_{i,\mathcal{A}_H}/(c\pi_i)}{\sum_{i\in s_H} y_{i,\mathcal{A}_H}/(c\pi_i)}. \tag{B.58}$$

All of the factors involving the size of $\mathcal{A}$ drop out of Equation B.58. This fact allows researchers to use groups defined by first names (e.g., people named Michael) in the probe alters $\mathcal{A}$, even if the size of these groups is not known (c.f., Salganik et al. (2011b)).

B.3.7 Estimating the size of the hidden population, N_H

We now make use of all of the results for the individual terms we derived above to present four different estimators for the size of the hidden population, N_H.

Result B.3.9 *The generalized scale-up estimator given by*

$$\widehat{N}_H = \frac{\widehat{y}_{F,H}}{\widehat{\bar{v}}_{H,F}} \tag{B.59}$$

is consistent and essentially unbiased for N_H if there are no false positive reports, if $\widehat{y}_{F,H}$ is consistent and unbiased for $y_{F,H}$, and if $\widehat{\bar{v}}_{H,F}$ is consistent and essentially unbiased for $\bar{v}_{H,F}$.

Proof: From the properties of a compound ratio estimator, we know that our estimator is consistent and essentially unbiased for $y_{F,H}/\bar{v}_{H,F}$ (Appendix B.5). By the argument in the main text given in Section 3.1, leading to Equation 2.10, this quantity is equal to N_H. ■

Result B.3.10 *The adjusted basic scale-up estimator given by*

$$\widehat{N}_H = \frac{\widehat{y}_{F,H}}{\widehat{\bar{d}}_{U,F}} \frac{1}{\widehat{\phi}_F} \frac{1}{\widehat{\delta}_F} \frac{1}{\widehat{\tau}_F} \tag{B.60}$$

is consistent and essentially unbiased for N_H if there are no false positive reports, and if each of the individual estimators is consistent and essentially unbiased.

Proof: From the results in Appendix B.5, we know that this compound ratio estimator will be consistent and essentially unbiased for $y_{F,H}/(\bar{d}_{U,F}\ \phi_F\ \delta_F\ \tau_F)$. The denominator is $\bar{v}_{H,F}$ by construction, leaving us with $y_{F,H}/\bar{v}_{H,F}$. By the argument in the main text given in Section 3.1, leading to Equation 2.10, this quantity is equal to N_H. ■

Result B.3.11 *The adjusted scale-up estimator*

$$\widehat{N}_H = \frac{\widehat{y}_{F,H}}{\widehat{\bar{d}}_{F,F}}\ \frac{1}{\widehat{\delta}_F}\ \frac{1}{\widehat{\tau}_F} \tag{B.61}$$

is consistent and essentially unbiased for N_H if there are no false positives, and if each of the individual estimators is consistent and essentially unbiased.

Proof: From the results in Appendix B.5, we know that this compound ratio estimator will be consistent and essentially unbiased for $y_{F,H}/(\bar{d}_{F,F}\ \delta_F\ \tau_F)$. The denominator is $\bar{v}_{H,F}$ by construction, leaving us with $y_{F,H}/\bar{v}_{H,F}$. By the argument in the main text given in Section 3.1, leading to Equation 2.10, this quantity is equal to N_H. ■

Result B.3.12 *The adjusted scale-up estimator*

$$\widehat{N}_H = \frac{\widehat{y}_{F,H}}{\widehat{\bar{d}}_{F,F}}\ \frac{1}{\widehat{\delta}_F}\ \frac{1}{\widehat{\tau}_F}\ \widehat{\eta}_F \tag{B.62}$$

is consistent and essentially unbiased for N_H if each of the individual estimators is consistent and essentially unbiased.

Proof: From the results in Appendix B.5, we know that this compound ratio estimator will be consistent and essentially unbiased for $(y_{F,H}\ \eta_F)/(\bar{d}_{F,F}\ \delta_F\ \tau_F)$. The numerator is $y^+_{F,H}$ by construction and the product of the denominators is $\bar{v}_{H,F}$ by construction, leaving us with $y^+_{F,H}/\bar{v}_{H,F}$. By the argument in Appendix B.1 this quantity is equal to N_H. ■

B.4 Sensitivity

All of the estimators that we propose require that specific conditions hold in order to produce consistent and essentially unbiased estimates. Result B.3.5 shows how we can quantify the extent to which one of these estimators is robust to violations of the conditions it depends upon.

Table B.2 shows results analogous to Result B.3.5 for all of the estimators we propose. We do not prove each one individually, since the derivations all follow the pattern of Result B.3.5 very closely. Researchers who wish to understand the how their estimates are affected by the assumptions they make can use Table B.2 to conduct a sensitivity analysis. Here, we use Table B.2 to show specific sensitivity results for the generalized scale-up estimator and the basic scale-up estimator. Note that in both cases, any problems with the sampling design—the probability sample from the frame for basic scale-up, and the probability sample from the frame or the relative probability sample from the hidden population for generalized scale-up—could result in problems with the estimates that do not get captured by these sensitivity analyses.

B.4.1 Generalized scale-up

The generalized scale-up estimator we recommend in Section 3.4 is $\widehat{N}_H = \widehat{y}_{F,H}/\widehat{\bar{v}}_{H,F}$. From the entry for $\widehat{\bar{v}}_{H,F}$ in Table B.2, and from Appendix B.1, we can see that if the conditions required by the generalized scale-up estimator do not hold, then its effective estimand is

$$\widehat{N}_H = \frac{\widehat{y}_{F,H}}{\widehat{\bar{v}}_{H,F}} \rightsquigarrow \frac{c_1\ \eta_F}{c_2\ c_3} N_H, \tag{B.63}$$

where η_F is the false positive rate, c_1 comes from error in the total for the groups of known size, c_2 comes from violating the accurate aggregate reports about visibility

condition, and c_3 comes from violating the probe alter condition (Table B.2 and Result B.3.2).

B.4.2 Basic scale-up

The basic scale-up estimator we recommend in Section 3.4 is $\widehat{N}_H = \widehat{y}_{F,H}/\widehat{\bar{d}}_{F,F}$. From the entry for $\widehat{\bar{d}}_{F,F}$, Appendix B.1, and Sections 3.1 and 3.2, we deduce that if the conditions for the basic scale-up estimator do not hold, then its effective estimand is

$$\widehat{N}_H = \frac{\widehat{y}_{F,H}}{\widehat{\bar{d}}_{F,F}} \rightsquigarrow \frac{c_1\ \eta_F}{\tau_F\ \delta_F\ c_2\ c_3} N_H, \tag{B.64}$$

where η_F is the false positive rate, τ_F is the true positive rate, δ_F is the degree ratio, c_1 comes from error in the total for the groups of known size, c_2 comes from violating the probe alter condition, and c_3 comes from violating the reporting condition (Table B.2 and Result B.2.3).

B.5 Approximate unbiasedness of compound ratio estimators

B.5.1 Overview

Several of the estimators we propose are nonlinear, which means that they are not design-unbiased (Sarndal et al., 2003). While ratio estimators are common in survey sampling and the bias of these estimators is commonly regarded as insignificant (Sarndal et al., 2003), several of the estimators we propose are somewhat more complex than standard ratio estimators. In fact, all of our nonlinear estimators turn out to all be special cases of a ratio of ratios (Table B.3), which is also known as a double ratio estimator (Rao and Pereira, 1968). Any double ratio can be written

Estimator	Imperfect assumptions	Effective estimand
$\widehat{\bar{d}}_{F,F}$ (Result B.2.3)	(i) $\widehat{N}_{\mathcal{A}} = c_1\ N_{\mathcal{A}}$ (ii) $\bar{d}_{\mathcal{A},F} = c_2\ \bar{d}_{F,F}$ (iii) $y_{F,\mathcal{A}} = c_3\ d_{F,\mathcal{A}}$	$\frac{c_2\ c_3}{c_1}\ \bar{d}_{F,F}$
$\widehat{\bar{d}}_{U,F}$ (Result B.2.4)	(i) $\widehat{N}_{\mathcal{A}} = c_1\ N_{\mathcal{A}}$ (ii) $\bar{d}_{\mathcal{A},F} = c_2\ \bar{d}_{U,F}$ (iii) $y_{F,\mathcal{A}} = c_3\ d_{F,\mathcal{A}}$	$\frac{c_2\ c_3}{c_1}\ \bar{d}_{U,F}$
$\widehat{\phi}_F$ (Result B.2.6)	(i) $\widehat{\bar{d}}_{F,F} \rightsquigarrow c_1\ \bar{d}_{F,F}$ (ii) $\widehat{\bar{d}}_{U,F} \rightsquigarrow c_2\ \bar{d}_{U,F}$	$\frac{c_1}{c_2}\ \phi_F$
$\widehat{\bar{v}}_{H,F}$ (Result B.3.2)	(i) $\widehat{N}_{\mathcal{A}\cap F} = c_1\ N_{\mathcal{A}\cap F}$ (ii) $\tilde{v}_{H,\mathcal{A}\cap F} = c_2\ v_{H,\mathcal{A}\cap F}$ (iii) $\frac{v_{H,\mathcal{A}\cap F}}{N_{\mathcal{A}\cap F}} = c_3\ \frac{v_{H,F}}{N_F}$	$\frac{c_3\ c_2}{c_1}\ \bar{v}_{H,F}$
$\widehat{\delta}_F$ (Result B.3.7)	(i) $\widehat{\bar{d}}_{H,F} \rightsquigarrow c_1\ \bar{d}_{H,F}$ (ii) $\widehat{\bar{d}}_{F,F} \rightsquigarrow c_2\ \bar{d}_{F,F}$	$\frac{c_1}{c_2}\ \delta_F$
$\widehat{\tau}_F$ (Result B.3.8)	(i) $\widehat{\bar{v}}_{H,F} \rightsquigarrow c_1\ \bar{v}_{H,F}$ (ii) $\widehat{\bar{d}}_{H,F} \rightsquigarrow c_2\ \bar{d}_{H,F}$	$\frac{c_1}{c_2}\ \tau_F$
$\widehat{N}_H$ (Result B.3.9)	(i) $\widehat{\bar{v}}_{H,F} \rightsquigarrow c_1\ \bar{v}_{H,F}$	$\frac{1}{c_1}\ N_H$
$\widehat{N}_H$ (Result B.3.11)	(i) $\widehat{\bar{d}}_{F,F} \rightsquigarrow c_1\ \bar{d}_{F,F}$ (ii) $\widehat{\delta}_F \rightsquigarrow c_2\ \delta_F$ (iii) $\widehat{\tau}_F \rightsquigarrow c_3\ \tau_F$	$\frac{1}{c_1\ c_2\ c_3}\ N_H$

Table B.2: Summary of estimators' robustness to imperfect assumptions. The first column lists several estimators we discuss in the main body and appendixes. The consistency and approximate unbiasedness of each estimator relies upon particular conditions being satisfied. These conditions are given in the second column, with a modification: we add a constant to each condition; if the constant is 1, then the original condition is satisfied. The estimand is then effectively changed to the quantity listed in the third column. (NB: we use the symbol $\rightsquigarrow$ as a shorthand for 'is consistent and essentially unbiased for'.) For example, the first row shows $\widehat{\bar{d}}_{F,F}$ and the three conditions that the estimator in Result B.2.3 relies upon. Suppose that the first and third hold, so that $c_1 = 1$ and $c_3 = 1$, but that the second does

$$R_d = \frac{R_1}{R_0} = \frac{\frac{\bar{y}_1}{\bar{x}_1}}{\frac{\bar{y}_0}{\bar{x}_0}} = \frac{\bar{y}_1 \bar{x}_0}{\bar{x}_1 \bar{y}_0}. \tag{B.65}$$

If we have unbiased estimators for each of the four terms, we can estimate R_d by

$$\widehat{r}_d = \frac{\widehat{\bar{y}}_1 \widehat{\bar{x}}_0}{\widehat{\bar{x}}_1 \widehat{\bar{y}}_0}. \tag{B.66}$$

In this appendix we investigate when we can expect the biases in our estimators to be small enough to be negligible.

B.5.2 The general case

We will focus on the relative bias in our estimator, $\widehat{r}_d$. The relative bias is given by

$$B_d = \frac{\mathbb{E}[\widehat{r}_d] - R_d}{R_d}. \tag{B.67}$$

B_d expresses the bias in our estimator $\widehat{r}_d$ in terms of the true value; a relative bias of 0.5, for example, means that our estimator is typically 0.5 times bigger than the true value. This is a natural quantity to consider because estimators that have small relative bias have small bias in substantive terms.

Our approach will be to follow Rao and Pereira (1968) in using a Taylor series to form an approximation to the relative bias. This is accomplished in Result B.5.1.

Result B.5.1 *(Rao and Pereira, 1968) If $\widehat{\bar{x}}_0$, $\widehat{\bar{x}}_1$, $\widehat{\bar{y}}_0$, and $\widehat{\bar{y}}_1$ are unbiased estimators, and $|(\widehat{\bar{x}}_1 - \bar{x}_1)/\bar{x}_1| < 1$ and $|(\widehat{\bar{y}}_0 - \bar{y}_0)/\bar{y}_0| < 1$, then the relative bias of the double ratio estimator, B_d, is approximated by*

$$B_d = \frac{\mathbb{E}[\widehat{r}_d] - R}{R} \approx B'_d = C_{\widehat{\bar{x}}_1, \widehat{\bar{y}}_0} - C_{\widehat{\bar{x}}_1, \widehat{\bar{y}}_1} - C_{\widehat{\bar{y}}_0, \widehat{\bar{y}}_1} - C_{\widehat{\bar{x}}_0, \widehat{\bar{x}}_1} - C_{\widehat{\bar{x}}_0, \widehat{\bar{y}}_0} + C_{\widehat{\bar{y}}_1, \widehat{\bar{x}}_0} + C^2_{\bar{y}_0} + C^2_{\bar{x}_1}, \tag{B.68}$$

Estimator	Reference	Form	$\widehat{\widehat{x}}_0$	$\widehat{\widehat{y}}_1$	$\widehat{\widehat{x}}_1$	$\widehat{\widehat{y}}_0$	Approx. rel. bias
$\widehat{\phi}_F$	Res. B.2.6	$K\widehat{\widehat{x}}_0/\widehat{\widehat{y}}_0$	$\sum_{i\in s_F} y_{i,\mathcal{A}_{F_1}}/\pi_i$	-	-	$\sum_{i\in s_F} y_{i,\mathcal{A}_{F_2}}/\pi_i$	$C^2_{\widehat{\widehat{y}}_0} - C_{\widehat{\widehat{y}}_0,\widehat{\widehat{x}}_0}$
$\widehat{\widehat{v}}_{H,F}$	Res. B.3.2	$K\widehat{\widehat{x}}_0/\widehat{\widehat{y}}_0$	$\sum_{i\in s_H} \tilde{v}_{i,\mathcal{A}_H\cap F}/c\pi_i$	-	-	$\sum_{i\in s_H} 1/c\pi_i$	$C^2_{\widehat{\widehat{y}}_0} - C_{\widehat{\widehat{y}}_0,\widehat{\widehat{x}}_0}$
$\widehat{\widehat{d}}_{H,F}$	Res. B.3.6	$K\widehat{\widehat{x}}_0/\widehat{\widehat{y}}_0$	$\sum_{i\in s_H} y_{i,\mathcal{A}_H\cap F}/c\pi_i$	-	-	$\sum_{i\in s_H} 1/c\pi_i$	$C^2_{\widehat{\widehat{y}}_0} - C_{\widehat{\widehat{y}}_0,\widehat{\widehat{x}}_0}$
$\widehat{\delta}_F$	Res. B.3.7	$K\widehat{\widehat{x}}_0/(\widehat{\widehat{y}}_0\ \widehat{\widehat{x}}_1)$	$\sum_{i\in s_H} y_{i,\mathcal{A}_H\cap F}/c\pi_i$	-	$\sum_{i\in s_F} y_{i,\mathcal{A}_F}/\pi_i$	$\sum_{i\in s_H} 1/c\pi_i$	$C^2_{\widehat{\widehat{y}}_0} + C^2_{\widehat{\widehat{x}}_1} - C_{\widehat{\widehat{y}}_0,\widehat{\widehat{x}}_0}$
$\widehat{\tau}_F$	Res. B.3.8	$K\widehat{\widehat{x}}_0/(\widehat{\widehat{y}}_0\ \widehat{\widehat{x}}_1)$	$\sum_{i\in s_H} \tilde{v}_{i,\mathcal{A}_H\cap F}/c\pi_i$	-	$\sum_{i\in s_H} y_{i,\mathcal{A}_H\cap F}/c\pi_i$	$\sum_{i\in s_H} 1/c\pi_i$	$C^2_{\widehat{\widehat{y}}_0} + C^2_{\widehat{\widehat{x}}_1} - C_{\widehat{\widehat{y}}_0,\widehat{\widehat{x}}_0}$
$\widehat{N}_H$	Res. B.3.9	$K\widehat{\widehat{y}}_1\widehat{\widehat{x}}_0/\widehat{\widehat{y}}_0$	$\sum_{i\in s_H} 1/c\pi_i$	$\sum_{i\in s_F} y_{i,H}/\pi_i$	-	$\sum_{i\in s_H} \tilde{v}_{i,\mathcal{A}_H\cap F}/c\pi_i$	$C^2_{\widehat{\widehat{y}}_0} - C_{\widehat{\widehat{y}}_0,\widehat{\widehat{x}}_0}$
$\widehat{N}_H$	Res. B.3.11	$K\widehat{\widehat{x}}_0/\widehat{\widehat{y}}_0$	$\sum_{i\in s_F} y_{i,H}/\pi_i$	-	-	$\sum_{i\in s_F}\sum_j y_{i,A_j}/\pi_i$	$C^2_{\widehat{\widehat{y}}_0} - C_{\widehat{\widehat{y}}_0,\widehat{\widehat{x}}_0}$

Table B.3: Description of the general form of the nonlinear estimators we propose. K is a constant, $\widehat{\widehat{y}}_1$ and $\widehat{\widehat{x}}_1$ are taken from s_F, while $\widehat{\widehat{x}}_0$ and $\widehat{\widehat{y}}_0$ are taken from s_H. Our nonlinear estimators are all special cases of the double ratio estimator, which we define and discuss below. Note that the estimator for $\widehat{N}_H$ that involves adjusting a basic scale-up estimate (Result B.3.11) would, in practice, take these adjustment factors from other studies; we therefore assume that these adjustment factors are independent of the quantities that go into the scale-up estimate, and treat them as constants.

where $C_{\widehat{\bar{x}},\widehat{\bar{y}}} = \frac{cov(\widehat{\bar{x}},\widehat{\bar{y}})}{\bar{x}\bar{y}}$ *is the relative covariance between* $\widehat{\bar{x}}$ *and* $\widehat{\bar{y}}$*, and* $C^2_{\widehat{\bar{y}}} = \frac{var(\widehat{\bar{y}})}{\widehat{\bar{y}}^2}$.

Proof: Define

$$\delta_{\widehat{\bar{x}}_0} = \frac{\widehat{\bar{x}}_0 - \bar{x}_0}{\bar{x}_0}, \tag{B.69}$$

with analogous definitions for $\delta_{\widehat{\bar{x}}_1}$, $\delta_{\widehat{\bar{y}}_1}$, and $\delta_{\widehat{\bar{y}}_0}$. We can express r_d as

$$\widehat{r}_d = R\frac{(1+\delta_{\widehat{\bar{y}}_1})(1+\delta_{\widehat{\bar{x}}_0})}{(1+\delta_{\widehat{\bar{y}}_0})(1+\delta_{\widehat{\bar{x}}_1})}. \tag{B.70}$$

The relative bias then becomes

$$B_d = \frac{\mathbb{E}[\widehat{r}_d] - R}{R} = \mathbb{E}\left[\frac{(1+\delta_{\widehat{\bar{y}}_1})(1+\delta_{\widehat{\bar{x}}_0})}{(1+\delta_{\widehat{\bar{y}}_0})(1+\delta_{\widehat{\bar{x}}_1})}\right] - 1. \tag{B.71}$$

The strategy is now to expand the two factors in the denominator and to then discard high-order terms. What remains will be an approximation to the true relative bias.

Recall that if $|x| < 1$ then $\frac{1}{1-x} = \sum_{i=0}^{\infty} x^i$ and, in particular, $\frac{1}{1+x} = 1 - x^2 + x^3 - \cdots$. We'll make use of this expansion for the two factors in the denominator of Equation B.71; that is, we assume that $|\delta_{\bar{y}_0}| < 1$ and $|\delta_{\bar{x}_1}| < 1$. Then we have

$$B_d = \mathbb{E}\left[(1+\delta_{\widehat{\bar{y}}_1})(1+\delta_{\widehat{\bar{x}}_0})(1-\delta_{\widehat{\bar{y}}_0}+\delta^2_{\widehat{\bar{y}}_0} - \cdots)(1-\delta_{\widehat{\bar{x}}_1}+\delta^2_{\widehat{\bar{x}}_1} - \cdots)\right] - 1 \tag{B.72}$$

If we multiply this out and retain only terms up to order 2, we obtain the following approximation:

$$B_d \approx \mathbb{E}\left[\delta_{\widehat{\bar{x}}_1}\delta_{\widehat{\bar{y}}_0} + \delta_{\widehat{\bar{x}}_0}\delta_{\widehat{\bar{y}}_1} - \delta_{\widehat{\bar{x}}_0}\delta_{\widehat{\bar{y}}_0} - \delta_{\widehat{\bar{x}}_0}\delta_{\widehat{\bar{x}}_1} - \delta_{\widehat{\bar{x}}_1}\delta_{\widehat{\bar{y}}_1} - \delta_{\widehat{\bar{y}}_0}\delta_{\widehat{\bar{y}}_1} + \delta_{\widehat{\bar{x}}_0} + \delta_{\widehat{\bar{y}}_1} - \delta_{\widehat{\bar{x}}_1} - \delta_{\widehat{\bar{y}}_0} - \delta^2_{\widehat{\bar{y}}_0} - \delta^2_{\widehat{\bar{x}}_1}\right]. \tag{B.73}$$

Since we assumed that the estimators for the individual components of r_d are unbiased, we know that

$$\mathbb{E}[\delta_{\widehat{\bar{x}}_1}] = 0, \tag{B.74}$$

We can also determine that

$$\mathbb{E}[\delta_{\widehat{\bar{x}}_1}\delta_{\widehat{\bar{y}}_1}] = \frac{\mathrm{cov}(\widehat{\bar{x}}_1, \widehat{\bar{y}}_1)}{\bar{x}_1\bar{y}_1}, \tag{B.75}$$

and, that

$$\mathbb{E}[\delta^2_{\widehat{\bar{x}}_1}] = \frac{\mathrm{var}(\widehat{\bar{x}}_1)}{\bar{x}_1^2}. \tag{B.76}$$

Applying these relationships to Equation B.73, we find

$$B_d \approx C_{\widehat{\bar{x}}_0,\widehat{\bar{y}}_1} + C_{\widehat{\bar{x}}_1,\widehat{\bar{y}}_0} - C_{\widehat{\bar{x}}_0,\widehat{\bar{x}}_1} - C_{\widehat{\bar{x}}_0,\widehat{\bar{y}}_0} - C_{\widehat{\bar{x}}_1,\widehat{\bar{y}}_1} - C_{\widehat{\bar{y}}_0,\widehat{\bar{y}}_1} + C^2_{\bar{x}_1} + C^2_{\bar{y}_0}, \tag{B.77}$$

which is our result. ■

Result B.5.1 is useful because it reveals the behavior of double ratio estimators in quite general contexts. To understand what it says a bit more intuitively, note that Result B.5.1 is framed in terms of the relative covariances and variances of the *estimators* $\widehat{\bar{x}}_0$, $\widehat{\bar{x}}_1$, $\widehat{\bar{y}}_0$, and $\widehat{\bar{y}}_1$. In the special case of simple random sampling with replacement, we can re-write the approximation in terms of the finite population variances and covariances and a constant, κ:

$$B'_d = \kappa\left[C_{x_1,y_0} - C_{x_1,y_1} - C_{y_0,y_1} - C_{x_0,x_1} - C_{x_0,y_0} + C_{y_1,x_0} + C^2_{y_0} + C^2_{x_1}\right], \tag{B.78}$$

where $\kappa = \left(\frac{1}{n} - \frac{1}{N}\right)$, n is our sample size, and N is the size of the population. In the case of simple random sampling, the relative bias depends upon the finite population variances of the underlying population values and the size of our sample.

For designs other than simple random sampling, there is no analogous expression as simple as Equation B.78. However, speaking roughly, if we have an idea that our sampling plan has a typical design effect (deff) for the quantities inside the square brackets in Equation B.78, then we can see that we would simply replace the κ in Equation B.78 by $(\kappa \cdot \text{deff})$ in order to get a sense of the approximate relative bias.

Notice, also, that Result B.5.1 is framed largely in terms of relative covariances. When we apply Result B.5.1, we will often make use of the fact that the relative covariances can be expressed in terms of correlations and coefficients of variation as follows:

$$C_{\widehat{\bar{x}},\widehat{\bar{y}}} = \frac{\text{cov}(\widehat{\bar{x}},\widehat{\bar{y}})}{\bar{x}\bar{y}} = \frac{\rho_{\widehat{\bar{x}},\widehat{\bar{y}}}\sqrt{\text{var}(\widehat{\bar{x}})}\sqrt{\text{var}(\widehat{\bar{y}})}}{\bar{x}\bar{y}} \tag{B.79}$$

$$= \rho_{\widehat{\bar{x}},\widehat{\bar{y}}}\ \text{cv}(\widehat{\bar{x}})\ \text{cv}(\widehat{\bar{y}}), \tag{B.80}$$

where $\rho_{\widehat{\bar{x}},\widehat{\bar{y}}}$ is the correlation between the estimators $\widehat{\bar{x}}$ and $\widehat{\bar{y}}$, and $\text{cv}(\widehat{\bar{x}}) = \frac{\sqrt{\text{var}(\widehat{\bar{x}})}}{\bar{x}}$ is the coefficient of variation of the estimator $\widehat{\bar{x}}$. We will also make use of the fact that $C^2_{\widehat{\bar{x}}} = \text{cv}(\widehat{\bar{x}})^2$.

B.5.3 Applying Result B.5.1 to scale-up

We now apply Result B.5.1 to understand the biases in the nonlinear estimators we propose for realistic situations. For each particular estimator, we can simplify the expression in Result B.5.1. In order to do so, we first remove terms that do not appear in the estimator itself (for example, in $\widehat{\delta}_F$, there is no $\widehat{\bar{y}}_1$). Additionally, we assume that the estimates produced from a sample from the frame population and a

sample from the hidden population will be independent of one another, meaning that their correlation will be 0. Table B.3 summarizes the nonlinear estimators we propose, along with the specific version of the approximate relative bias from Result B.5.1 that applies.

Finally, in order to give a sense of the magnitude of the coefficients of variation and correlations found in real studies, we estimated the quantities that go into the approximate relative bias from the studies available to us. Table B.4 shows the coefficients of variation for the estimated degree (the values of $\widehat{\bar{x}}_1$ for $\widehat{\delta}_F$) in surveys from Rwanda, the United States, and Curitiba, Brazil. Further, Tables B.5 and B.6 show the relevant coefficients of variation and pairwise correlations for all remaining quantities using data from Curitiba, Brazil (currently, the only setting where we have data from a sample of the hidden population). For all values in these tables, the estimated variance of the estimators is calculated using the bootstrap methods presented in Section B.6.1.

Since we have both a sample from the frame population and a sample from the hidden population in Curitiba, we can compute numerical estimates of the bias of each nonlinear estimator in the context of that study. We can see that in this study bias caused by the nonlinearity of the estimator was not a big problem: in each case, the estimated approximate bias was less than one percent of the estimate (Table B.7).

To conclude, we began by deriving an expression for the approximate relative bias in double ratio estimators in general. We then simplified the approximation for each specific nonlinear estimator that we propose. Finally, we used data from a real scale-up study in Curitiba, Brazil to estimate magnitude of the biases caused by the non-linearity of the estimators in a specific scale-up study. From these results, we conclude that theses estimators are essentially unbiased, and that sampling error and non-sampling error will dominate any bias introduced by the nonlinear form of the estimators.

$\widehat{\text{cv}}(\widehat{\bar{d}})$	source
0.04	Rwanda
0.09	Curitiba
0.02	US

Table B.4: Estimated coefficients of variation for the average degree from 3 different scale-up surveys. These play a role in the approximate relative bias for the estimate of $\widehat{\delta}_F$. Our approximation tells us that the larger these values are, the worse the relative bias will be. The estimates were computed using the rescaled bootstrap procedure.

	estimated coef. of variation
$\sum_{i \in s_H} y_{i,\mathcal{A} \cap F} / c\pi_i$	0.08
$\sum_{i \in s_H} \tilde{v}_{i,\mathcal{A} \cap F} / c\pi_i$	0.08
$\sum_{i \in s_H} 1 / c\pi_i$	0.06

Table B.5: Estimated coefficients of variation for quantities derived from a sample from the hidden population. These quantities play a role in the approximate relative bias for the estimate of all of the nonlinear estimators we propose. The estimates were computed using the respondent-driven sampling bootstrap procedure (Salganik, 2006).

B.6 Variance estimation

B.6.1 Variance estimation with a sample from F

The goal of a bootstrap variance estimation procedure is to put a confidence interval around an estimate $\widehat{N}_H$ that is derived from a sample s_F. The most standard bootstrap procedure has three steps. First, researchers generate B replicate samples, $s_F^{(1)}, s_F^{(2)}, \ldots, s_F^{(B)}$ by randomly sampling with replacement from s_F. Second, these replicate samples are then used to produce a set of replicate estimates, $\widehat{N}_H^{(1)}, \widehat{N}_H^{(2)}, \ldots, \widehat{N}_H^{(B)}$. Finally, the replicate estimates are combined to produce a confidence interval; for example, by the percentile method which chooses the 2.5th and 97.5th percentiles of the B estimates (Fig. B.3) (Efron and Tibshirani, 1993).

When the original sample can be modeled as a simple random sample, this standard bootstrap procedure is appropriate. For example, consider the scale-up study of McCarty et al. (2001) that was based on telephone survey of 1,261 Americans

	estimated correlation
$\widehat{\text{cor}}(\sum_{i \in s_H} y_{i,\mathcal{A} \cap F}/c\pi_i, \sum_{i \in s_H} \tilde{v}_{i,\mathcal{A} \cap F}/c\pi_i)$	0.92
$\widehat{\text{cor}}(\sum_{i \in s_H} y_{i,\mathcal{A} \cap F}/c\pi_i, \sum_{i \in s_H} 1/c\pi_i)$	0.71
$\widehat{\text{cor}}(\sum_{i \in s_H} \tilde{v}_{i,\mathcal{A} \cap F}/c\pi_i, \sum_{i \in s_H} 1/c\pi_i)$	0.68

Table B.6: Estimated pairwise correlations for quantities derived from a sample from the hidden population. These quantities play a role in the approximate relative bias for the estimate of all of the nonlinear estimators we propose.

	approx. rel. bias, B_d	estimate	estimated absolute bias
$\widehat{\tau}_F$	0.0005	0.77	0.0004
$\widehat{\delta}_F$	0.0086	0.69	0.0059
$\widehat{N}_H$	0.0027	114498	304

Table B.7: Approximate relative bias in the estimates of the nonlinear quantities using data taken from the Curitiba study, the point estimates produced by the Curitiba study, and the estimated implied absolute bias. For each quantity, the bias is very small.

selected via random digit dialing.[4] We can approximate the sampling design as simple random sampling, and draw $B = 10,000$ replicate samples of size 1,261. In this case the bootstrap confidence intervals are, as expected, larger than the confidence

[4]The original data file includes 1,375 respondents. From these cases, 113 respondents who had missing data for some of the aggregated relational data questions and 1 respondent who answered 7 for all questions (see Zheng et al. (2006)). Further, consistent with common practice (e.g., Zheng et al. (2006)), we top coded all responses at 30, affecting 0.26% of responses.

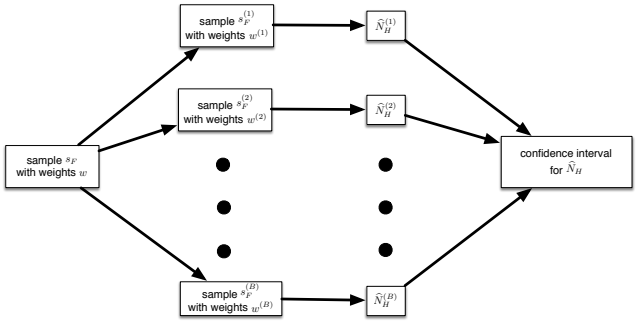


Figure B.3: Schematic of the bootstrap procedure to put a confidence interval around $\widehat{N}_H$ when there is a sample from the frame s_F.

intervals from Equation 3.16, since they account for the clustering of responses with respondent; on average, they are 2.05 times wider.

This standard bootstrap procedure, however, can perform poorly when the original data are collected with a complex sample design (Shao, 2003). To deal with this problem Rust and Rao (1996), propose a modified bootstrap procedure that works well when the data are collected with a general multistage sampling design, a class of designs that includes most designs that would be used for face-to-face scale-up surveys. For example, it includes stratified two-stage cluster sampling with oversampling (as was used in a recent scale-up study in Rwanda (Rwanda Biomedical Center, 2012)) and three-stage element sampling (as was used in a recent scale-up study in Curitiba, Brazil (Salganik et al., 2011a)); a full description of the designs included in this class is presented in Rust and Rao (1996).

This bootstrap procedure proposed by Rust and Rao (1996) includes two conceptual changes from the standard bootstrap. First, we approximate the actual sampling design by a closely related one that is much easier to work with. In particular, if we assume that primary sampling units (PSUs) are selected with replacement and that all subsequent stages of sampling are conducted independently each time a given PSU is selected, then we can use the with-replacement sampling framework in which variance estimation is much easier; see Sarndal et al. (2003) Result 4.5.1 for a more formal version of this claim. It is important to note that this approximation is generally conservative because with-replacement sampling usually results in higher variance than without-replacement sampling. Therefore, we will be estimating the variance for a design that has higher variance than the actual design. In practice, this difference is usually small because the sampling fraction in each stratum is usually small (Rao et al., 1992; Rust and Rao, 1996); see Sarndal et al. (2003) Section 4.6 for a more formal treatment. To estimate the variance in this idealized with-replacement design,

resampling should be done independently in each stratum and the units that are resampled with replacement should be entire PSUs, not respondents.

This change—resampling PSUs, not respondents—introduces the need for a second change in the resampling procedure. It is known that the standard bootstrap procedure is off by a factor of $(n-1)/n$ where n is the sample size (Rao and Wu, 1988b). Thus, when the sample size is very small, the bootstrap will tend to underestimate the variance. While this issue is typically ignored, it can become important when we resample PSUs rather than respondents. In particular, the number of sampled PSUs in stratum h, n_h, can be small in complex sample designs. At the extreme, in a design with two sampled PSUs per stratum, which is not uncommon, the uncorrected bootstrap would be expected to produce a 50% underestimate of the variance. Therefore, Rao et al. (1992) developed the rescaled bootstrap, whereby the bootstrap sample size is slightly smaller than the original sample size and the sample weights are rescaled to account for this difference. Rust and Rao (1996) recommend that if the original sample includes n_h PSUs in strata h, then researchers should resample $n_h - 1$ PSUs and rescale the respondent weights by $n_h/(n_h - 1)$. That is, the weight for the jth person in PSU i in the bth replicate sample is

$$w_{ij}^{(b)} = w_{ij} \times \frac{n_h}{(n_h - 1)} \times r_i^{(b)} \tag{B.81}$$

where w_{ij} is the original weight for the jth unit in the ith PSU, n_h is the number of PSUs in strata h, and $r_i^{(b)}$ is the number of times the ith PSU was selected in replicate sample b.

In Figure 5.1 in the main text, we compared the three different procedures for putting confidence intervals around the basic scale-up estimator: the current scale-up variance estimator (Killworth et al., 1998b), the standard bootstrap, and the rescaled bootstrap. We made this comparison using data from recent scale-up studies

in the United States, Rwanda,[5] and Curitiba, Brazil.[6] As expected, the rescaled bootstrap produced confidence intervals that are larger than those from the standard bootstrap, which in turn are larger than those from the current scale-up variance estimation procedure. In the study from Curitiba, the rescaled bootstrap procedure produced confidence intervals 1.17 times larger than the standard bootstrap and 2.84 times larger than the current variance estimator. In the Rwanda case, the rescaled bootstrap procedure produced confidence intervals 1.58 times larger than the standard bootstrap and 2.95 times larger than the current variance estimator.

The standard bootstrap procedure for simple random samples and the rescaled bootstrap procedure for complex sample designs are both well-grounded statistically, but they are not perfect. Most importantly, they only capture sampling error and do not capture other forms of survey error such as response error, sampling frame error, and interviewer effects. Therefore, these procedures should be expected to produce 95% intervals that do not have the desired coverage probabilities. In fact, this is what we find in practice: Figure 3.1 shows the estimated confidence intervals for the groups of known size in the three studies described above. The coverage rates for the boot-

[5]The scale-up study in Rwanda used stratified two-stage cluster sampling with unequal probability of selection across strata in order to oversample urban areas. Briefly, the sample design divided Rwanda into five strata: Kigali City, North, East, South, and West. At the first stage, PSUs—in this case villages—were selected with probability proportional to size and without replacement within each stratum with oversampling in the Kigali City stratum. This approach resulted in a sample of 130 PSUs: 35 from Kigali City, 24 from East, 19 from North, 26 from South, and 26 from West. At the second stage, 20 households were selected via simple random sampling without replacement from each PSU in Kigali City and 15 households from each PSU in other strata. Finally, all members of the sampled household over the age of 15 were interviewed. For full details see Rwanda Biomedical Center (2012). The original data file includes 4,669 respondents. From these cases, we removed 6 respondents who had missing data for some of the aggregated relational data questions. Further, consistent with common practice (e.g., Zheng et al. (2006)), we top coded all responses at 30, affecting 0.23% of responses.

[6]The scale-up study in Curitiba, Brazil used two-stage element sampling where 54 primary sampling units (PSUs)—in this case census tracks—were selected with probability proportional to their estimated number of housing units and without replacement. Then, within each cluster, eight secondary sampling units (SSUs)—in this case people—were selected with equal probability without replacement. For full details see Salganik et al. (2011a). The original data file includes 500 respondents. From these cases, we removed no respondents who had missing data for some of the aggregated relational data questions. Further, consistent with common practice (e.g., Zheng et al. (2006)), we top coded all responses at 30, affecting 0.58% of responses.

strap confidence intervals for the US, Rwanda, and Curitiba, are 3.4%, 13.6%, 15.0%. While this is far from ideal, we note that it is slightly better than the currently used procedure (Equation 3.16), which produced coverage rates of 0.0%, 9.1%, 5.0%, and it is also slightly better than the standard bootstrap, which produced coverage rates of 3.4%, 9.1%, and 10.0%. We believe that these theoretical and empirical arguments demonstrate three conclusions. First, bootstrap confidence intervals, which do not assume the basic scale-up model, are preferable to the current variance estimation procedure; second, until there is a substantially better understanding of non-sampling errors, researchers should expect that the bootstrap confidence intervals will understate true uncertainty; and, third, variance estimation for the scale-up method is an important topic for future research.

B.6.2 Variance estimation with sample from F and H

In this paper we showed that the scale-up estimates could be improved if information from people on the frame was combined with information from people in the hidden population. Producing confidence intervals around the generalized scale-up estimator is more difficult because the generalized estimator has uncertainty from two different samples. To capture all of this uncertainty, we propose combining replicate samples from the frame population with independent replicate samples from the hidden population in order to produce a set of replicate estimates. More formally, given s_F, a sample from the frame population, and an independent sample s_H from the hidden population, we seek to produce a set of B bootstrap replicate samples for s_F and s_H, $s_F^{(1)}, s_F^{(2)}, \ldots, s_F^{(B)}$ and $s_H^{(1)}, s_H^{(2)}, \ldots, s_H^{(B)}$, which are then combined to produce a set of B bootstrap estimates: $\widehat{N}_H^{(1)} = f(s_F^{(1)}, s_H^{(1)})$, $\widehat{N}_H^{(2)} = f(s_F^{(2)}, s_H^{(2)}), \ldots \widehat{N}_H^{(B)} = f(s_F^{(B)}, s_H^{(B)})$. Finally, these B replicate estimates are converted into a confidence interval using the percentile method (Fig. B.4).

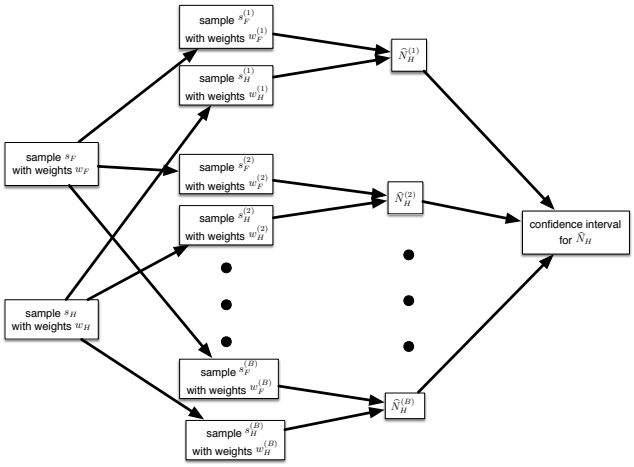


Figure B.4: Schematic of the bootstrap procedure to put a confidence interval around $\widehat{N}_H$ when there is a sample from the frame s_F and a sample from the hidden population s_H.

Because of the challenges involved in sampling hard-to-reach populations, the two most likely sampling designs for s_H will be time-location sampling and respondent-driven sampling. If s_H is selected with time-location sampling, we recommend treating the design as a two-stage element sample (see Karon and Wejnert (2012)) and using the procedure of Rust and Rao (1996). If s_H was selected with respondent-driven sampling, as was done in a recent study of heavy drug users in Curitiba, Brazil (Salganik et al., 2011b), we recommend using the best available bootstrap method for respondent-driven sampling data, which at the present time is the procedure introduced in Salganik (2006). One implementation detail of this bootstrap procedure is that it requires researchers to divide the sample of the hidden population into two mutually exclusive groups. In this case, we recommend dividing the hidden population into those who are above and below the median of their estimated visibility $\widehat{v}_{i,F}$ in order to capture some of the extra uncertainty introduced if there are strong tendencies for more hidden members of the hidden population to recruit each other.

Because the generalized scale-up method has never been used for groups of known size, we cannot explore the coverage rate of the proposed procedure. However, based on experience with respondent-driven sampling, we suspect that variance estimation

procedures for hidden populations will underestimate the actual uncertainty in the estimates (Goel and Salganik, 2009, 2010; Yamanis et al., 2013; Verdery et al., 2013). If this is the case, then the intervals around the generalized scale-up estimates will be too small.

In conclusion, Sec. B.6.1 presents a bootstrap procedure for simple and complex sample designs from the sampling frame, and Sec. B.6.2 extends these results to account for the sampling variability introduced by having a sample from the hidden population. We have shown that the performance of these procedures on three real scale-up datasets is consistent with theoretical expectations. Additional research in this area, which is beyond the scope of this paper, could adopt a total survey error approach and attempt to quantify all sources of uncertainty in the estimates, not just sampling uncertainty. Additional research could also explore the properties and sensitivity of these variance estimation procedures though simulation.

B.7 Simulation study

B.7.1 Overview

In this appendix, we describe a simulation study comparing the performance of the generalized and basic network scale-up estimators. The results of these simulations confirm and illustrate several of the analytical results in Section 3.2 of the paper. Most importantly, the simulations show that the generalized network scale-up estimator is unbiased for all of the situations explored by the simulation, while the basic network scale-up estimator is biased for all but a few special cases. Moreover, our analytical results correctly predict the bias of the basic network scale-up estimator in each case.

B.7.2 Simulation design

Our simulation study is intentionally simple in order to clearly illustrate our analytical results; it is not designed to be a realistic model of any scale-up study. Concretely, our simulations compare the performance of generalized and basic scale-up estimators as three important quantities vary: (1) the size of the frame population F, relative to the size of the entire population, U; (2) the extent to which people's network connections are not formed completely at random, also called the amount of inhomogenous mixing; and (3) the accuracy of reporting, as captured by the true positive rate τ_F (see Equation 3.13).

We simulate populations consisting of $N = 10,000$ people, using a stochastic block-model (White et al., 1976; Wasserman and Faust, 1994) to randomly generate networks with different amounts of inhomogenous mixing. Stochastic block models assume population members can be grouped into different *blocks*. For any pair of people, i and j, the probability that there is an edge between i and j is completely determined by the block memberships of i and j.

In our simulation model, each person can be either in or out of the frame population F and each person can also be either in or out of the hidden population H, producing four possible blocks: FH, $F\neg H$, $\neg F\neg H$, and $\neg FH$. (Here, we use the logical negation symbol, $\neg$, to denote not being in a group.) The probability of an edge between any two people i and j is then governed by a Bernoulli distribution whose mean is a function of the two block memberships:

$$\Pr(i \leftrightarrow j) \sim \text{Bernoulli}(\mu_{g(i),g(j)}), \tag{B.82}$$

where $g(i)$ is the block containing i, $g(j)$ is the block containing j, $i \leftrightarrow j$ denotes an undirected edge between i and j, and $\mu_{g(i),g(j)}$ is the probability of an edge between a member of group $g(i)$ and a member of group $g(j)$. In a network with a

no inhomogenous mixing (equivalent to an Erdos-Renyi random graph), $\mu_{g(i),g(j)}$ will be the same for all i and j. On the other hand, in a network with a high level of inhomogenous mixing, $\mu_{g(i),g(j)}$ will be relatively small when $g(i) \neq g(j)$ and $\mu_{g(i),g(j)}$ will be relatively large when $g(i) = g(j)$. We use the `igraph` library to conduct our simulations (Csárdi and Nepusz, 2006).

Each random network drawn under our simulation model depends on six parameters. The first three parameters describe population size and group memberships; they are:

- N, the size of the population
- p_F, the fraction of people in the frame population
- p_H, the fraction of people in the hidden population

For simplicity, we assume that each person's membership in the frame population (F) and the hidden population (H), is independent. This means that, for example, the number of people who are in both F and H will be $N \times p_F \times p_H$.

The next two parameters govern the amount of inhomogenous mixing in the network that connects people to each other; they are:

- ζ, the probability of an edge between two people who are both in the same block.
- ρ, the relative probability of an edge between two vertices that are not in the same group. For example, a value of 0.8 would mean that the chances of having a connection between a particular person in F and a particular person not in F is 80% of the chance of a connection between two members of F or two members of $\neg F$.

Together, the parameters ζ and ρ are used to construct the mixing matrix M (Figure B.5), which is used as the basis for the probability of randomly creating an edges

$$\mathbf{M} = \begin{array}{c} \\ F\ H \\ F\neg H \\ \neg F\ H \\ \neg F\neg H \end{array} \begin{array}{c} \begin{array}{cccc} F\ H & F\neg H & \neg F\ H & \neg F\neg H \end{array} \\ \begin{pmatrix} \zeta & \rho\cdot\zeta & \rho\cdot\zeta & \rho^2\cdot\zeta \\ \rho\cdot\zeta & \zeta & \rho^2\cdot\zeta & \rho\cdot\zeta \\ \rho\cdot\zeta & \rho^2\cdot\zeta & \zeta & \rho\cdot\zeta \\ \rho^2\cdot\zeta & \rho\cdot\zeta & \rho\cdot\zeta & \zeta \end{pmatrix} \end{array} \tag{B.83}$$

Figure B.5: The mixing matrix used to generate a random network using the stochastic block model. Entry (i, j) in the matrix describes the probability of an edge between two people, one of whom is in group i and one in group j. The probabilities are governed by ζ and ρ. In our simulations, we generate networks with different amounts of inhomogenous mixing between the four groups by fixing ζ and varying ρ from 0.1 (extreme inhomogenous mixing) to 1 (perfectly random mixing).

between each pair of people in the population. Note that varying the parameter ρ will change several structural features of the network in addition to the amount of inhomogenous mixing; for example, changing ρ will alter the degree distribution. Our analytical results show that the generalized network scale-up estimator is robust to changes in these structural features.

The final parameter, τ_F, is used to control the amount of imperfect reporting. After randomly drawing a network using the stochastic block model, we generate a reporting network as follows:

1. convert all undirected edges $i \leftrightarrow j$ in the social network into two directed reporting edges in the reporting network: one $i \rightarrow j$ and one $j \rightarrow i$

2. select a fraction, $1 - \tau_F$, of the edges that lead from members of the frame population to members of the hidden population uniformly at random and remove them from the reporting graph.

Given a census of the relevant populations, the reporting graph is then used to compute the basic and generalized scale-up estimates for the size of the hidden population.

Across our simulations, we fix three of the parameters at constant values ($N = 10,000$; $p_F = 0.03$; $\zeta = 0.05$). We systematically explore varying the remaining three parameters, p_F, ρ, and τ_F, from 0.1 to 1 in increments of 0.1. For each combination

of the parameter values, we repeated our simulation 10 times, for a total of $10 \times 10 \times 10 \times 10 = 10,000$ iterations. For each unique combination of parameters, we averaged the results across the replications.

B.7.3 Results

In this section, we focus on a representative sample of the simulation results in order to illustrate the main conclusions while keeping the plots relatively simple. The complete simulation results are consistent with the analysis presented here.

Figure B.6 shows the estimated size of the hidden population for the generalized and basic scale-up estimators. The generalized scale-up estimator is exactly correct for all parameter combinations. The basic scale-up estimator works well when the assumptions it relies upon hold. In our simulations, these assumptions hold in special cases, such as when $p_F = 1$, $\rho = 1$, and $\tau_F = 1$. With these parameter values, (1) the social network is an Erdos-Renyi random graph (since $\rho = 1$), meaning that in expectation $\bar{d}_{H,F}/\bar{d}_{F,F} = \delta_F = 1$; (2), $U = F$ (since $p_F = 1$), meaning $\bar{d}_{U,F} = \bar{d}_{F,F}$; and, (3), there is perfect reporting (since $\tau_F = 1$), meaning $\bar{v}_{H,F} = \bar{d}_{H,F}$. Together, (1), (2), and (3) mean that the basic scale-up estimator's assumptions hold for this particular combination of parameters (Equation 3.10). In general, however, Figure B.6 shows that the basic scale-up estimator can be biased (as described in Section 3.2).

Further, our analytical results can be used to predict the bias in the basic scale-up estimator. In Section 3.2, we propose a decomposition of the difference between the basic and generalized scale-up estimators, and derive an expression for the bias in the basic scale-up estimator (Equation 3.15). Figure B.7 illustrates the empirical bias in the basic scale-up estimator from our simulations for the same combination of parameter values shown in Figure B.6. Our analytical expression (Equation 3.15) predicts the empirical bias (Figure B.7) exactly in all cases.

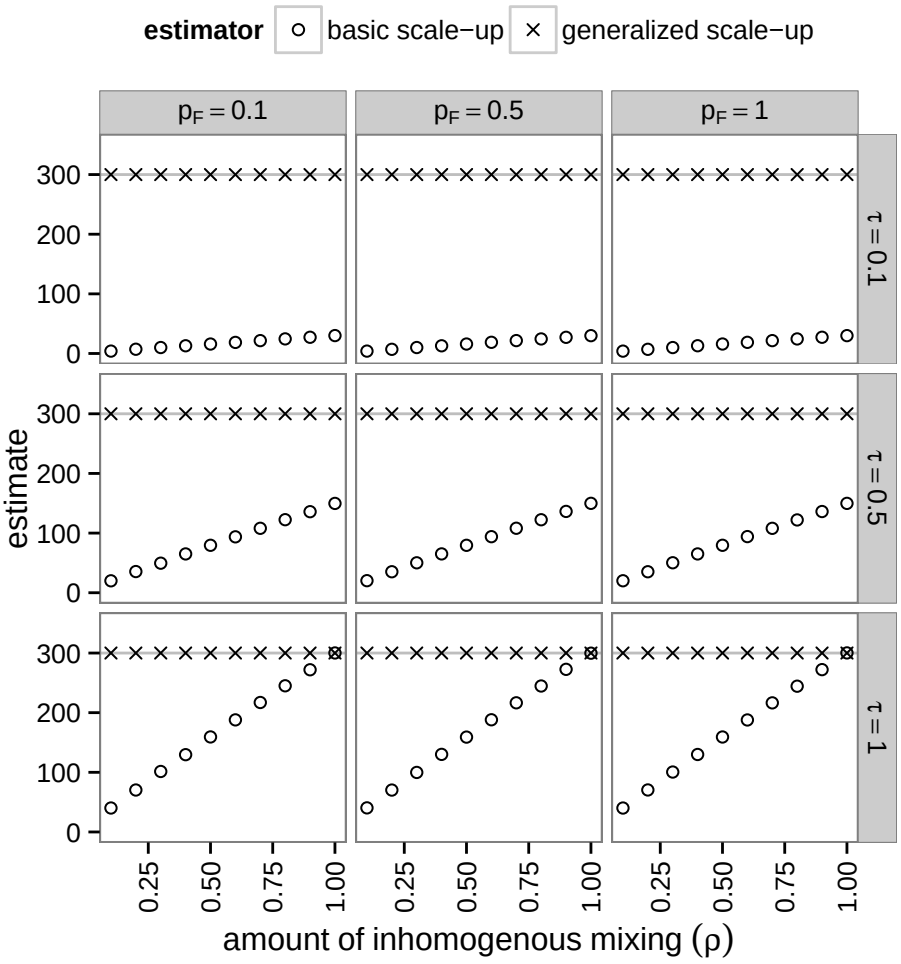


Figure B.6: Estimated size of the hidden population for the generalized and basic scale-up estimators. Each panel shows how the two estimators change as the amount of inhomogenous mixing is varied from high ($\rho = 0.1$) to low ($\rho = 1$). The columns show results for different sizes of the frame population, from small (left column, $p_F = 0.1$), to medium (middle column, $p_F = 0.5$), and large (right column, $p_F = 1$). The rows show results for different levels of reporting accuracy, from a small amount of true positives (top row, $\tau_F = 0.1$), to a moderate amount of true positives (middle row, $\tau_F = 0.5$), and perfect reporting (bottom row, $\tau_F = 1$). For example, looking at the middle of the center panel, when $p_F = 0.5$, $\tau_F = 0.5$, and $\rho = 0.5$, we see that the basic scale-up estimate is about 100, while the generalized scale-up estimate is 300. The generalized scale-up estimator is exactly correct for all parameter combinations, while the basic scale-up estimator is only correct for certain special cases (e.g., when $\rho = 1$, and $\tau_F = 1$). These results confirm and illustrate the discussion in Section 3.2.

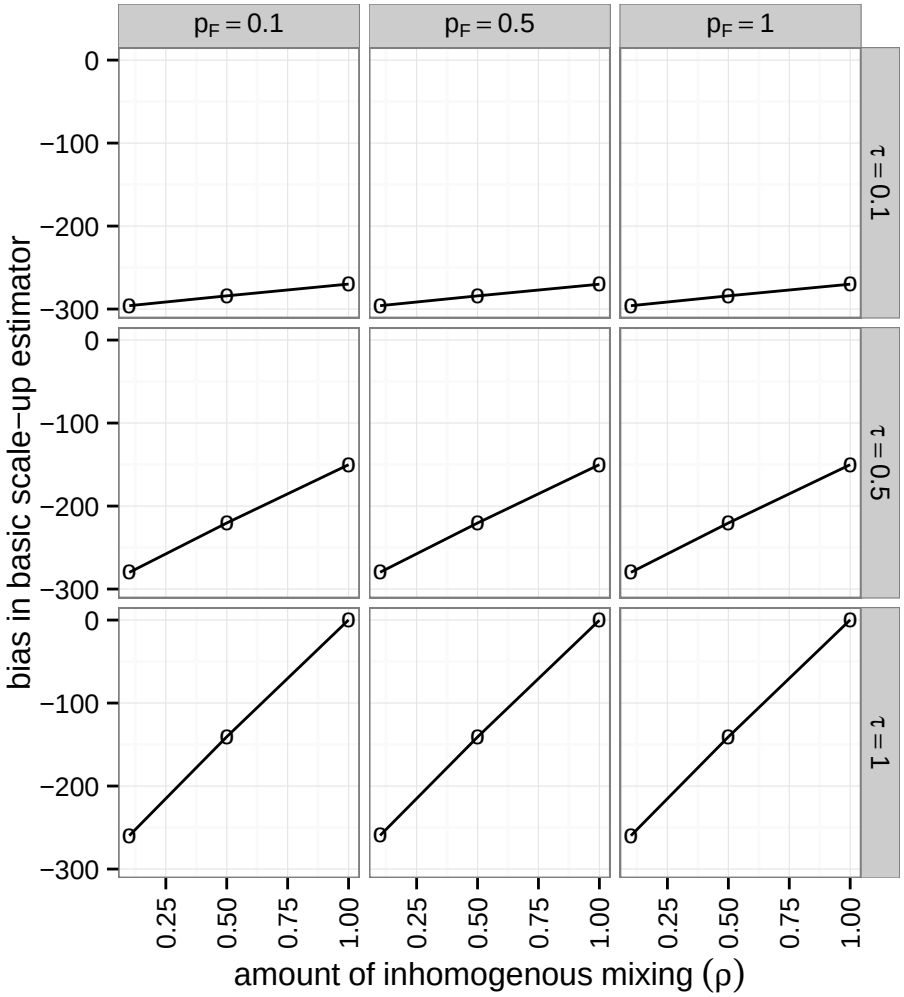


Figure B.7: Bias (solid line) and predicted bias (open circles) in the basic scale-up estimates for the same parameter configurations depicted in Figure B.6. The predicted bias from our analytical results (Equation 3.15) matches the bias observed in our simulation study exactly.

Appendix C

Tie definition appendices

C.1 Data collection and evaluation

In Appendix C.1, we give additional details about data collection and evaluation. First, we describe how we obtained information about the groups of known size for Rwanda. Next, we discuss several balance checks we conducted to ensure that randomizing the tie definition followed the study plan. Finally, we describe a manipulation check which demonstrates that respondents pay attention to the definition of a tie they are asked to report about.

C.1.1 Groups of known size

We collected information about groups of known size for Rwanda from several different sources (Table 4.2). Although there is currently no general method that researchers can use to compile their own list of groups of known size in a different setting, our experience in Rwanda suggests that this would also be feasible in other developing countries. Our goal was to follow previous scale-up studies in finding about 20 groups of known size whose prevalence in the general population varied from about 0.1% to

3% to be consistent with earlier studies. As an illustration, we briefly describe the three main strategies we used to compile groups of known size for this study.

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

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

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

C.1.2 Balance checks

In order to confirm that households were assigned to the different tie definitions according to the randomization procedure required by our study plan, we checked the balance between the two arms. These checks suggest that there is no reason to suspect that the randomization of households to tie definitions was not conducted according to the study design.

Table C.1 compares the distribution of five household-level variables between households assigned to the meal definition and households assigned to the acquaintance definition. No substantively meaningful differences are apparent. We used random permutation tests, accounting for our blocked and clustered randomization, to compute p values for the difference in household covariates between each arm; Table C.1 has the results. However, these p values are difficult to interpret because they are not independent and they involve multiple comparisons. Therefore, we also conducted an omnibus test for imbalance proposed by Hansen and Bowers (2008). This test accounts for the blocked and clustered nature of respondents' random assignment to a tie definition; furthermore, the omnibus test avoids the problem of testing multiple hypotheses, which arises when separately testing several individual covariates for balance. Finally, the omnibus test permits us to test both individual and household-level characteristics. The omnibus test included as covariates the number of respondents in the household, the five wealth quintiles, the total age of each household, the total number of females in each household, and the total number of people in each of the three education categories in each household. The Hansen and Bowers (2008) omnibus test did not find any significant evidence of imbalance (χ^2 test on 9 degrees of freedom produced a two-sided $p = 0.22$).

Together, the results of these balance checks lead us to conclude that the randomization of households to different definitions of a network tie was conducted according to the study design.

Table C.1: Comparison between households assigned to the acquaintance tie definition and households assigned to the meal tie definition. All values are fractions, with the exception of the average number of respondents. P-values are from non-independent random permutation tests. Sampling weights were not used in computing these means.

	Acquaintance	Meal	two-sided p
Num. respondents in hh	2.263	2.209	0.335
Lowest wealth quintile	0.088	0.116	0.054
Second wealth quintile	0.151	0.153	0.946
Third wealth quintile	0.203	0.208	0.957
Fourth wealth quintile	0.210	0.210	1
Highest wealth quintile	0.348	0.312	0.086

C.1.3 Manipulation checks

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

C.2 Tie strength and reporting accuracy: formal models

In this appendix, we describe two mathematical models that lead to a trade-off between the quantity and quality of reports that is qualitatively similar to the behavior illustrated in Figure 4.1. We view this model as a starting point for reasoning about the potential trade-offs involved in choosing a tie definition to ask respondents to report about; we hope that future empirical work will complement and improve these formal models.

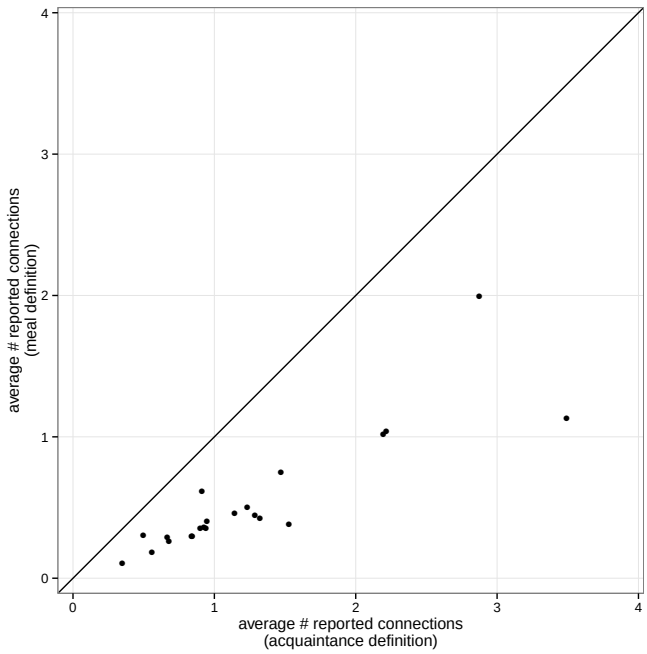


Figure C.1: Average number of connections by tie definition for 22 groups of known size (Table 4.2). On the x axis is the average number of connections from the experimental arm that used the acquaintance definition of a network tie, and on the y axis is the average number of connections from the arm that used the meal definition. The average number of reported connections for the meal definition is consistently lower than for the acquaintance definition.

We are interested in estimating the size of group H, N_H, using an estimator $\widehat{N}_H$. Since we wish to analyze how total error is related to the tie definition respondents are asked to report about, we will consider the sampling design and the sample size fixed.

The *total error* of the estimator $\widehat{N}_H$, quantified as mean squared error (MSE), is

$$\text{MSE}(\widehat{N}_H) = \mathbb{E}[(\widehat{N}_H - N_H)^2] \tag{C.1}$$

$$= \sigma^2 + \beta^2, \tag{C.2}$$

where the expectation is taken over the sampling distribution. Equation C.2 is sometimes called the *bias-variance decomposition*. This decomposition shows how total error (MSE) is the sum of sampling error (σ^2) and non-sampling error (β^2). The

sampling error is captured by the sampling variance of the estimator,

$$\sigma^2 = \mathbb{E}[(\widehat{N}_H - \mathbb{E}[N_H])^2], \tag{C.3}$$

and the nonsampling error is captured by the squared bias of the estimator,

$$\beta^2 = (\mathbb{E}[\widehat{N}_H] - N_H)^2. \tag{C.4}$$

Our formal model starts from the identity in Equation C.2. The idea is to add assumptions about how each of the two terms—the squared bias and the variance—will react as the strength of the tie definition t is increased or decreased.

Granovetter (1973) provides a commonly used qualitative definition of the concept of tie strength. We will not provide a rigorous formalization of a quantitative notion of tie strength itself here; we consider this topic an interesting direction for further work[1]. Therefore, we start by assuming that the strength of a network tie has been formalized so that a particular tie definition can be assigned a value $t \in T$, where T is a range of possible tie strengths $T = [t_{\min}, t_{\max}]$. For convenience, we take $t_{\min} = 1$. Note that, somewhat counter-intuitively, we take lower values of t to mean stronger ties, and higher values of t to mean weaker ties. We formalize our assumptions about the relationship between tie strength, the squared bias, and the variance using functions that map tie strength to error. Specifically, we model the nonsampling error β^2 with $f_n(t)$ and the sampling error σ^2 with $f_s(t)$. The error of our estimator $\widehat{N}_H$ is thus a function of t:

$$f(t) = f_n(t) + f_s(t) \tag{C.5}$$

[1]Such an approach could follow the example of theoretical economists, who have developed the machinery needed to understand precisely when a set of preference relations—or, in our case, tie strength relations— can be embedded as continuous function on the real line; see Varian (1992) or Mas-Colell et al. (1995) for examples.

for $t \in T$.

Given the general structure of this model, we now show how adding more detailed assumptions permits us to learn more and more from it. Each of the assumptions could be investigated by future empirical research, leading to improvements to our model.

First, assume only that $f_n(t)$ and $f_s(t)$ are smooth (i.e., twice-differentiable) functions of t. We wish to understand when we can expect $f(t)$ to be at a minimum; at such a minimum, our estimator will have the smallest mean squared error. At a tie strength t^* which is a local minimizer of f, the first derivative of f will be zero, i.e., $f'(t^*) = f'_n(t^*) + f'_s(t^*) = 0$. So we therefore have the first order condition:

$$f'_s(t^*) = -f'_n(t^*). \tag{C.6}$$

The second-order condition for t^* to be a minimizer is that

$$f''(t^*) = f''_n(t^*) + f''_s(t^*) \geq 0. \tag{C.7}$$

For any smooth functions f_n and f_s, a point t^* that satisfies Equation C.6 and Equation C.7, will be a local minimum. Moreover, if there is no such point t^* in the interior of T, then we are guaranteed that one of the endpoints $t_{\min}$ or $t_{\max}$ will be a minimum.

This fact alone—which is based on minimal assumptions—suggests that it may be worthwhile to explore different definitions of a network tie: any non-trivial shape to $f_s(t)$ or $f_n(t)$ could lead to different total errors as a function of tie strength. Even if finding a global minimum—i.e., the best tie definition— is difficult, researchers can still seek to improve network reporting estimates by identifying a tie definition that produces locally minimal total error. Note also that if we make the additional assumption that both f_n and f_s are convex in t, so that $f''_n(t) > 0$ and $f''_s(t) > 0$ for all t, then we are guaranteed to have a unique global minimum t^*. Therefore, a

quite general formal model suggests that there is good reason to explore alternate tie definitions, and that additional, potentially testable assumptions (the convexity of $f_n(t)$ and $f_s(t)$) can lead to a global minimum.

C.2.1 Parametric model

We now provide an example of how adding plausible, more concrete (but not-yet-tested) assumptions about the shapes of f_s and f_n allows us to learn more about potential trade-offs between the quantity and quality of information produced by network reports.

For a given estimator and sampling design, the sampling variance is typically a function of the sample size. In our model, we hold sample size – the number of people interviewed by the survey – fixed. However, if we interview n people using a network reporting method, we actually get information about more (potentially many more) than just those n people. We call the size of this set of people we learn about the *effective sample size*.

f_s' is the rate of change in the sampling error (i.e., sampling variance) that comes from weakening the strength of the tie. Empirical evidence suggests that networks formed from stronger ties tend to have fewer total connections than networks of weaker ties, meaning that we expect weakening the tie definition will produce more information from each respondent. In that sense, the effective sample size will increase and therefore we expect the sampling error to decrease; in other words, we expect $f_s'(t) < 0$.

The exact form of $f_s'(t)$ will likely be complex, as it will depend on the structure of the underlying network and possibly several other factors. However, from the theory of sampling, we know that formulas for sampling variance usually have the form $O(n^{-1})$, where n is the sample size. We predict that there will be a similar dynamic at play with effective sample size: as the tie definition weakens, the increased size of

respondents' networks leads to an increase in the effective sample size, and therefore a decrease in sampling error. To be concrete, we will model the sampling error as

$$f_s(t) = K_s \, t^{-1}. \tag{C.8}$$

Equation C.8 says that sampling error behaves like t^{-1}. Moreover, if we choose the scale for t so that $t = 1$ is asking the respondents' to report about themselves—in a sense, the strongest possible tie—then the constant K_s can be interpreted as the sampling error in the standard survey estimator where respondents report about themselves. Further, we are also free to scale the quantitative representation of tie definition in any way we please, and it may be conceptually helpful to imagine that the units of t have been chosen so that increasing t by 1 (at least at a particular reference value of t, t_r) results in a change in sampling variance equal to adding a single extra survey respondent.

f_n' is the rate of change in the nonsampling error that comes from weakening the strength of the tie. Empirical evidence suggests that respondents tend to know more about people they are connected to through stronger ties. Assuming respondents answer survey questions honestly, this implies that weakening the tie definition leads to less accurate reports, and more nonsampling error; in other words, we expect $f_n'(t) > 0$. Additionally, we might speculate that this change in nonsampling error as tie strength weakens will not be uniform; instead, we predict that moving from a very strong tie to a moderately strong tie will make more of a difference than moving from a moderately strong tie to a weak tie. Mathematically, this means that we might expect the change in nonsampling error to behave like a concave function. We will model the nonsampling error as

$$f_n(t) = K_n \left(\log t - 1 \right). \tag{C.9}$$

We stress that Equations C.8 and C.9 are not reality. However, they both capture a specific set of assumptions that we consider plausible in many situations. Of course, empirical work could check and improve them.

Under this model, the total error is

$$f(t) = f_s(t) + f_n(t) \tag{C.10}$$

$$= K_s t^{-1} + K_n(\log t - 1) \tag{C.11}$$

Substituting $f_s'(t)$ and $f_n'(t)$ into the first order conditions we derived above, we have that at a minimum tie strength t^*,

$$\frac{K_n}{t^*} = \frac{K_s}{t^{*2}} \tag{C.12}$$

$$\iff t^* = \frac{K_s}{K_n}. \tag{C.13}$$

Checking the second-order conditions, we have

$$f_n''(t^*) + f_s''(t^*) = \frac{1}{t^{*2}}\left[\frac{2K_s}{t^*} - K_n\right] \geq 0. \tag{C.14}$$

Whenever Equation C.14 is satisfied, our solution $t^* = \frac{K_s}{K_n}$ is a minimum. But this is always the case, since $\frac{1}{t^{*2}} > 0$ and $2K_s - t^*K_n = 2K_s - K_s > 0$ as well.

In sum, this much stronger set of assumptions—i.e., particular parametric forms for $f_n(t)$ and $f_s(t)$—led us to a much more specific conclusion: here, tie strength will be at a global minimum at $t^* = \frac{K_s}{K_n}$. Note that we obtained this result even though we had a concave functional form for the nonsampling error $f_n(t)$. Of course, other choices for the functional forms would not have produced such a clean result. Further formal development could be accompanied by empirical work to evaluate and improve the specifications for $f_n(t)$ and $f_s(t)$.

C.3 Comparisons with other estimates

In Appendix C.3, we provide additional details needed to compare our estimates to earlier estimates from Rwanda and to benchmarks from the Joint United Nations Programme on HIV/AIDS (UNAIDS). First, we present the UNAIDS benchmarks and describe how they are constructed (Section C.3.1). Next, we give additional detail about the comparison with previous studies that was discussed in the main text (Section C.3.2). Finally, we compare the estimates from our study with both the UNAIDS benchmarks and the previous studies from Rwanda (Section C.3.3).

C.3.1 UNAIDS benchmark estimates

The UNAIDS benchmark estimates are derived from published literature on key population size estimates from around the world. UNAIDS benchmark estimates for the sizes of hidden populations are presented as a prevalence within some portion of the entire population. For example, the benchmark estimated number of sex workers in a country in Sub-Saharan Africa is 0.4% to 4.3% of females age 15-49 (UNAIDS, 2014). Table C.2 reproduces UNAIDS benchmark ranges suggested for Rwanda for all of the hidden populations except for men who have sex with men, for which no benchmarks are available for African countries. For injecting drug users, the UNAIDS guidelines do not specify the portion of the population to which the estimates apply, so we assume it to be both men and women aged 15-49. In order to estimate quantities such as the number of women between 15-49, we take the age-sex distribution from the United Nations Population Division and apply it to the total population estimate for Rwanda in 2011, which was provided to the study team by the National Institute of Statistics of Rwanda (Table C.3). The resulting benchmark estimates for Rwanda are shown in Table C.4.

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

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

Table C.2: UNAIDS benchmark ranges for the populations most at-risk for HIV/AIDS from UNAIDS (2014, pg. 22).

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

Table C.3: Population estimates used to produce UNAIDS benchmark values, and also to produce hidden population size estimates. UNPD is the United Nations Population Division estimate of Rwanda's age-sex distribution in 2010 (UN, 2012), and NISR is the National Statistical Institute of Rwanda, who provided the study team with an estimate for Rwanda's total population in 2011.

C.3.2 Previous studies of key populations at risk for HIV in Rwanda

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

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

Table C.4: Benchmark estimates for hidden populations in Rwanda computed from the information in Table C.2 and Table C.3.

Table C.5 compares the definitions of the hidden population used in our study and in the three other studies to which we compare. For female sex workers, there are two existing estimates. The first is from a capture-recapture study. The capture was the 2010 Behavioural Surveillance Survey (BSS) of female sex workers (Rwanda Biomedical Center, 2010). The recapture was part of a mapping of sex workers in which enumerators asked sex workers whether or not they participated in the BSS (RBC/IHDPC et al., 2012). The second existing estimate for the number of female sex workers is based on a participatory mapping of female sex workers in Rwanda in 2011 and 2012 (RBC/IHDPC et al., 2012).

For the number of male clients of sex workers, the existing estimate comes from a direct question posed to a nationally representative sample of men in the 2005 Rwanda DHS (INSR and ORC Macro, 2006). The survey asked men whether they had paid for sex in the past 12 months and also whether they had ever paid for sex. The DHS survey requires that these questions be asked when no other person is within hearing distance of the interview. Despite this precaution, men are likely to underreport this behavior while being interviewed in their own homes. The survey found that 0.83 percent of men reported having paid for sex in the 12 months prior to the survey. Assuming that this proportion has not changed between 2005 and 2011, we can multiply 0.83 percent by the male population aged 15-49 (Table C.3) to estimate that approximately 20,142 men 15-49 paid for sex in the preceding 12 months (RBC/IHDPC et al., 2012).

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

C.3.3 Comparison with UNAIDS benchmarks and previous Rwanda studies

Finally, we compare our study's estimates to both the UNAIDS benchmarks and to the results of the previous studies (Figure C.2). Our estimates are generally higher than the ones produced by the three comparison studies, but lower than the UNAIDS benchmarks, with the exception of sex workers, where our estimates are within the UNAIDS benchmark range, and men who have sex with men, where no comparisons and no UNAIDS benchmarks are available.

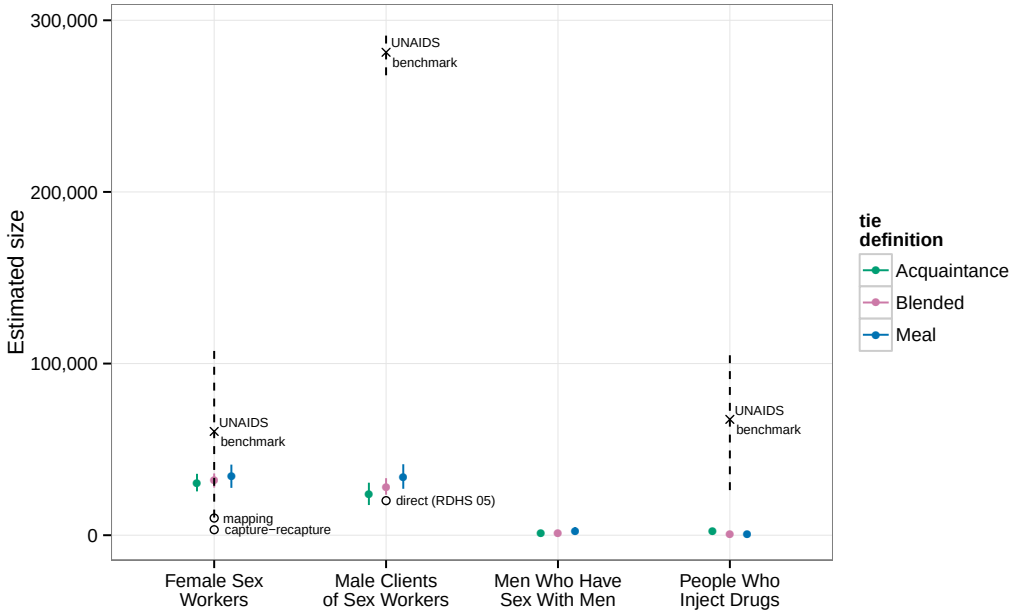


Figure C.2: Comparison between our estimates, the UNAIDS benchmarks, and the estimates from other studies in Rwanda. Our estimates are higher than all earlier estimates from Rwanda and are either lower than or within the bounds given by the UNAIDS benchmarks.

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

Table C.5: Question wording used to estimate the size of four key populations at risk for HIV in Rwanda. Differences in the precise definition of membership in the hidden population may contribute to discrepancies in size estimates. Fsw is female sex workers, csw is male clients of sex workers, msm is men who have sex with men and idu is injecting drug users.

C.4 Linear blending

In this section, we describe our strategy for combining two arms of a survey experiment. Our approach is similar to methods used in meta-analysis (Hedges and Olkin, 1985), and the result is similar to a Bayesian shrinkage estimator.

Result C.4.1 *Suppose we have two estimators for N_H, $\widehat{N}_H^{(A)}$ and $\widehat{N}_H^{(B)}$. Suppose that the two estimators have sampling variance σ_A^2 and σ_B^2, and that they are unbiased, so that $\mathbb{E}[N_H^{(A)}] = N_H$ and $\mathbb{E}[N_H^{(B)}] = N_H$. Finally, suppose that the two estimators are independent, so that $cov(\widehat{N}_H^{(A)}, \widehat{N}_H^{(B)}) = 0$. Consider the set of all possible estimates*

$$\widehat{N}_H = w\ \widehat{N}_H^{(A)} + (1-w)\ \widehat{N}_H^{(B)}, \tag{C.15}$$

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

$$w^\star = \frac{\sigma_B^2}{\sigma_A^2 + \sigma_B^2}. \tag{C.16}$$

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

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

$$\text{var}(\widehat{N}_H) = \text{var}(w\widehat{N}_H^{(A)} + (1-w)\widehat{N}_H^{(B)}) \tag{C.17}$$

$$= w^2\sigma_A^2 + (1-w)^2\sigma_B^2. \tag{C.18}$$

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

$$\mathrm{MSE}(\widehat{N}_H) = w^2\sigma_A^2 + (1-w)^2\sigma_B^2. \tag{C.19}$$

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

$$\frac{\partial \mathrm{MSE}(\widehat{N}_H)}{\partial w} = 2w\sigma_A^2 - 2(1-w)\sigma_B^2 \tag{C.20}$$

$$= 2w(\sigma_A^2 + \sigma_B^2) - 2\sigma_B^2. \tag{C.21}$$

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

$$w^\star = \frac{\sigma_B^2}{\sigma_A^2 + \sigma_B^2}. \tag{C.22}$$

We did not formally include the constraint that $w \in [0,1]$ in our derivation, but since variances are always non-negative, $w^\star$ in Equation C.22 will always satisfy this condition. We can confirm that $w^\star$ is a minimum by differentiating Equation C.21 again to obtain

$$\frac{\partial^2 \mathrm{MSE}(\widehat{N}_H)}{\partial w^2} = 2(\sigma_A^2 + \sigma_B^2). \tag{C.23}$$

As long as we are not in the degenerate case where both variances are identically zero, Equation C.23 is always greater than 0, meaning that Equation C.22 is indeed a minimum. (In the degenerate case, the two estimators would always produce identical, exactly correct estimates, since they would both be unbiased estimators with 0 variance.) ■

Finally, we address the question of when it is advantageous to blend the estimates from both arms of an experiment, instead of just using the estimate from one arm. In general, we prefer the blended estimator when we expect it to produce smaller mean

squared error than either of the individual arms. Without loss of generality, suppose that experimental condition A outperformed condition B, meaning that we estimate $\text{MSE}_A \leq \text{MSE}_B$. We want to blend as long as $\text{MSE} \leq \text{MSE}_A$. Plugging $w^\star$ into our expression for the mean squared error, Equation C.19, we have

$$\text{MSE} = (w^\star)^2\sigma_A^2 + (1 - w^\star)^2\sigma_B^2 \tag{C.24}$$

$$= \sigma_B^2 \frac{\sigma_A^2\sigma_B^2}{(\sigma_A^2 + \sigma_B^2)^2} + \sigma_A^2 \frac{\sigma_A^2\sigma_B^2}{(\sigma_A^2 + \sigma_B^2)^2} \tag{C.25}$$

$$= \frac{\sigma_A^2\sigma_B^2}{\sigma_A^2 + \sigma_B^2}. \tag{C.26}$$

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

$$\frac{\sigma_A^2\sigma_B^2}{\sigma_A^2 + \sigma_B^2} \leq \sigma_A^2 \tag{C.27}$$

$$\iff \frac{\sigma_B^2}{\sigma_A^2 + \sigma_B^2} \leq 1. \tag{C.28}$$

Since $\sigma_A^2 > 0$ and $\sigma_B^2 > 0$ the inequality in Equation C.28 will always hold. Our conclusion is that, assuming the estimators from each arm are unbiased and independent, blending always produces lower (or equal) expected mean squared error than just choosing the estimate from one arm.

C.5 Framework for sensitivity analysis

The scale-up literature has long discussed different types of potential bias in basic scale-up estimates (Feehan and Salganik, 2015; Maltiel et al., 2015; Bernard et al., 2010; McCormick and Zheng, 2007; Killworth et al., 2003, 2006; Shelley et al., 2006, 1995; Zheng et al., 2006; Cowan, 2014). In Section 3.2 and Appendix B.1, we developed a framework for assessing the sensitivity of scale-up estimates to these potential sources of bias. The framework decomposes the multiplicative (see Equations 4.5 and

4.6):

$$N_H = \widehat{N}_H \left(\frac{\eta_F}{\tau_F}\right)\left(\frac{1}{\phi_F \delta_F}\right) \tag{C.29}$$

The estimates produced by the basic scale-up estimator – which was used in many previous studies, and which we use throughout the main text – make the assumption that the product of these adjustment factors is 1. Equation 3.10 shows that if this assumption does not hold, the estimates will be biased. For example, the results of the internal consistency checks in Figure 4.3 show that our internal consistency check estimate for the number of teachers is too high. This result suggests that the product of the four adjustment factors is less than 1. This could happen if, for example, teachers have more average connections to members of the frame population than members of the frame population have to each other, leading the degree ratio, $\delta_F = \bar{d}_{H,F}/\bar{d}_{F,F}$ to be greater than 1.

Previous research has discussed several potential sources of bias in basic scale-up estimates, including (i) structural differences between the networks of hidden population members and the general population, which have been called *barrier effects* (Killworth et al., 2006; Zheng et al., 2006; Maltiel et al., 2015); (ii) respondents who are not perfectly aware of the characteristics of their network alters, which has been called *transmission error* (Shelley et al., 1995, 2006; Killworth et al., 2006; Salganik et al., 2011b; Maltiel et al., 2015); and (iii) respondents who make errors in their network reports, which is called *recall error* (Killworth et al., 2003, 2006; McCormick and Zheng, 2007; Maltiel et al., 2015). All three of these possible sources of bias are accounted for by the framework summarized in Equation 3.10: barrier effects will impact the degree ratio δ_F; and transmission and recall errors will impact the reporting terms, η_F and τ_F.

The results above and the results in the main text all assume that the conditions for the known population estimator are met (see Chapter 3 and Feehan and Salganik (2015)). If those conditions are not met, then researchers could use the robustness results in Chapter 3 to construct additional adjustment factors.

C.5.1 Producing estimates under alternative assumptions

Researchers can use the theoretical results from Chapter 3, which are summarized in Equation 3.10, and the empirical results from our study (Table C.6) to produce their own estimates under different assumptions about the magnitude of potential biases. Researcher can also apply the blending weights derived in Section C.4 to estimates based on their own assumptions to produce a blended estimate, $\widehat{N}_H^{\dagger}$.

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

Table C.6: Quantities needed for sensitivity analysis.

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

Table C.7: Possible assumptions a researcher might make about ϕ_F, δ_F, τ_F, and η_F for estimating the number of female sex workers from each arm of the study.

As a concrete example, suppose a researcher thinks that the most plausible values for each ϕ_F, δ_F, τ_F, and η_F, are the ones in Table C.7. Using these values, the implied

value for the overall adjustment factor for the arm with the meal definition, $\alpha_F^{(A)}$ is

$$\alpha_F^{(A)} = \frac{\eta_F^{(A)}}{\phi_F^{(A)}\ \delta_F^{(A)}\ \tau_F^{(A)}} \tag{C.30}$$

$$= \frac{0.99}{1.30 \times 0.90 \times 0.80} \tag{C.31}$$

$$\approx 1.06 \tag{C.32}$$

This means that the unbiased estimator produces

$$\widehat{N}_H^{(A)\dagger} = \alpha_F^{(A)} \times \widehat{N}_H^{(A)} \tag{C.33}$$

$$= 1.06 \times 34,213 \tag{C.34}$$

$$\approx 36,187 \tag{C.35}$$

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

$$\alpha_F^{(B)} = \frac{\eta_F^{(B)}}{\phi_F^{(B)}\ \delta_F^{(B)}\ \tau_F^{(B)}} \tag{C.36}$$

$$= \frac{0.95}{1.30 \times 0.50 \times 0.40} \tag{C.37}$$

$$\approx 3.65 \tag{C.38}$$

In this case, the unbiased estimator produces

$$\widehat{N}_H^{(B)\dagger} = \alpha_F^{(B)} \times \widehat{N}_H^{(B)} \tag{C.39}$$

$$= 3.65 \times 30,466 \tag{C.40}$$

$$\approx 111,320 \tag{C.41}$$

Next, we note that, based on the estimated sampling variances, the optimal blending weight, $w^\star$ (Equation C.16), is:

$$w^\star = \frac{\sigma_B^2}{\sigma_A^2 + \sigma_B^2} \tag{C.42}$$

$$= \frac{6,810,654}{11,894,878 + 6,810,654} \tag{C.43}$$

$$\approx 0.36 \tag{C.44}$$

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

$$\widehat{N}_H^\dagger = w\widehat{N}_H^{(A)\dagger} + (1-w)\widehat{N}_H^{(B)\dagger} \tag{C.45}$$

$$= 0.36 \times 36,187 + (1 - 0.36) \times 111,320 \tag{C.46}$$

$$= 83,964 \tag{C.47}$$

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

C.6 Bootstrap procedure

In order to compare the internal consistency results for the two tie definitions, and to put intervals around our estimates, we obtained 2,000 bootstrap resamples of our dataset using the rescaled bootstrap method, which accounts for our study's complex sample design (Rust and Rao, 1996; Rao and Wu, 1988b; Feehan and Salganik, 2015). We constructed 95% confidence intervals from the central 95% of the bootstrap-resampled blended estimates. From these 2,000 bootstrap resamples we also estimated the sampling variance of each estimate. These estimated sampling variances were used

to compute the estimated optimal weights (Equation 4.4), and, finally, we produce a single blended estimate using linear blending (Equation 4.3).

The procedure for producing a measure of uncertainty of the blended estimate was a bit more complex (Figure C.3). We took $M = 2,000$ bootstrap resamples $s_A^{(1)}, \ldots, s_A^{(M)}$ and $s_B^{(1)}, \ldots, s_B^{(M)}$ from each experimental arm. Within the ith bootstrap resample for each arm, we computed a size estimate $\widehat{N}_H^{(A)(i)}$ and $\widehat{N}_H^{(B)(i)}$. Also within the ith bootstrap resample for each arm, we obtained M bootstrap resamples $s_B^{\star\star(i,1)}, \ldots, s_B^{\star\star(i,M)}$ and $s_B^{\star\star(i,1)}, \ldots, s_B^{\star\star(i,M)}$, which we use to estimate the sampling variance $\widehat{\sigma}_A^{2\ (i)}$ and $\widehat{\sigma}_B^{2\ (i)}$. We use Equation C.16 to estimate the blending weight $w^{(i)}$ from these estimated sampling variances. Finally, we produce a blended estimate $\widehat{N}_H^{(i)}$ for the ith bootstrap resample using $w^{(i)}$. The uncertainty in our final blended estimate $\widehat{N}_H$ is computed using the percentile method: we take the upper and lower boundaries of the central 95% of our bootstrap estimates $\widehat{N}_H^{(1)}, \ldots, \widehat{N}_H^{(M)}$ as our uncertainty interval (Efron and Tibshirani, 1993). At all steps we used the rescaled bootstrap method, which accounts for the study's complex sample design (Rust and Rao, 1996)

We also performed the internal consistency checks on each set of $M = 2,000$ bootstrap resamples to reach our conclusion that the meal definition produces lower error than the acquaintance definition by three error metrics: the mean squared error, the average relative error, and the mean absolute error. For each bootstrap resample, for each arm, and for each error metric, we produce a summary of the error across the groups of known size. Figure C.4 show our results. For all three error metrics, and for all bootstrap resamples, the meal definition attains a lower error than the acquaintance definition ($p < 0.001$).

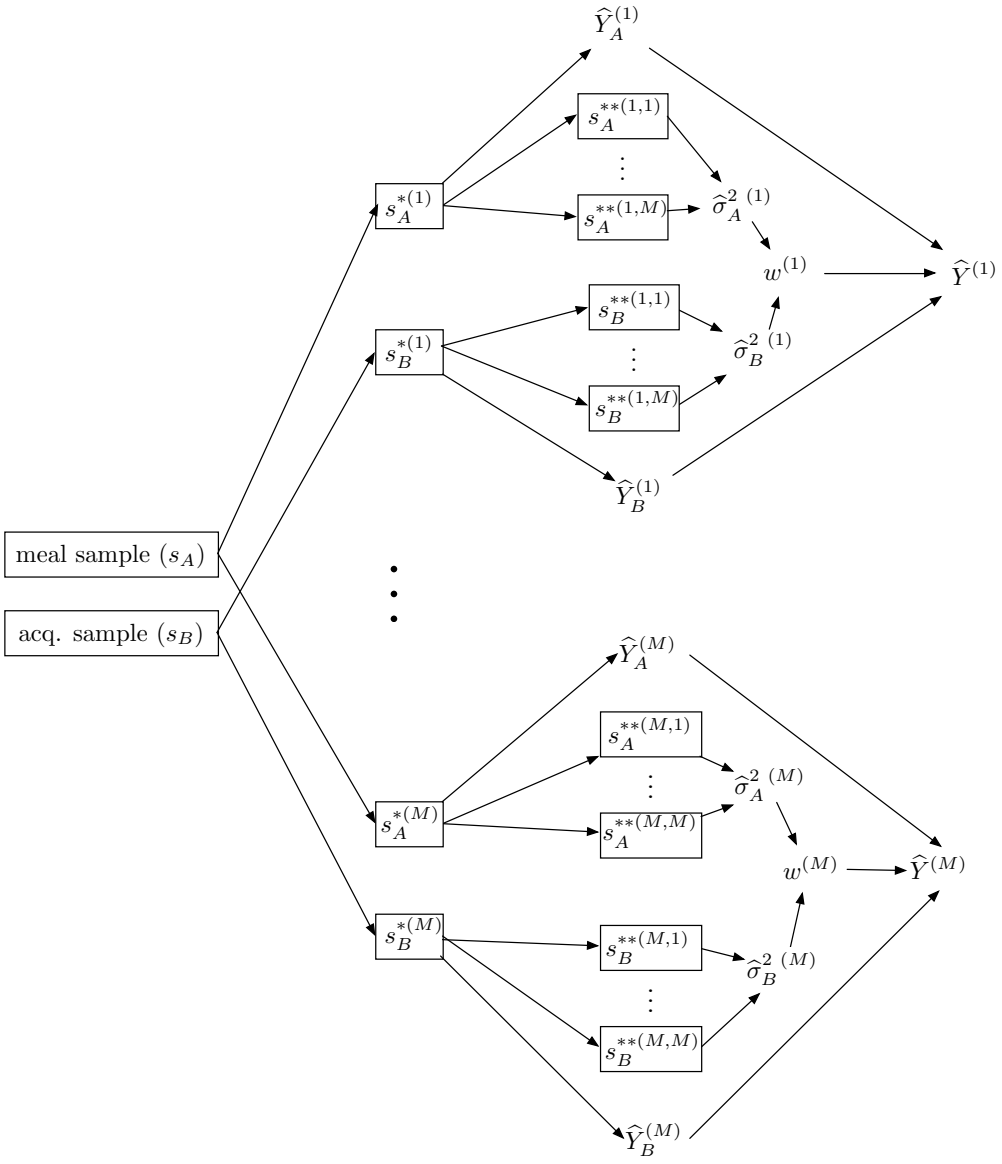


Figure C.3: Illustration of the procedure used to estimate the sampling uncertainty in the blended estimates. We took $M = 2,000$ bootstrap resamples from each experimental arm. Within each bootstrap resample, a size estimate is computed. Also within each bootstrap resample, an additional M resamples are taken to estimate the sampling variance. The blending weight is computed from the estimated sampling variance for each arm, and a blended estimate is formed. We end up with M blended estimates, and we take the middle 95% for our uncertainty interval.

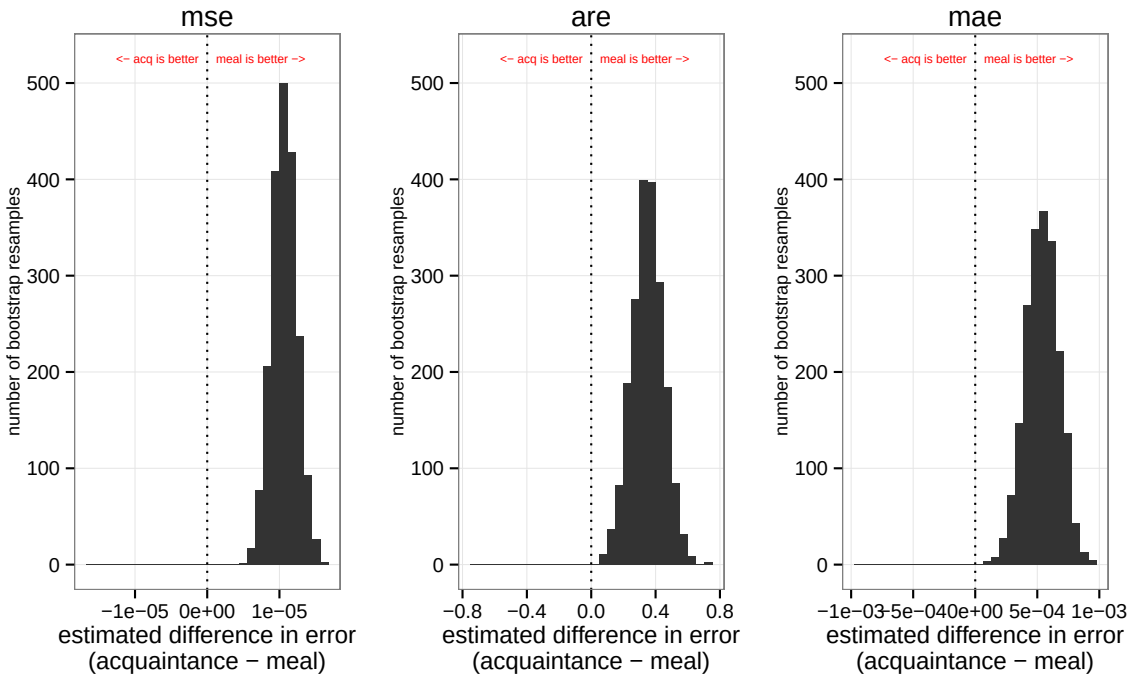


Figure C.4: Three summaries of the errors in estimates from the internal consistency checks. Results are shown for, from left to right, mean squared error, average relative error, and mean absolute error. If the true known population sizes are $X_1, \ldots, X_K$ and our estimated sizes are $\widehat{X}_1, \ldots, \widehat{X}_K$, then the mean square error is $\text{MSE} = \frac{1}{K}\sum_{i=1}^{K}(\widehat{X}_i - X_i)^2$; the average relative error is $\text{ARE} = \frac{1}{K}\sum_{i=1}^{K}\frac{\widehat{X}_i}{X_i}$; and the mean absolute error is $\text{MAE} = \frac{1}{K}\sum_{i=1}^{K}|\widehat{X}_i - X_i|$. Each panel shows the distribution, across bootstrap resamples, of the difference in estimated total error between the acquaintance and meal definitions. The dashed vertical line is at 0, which is what we would expect to see if there was no difference in error between the two tie definitions. Positive values, shown to the right of the dashed line, indicate bootstrap resamples for which the meal definition performed better by attaining lower error. The meal definition outperforms the acquaintance definition on all three error criteria.

Appendix D

Rwanda mortality appendices

D.1 Estimating personal network size

The network survival estimator uses the personal networks of survey respondents in demographic group α to estimate the visibility of deaths in demographic group α. This approach requires a method for estimating the average personal network size of survey respondents in demographic group α, $\bar{d}_{F_\alpha,F}$. In the first part of this appendix, we review some concepts that are needed to understand a probability sample, which we use throughout the remaining appendices. Then, in the next part, we adapt an existing personal network size estimator, called the known population method, so that it can be used to estimate $\bar{d}_{F_\alpha,F}$. In the final section, we describe the known populations that we used in the Rwanda study. Most of the contents of this appendix closely parallel the formal analysis of the known population estimator in Feehan and Salganik (2015, Appendix B.3).

D.1.1 Probability samples

Many of our results require that we have a probability sample taken from the frame population. We repeat Sarndal et al. (2003)'s definition of a probability sample here

for convenience. Suppose we have a set of possible samples, $\{s_1, \ldots, s_j, \ldots, s_{\max}\}$, with each $s_j \subset F$. Furthermore, suppose $p(s_j)$ gives the probability of selection for each possible sample s_j. If we select a sample s at random using a process that will produce each possible sample s_j with probability $p(s_j)$, and if every element $i \in F$ has a nonzero probability of inclusion $\pi_i > 0$, then we say that we have selected a *probability sample* and we call $p(\cdot)$ the *sampling design*.

We also need some more notation for the results developed here and in later appendices. Let the size of the frame population F be N_F, and let the number of frame population members who are also in demographic subgroup α be N_{F_α}. Also, for a particular demographic group α, and a sample s, let s_α be the members of the sample who are also members of α.

Some of our results are based on ratio estimators. These are standard in survey research, and a discussion of them can be found in many texts (see, e.g. Sarndal et al., 2003, chap. 5). Ratio estimators are not, strictly speaking, unbiased. However, there is a large literature that confirms that the bias in ratio estimators is typically very small when samples are not too small; this is what we use by the term *essentially unbiased* instead of unbiased in several of the results below. For example, Sarndal et al. (2003, chap. 5) shows that, under simple random sampling, the bias ratio of the standard ratio estimator goes to 0 as $n^{-1/2}$, where n is the sample size. Feehan and Salganik (2015, Appendix E) and Rao and Pereira (1968) have analogous analyses for the compound ratio estimator. Furthermore, Feehan and Salganik (2015) has an empirical study from a practical network reporting context; this empirical study suggests that any ratio bias can be expected to be extremely small.

D.1.2 Reporting about multisets

We now introduce the concept of reporting about multisets, which is important for understanding the known population method (Feehan and Salganik, 2015, Appendix

B.3). Suppose we have groups of known size $A_1, \ldots, A_J$, with $A_j \subset U$ for each j. Imagine concatenating all of these groups together, repeating each individual once for each known population she is a member of. The result, which we call the *probe alters*, $\mathcal{A}$, is a multiset. The size of $\mathcal{A}$ is $N_{\mathcal{A}} = \sum_j N_{A_j}$, where N_{A_j} is the size of known population j.

Let y_{i,A_j} be the number of members of group A_j that respondent i reports having in her personal network; let $y_{i,\mathcal{A}} = \sum_j y_{i,A_j}$ be the sum of respondent i's reported connections to all of the probe alters; and let $y_{F,\mathcal{A}} = \sum_{i \in F} \sum_j y_{i,A_j}$ be the total number of reports from F to $\mathcal{A}$ when all of F is interviewed (ie, in a census). Similarly, for network degrees, let $d_{i,\mathcal{A}} = \sum_j d_{i,A_j}$, the sum of the number of network connections between i and each of the known populations; also, let $d_{F,\mathcal{A}} = \sum_{i \in F} d_{i,\mathcal{A}}$. For all of the quantities above, we will refer to the subset of F that is also in demographic group α as F_α, and we will refer to the subset of the sample, s, that is also in demographic group α as s_α. So, for example, $y_{s_\alpha,\mathcal{A}} = \sum_{i \in s \cap \alpha} y_{i,\mathcal{A}}$.

D.1.3 Adapted known population estimator

The known population estimator of Killworth et al. (1998a) uses aggregate relational data about respondents' connections to groups whose size is known to estimate the average size of personal networks for the entire frame population ($\bar{d}_{F,F}$). In this section, we adapt the known population estimator so that it can also produce estimates of the average size of personal networks for subgroups of the frame population ($\bar{d}_{F_\alpha,F}$). But, before the adapted known population estimator can be introduced, it is necessary to derive a general property of reports about a multiset, $\mathcal{A}$. This property says that we can use reports about connections to the known populations $\mathcal{A}$ from survey respondents in demographic group α to estimate the total number of connections from frame population members in demographic group α to the known populations.

Property D.1.1 *Suppose we have a probability sample s taken from the frame population with known probabilities of inclusion π_i. Then for a particular demographic group α,*

$$\widehat{y}_{F_\alpha,\mathcal{A}} = \sum_{i \in s_\alpha} y_{i,\mathcal{A}}/\pi_i \tag{D.1}$$

is a consistent and unbiased estimator for $y_{F_\alpha,\mathcal{A}}$.

Proof: If we define $a_i = \sum_j y_{i,A_j}$, the sum of the responses to each A_j for individual i, then we can write our estimator as

$$\widehat{y}_{F_\alpha,\mathcal{A}} = \sum_{i \in s_\alpha} a_i/\pi_i. \tag{D.2}$$

This is a Horvitz-Thompson esimator (see, e.g., Sarndal et al., 2003, chap. 2); it is unbiased and consistent for the total $\sum_{i \in F_\alpha} a_i = y_{F_\alpha,\mathcal{A}}$. ■

We can now use Property D.1.1 in Result D.1.2, the adapted known population estimator.

Result D.1.2 *Suppose we have a probability sample s taken from the frame population with known probabilities of inclusion π_i. Further, suppose we have a multiset of probe alters (also called known populations) $\mathcal{A}$ that have been chosen so that two conditions hold:*

- $y_{F_\alpha,\mathcal{A}} = d_{F_\alpha,\mathcal{A}}$ *(reporting condition)*
- $\bar{d}_{\mathcal{A},F_\alpha} = \bar{d}_{F,F_\alpha}$ *(probe alter condition).*

Then the adapted known population estimator

$$\widehat{\bar{d}}_{F_\alpha,F} = \frac{\sum_{i \in s_\alpha} y_{i,\mathcal{A}}/\pi_i}{\sum_j N_{A_j}} \frac{N_F}{N_{F_\alpha}} = \frac{\widehat{y}_{F_\alpha,\mathcal{A}}}{N_{\mathcal{A}}} \frac{N_F}{N_{F_\alpha}} \tag{D.3}$$

is consistent and unbiased for $\bar{d}_{F_\alpha,F}$.

Proof: By Property D.1.1, $\widehat{y}_{F_\alpha,\mathcal{A}}/N_{\mathcal{A}}$ is consistent and unbiased for $y_{F_\alpha,\mathcal{A}}/N_{\mathcal{A}}$. By the reporting condition, $y_{F_\alpha,\mathcal{A}}/N_{\mathcal{A}} = d_{F_\alpha,\mathcal{A}}/N_{\mathcal{A}}$. Re-writing this quantity, we have

$$\frac{d_{F_\alpha,\mathcal{A}}}{N_{\mathcal{A}}} = \frac{d_{\mathcal{A},F_\alpha}}{N_{\mathcal{A}}} = \bar{d}_{\mathcal{A},F_\alpha}. \tag{D.4}$$

Now, using the probe alter condition,

$$\bar{d}_{\mathcal{A},F_\alpha} = \bar{d}_{F,F_\alpha}. \tag{D.5}$$

So we have shown that, assuming the reporting condition and the probe alter condition hold, $\widehat{y}_{F_\alpha,\mathcal{A}}/N_{\mathcal{A}}$ is consistent and unbiased for $\bar{d}_{F,F_\alpha}$. Now we can re-write $\bar{d}_{F,F_\alpha}$ as

$$\bar{d}_{F,F_\alpha} = \frac{d_{F,F_\alpha}}{N_F} = \frac{d_{F_\alpha,F}}{N_F}. \tag{D.6}$$

So we conclude that the estimator is consistent and unbiased for

$$\frac{d_{F_\alpha,F}}{N_F}\ \frac{N_F}{N_{F_\alpha}} = \frac{d_{F_\alpha,F}}{N_{F_\alpha}} = \bar{d}_{F_\alpha,F}. \tag{D.7}$$

■

Appendix B.2.4 offers suggestions for how to choose probe alters for the known population estimator; these suggestions carry over to the adapted estimator (Result D.1.2) with some modifications to accommodate the specific reporting condition and probe alter condition required by the adapted known population estimator.

D.2 The network survival estimator

In this appendix, we provide additional information about the network survival estimator. The main idea is to form an estimate for the death rate M_α using a ratio approach: we divide an estimate for the number of deaths, O_α, by an estimate of

the amount of exposure, E_α (Section 5.3). We begin this appendix by presenting and discussing several technical results that establish precisely which conditions must hold in order for each of the estimator for M_α's components to work well. Then we present a combined estimator for M_α based on the estimators for E_α and O_α. Finally, we also show that in the special case where the number of frame population members in group α (N_{F_α}) can be used as an estimate of N_α, then a simplified estimator for death rates is possible.

Several of the results in this appendix closely follow the analysis of the generalized scale-up estimator found in Chapter 3.

D.2.1 Estimating the amount of exposure, E_α

Result D.2.1 shows that, with a probability sample, we can estimate the size of the frame population in a demographic group α (N_{F_α}) using the information in the sampling design. In the Rwanda results, we make the assumption that the frame population is *complete*, meaning that all of the adults in Rwanda could potentially have been interviewed. This assumption permits us to use Result D.2.6 to estimate death rates. Researchers who use a different estimator for $\widehat{E}_\alpha$ can produce estimates using Result D.2.5. See Appendix D.2.3 more more information.

Result D.2.1 *Suppose we have a probability sample s taken from the frame population with known probabilities of inclusion π_i. Let the subset of the sample respondents who are in a particular demographic group α be $s_\alpha \subset s$. Then*

$$\widehat{N}_{F_\alpha} = \sum_{i \in s_\alpha} 1/\pi_i \tag{D.8}$$

is consistent and essentially unbiased for N_{F_α}, the number of people in the frame population in demographic group α.

Proof: For consistency, note that if the entire frame is sampled in a census, then $\pi_i = 1$ for all $i \in F$; the estimator is then $\sum_{i \in F_\alpha} 1 = N_{F_\alpha}$. For unbiasedness, let I_i be the random variable that denotes whether or not i is included each time a sample is drawn; then $\mathbb{E}[I_i] = \pi_i$, by the definition of π_i (see, e.g. Sarndal et al., 2003, chap. 2.6). In a particular sample, the estimator will have the value $\widehat{N}_{F_\alpha} = \sum_{i \in F} I_i/\pi_i$. Over all samples, the expected value of $\widehat{N}_{F_\alpha}$ is then

$$\mathbb{E}[\widehat{N}_{F_\alpha}] = \sum_{i \in s_\alpha} \mathbb{E}[I_i]/\pi_i = \sum_{i \in s_\alpha} 1 = N_{F_\alpha}. \tag{D.9}$$

So the estimator is unbiased. ■

D.2.2 Estimating the number of deaths, O_α

Equation 5.4 shows that the two components of the estimated number of deaths are: (i) the total number of reports about deaths, y_{F,O_α}; and (ii) the average visibility of deaths, $\bar{v}_{O_\alpha,F}$. Estimators for each of these two components can be combined to estimate O_α using a standard ratio estimator. We formally show that this is the case in Result D.2.4, below. However, we must first present two other results before we arrive at Result D.2.4: one about y_{F,O_α} (Result D.2.2), and one about $\bar{v}_{O_\alpha,F}$ (Result D.2.3).

First, Result D.2.2, shows that estimating y_{F,O_α} from survey reports is a straightforward application of standard survey techniques.

Result D.2.2 *Suppose we have a probability sample s taken from the frame population with known probabilities of inclusion π_i. Then*

$$\widehat{y}_{F,O_\alpha} = \sum_{i \in s} y_{i,O_\alpha}/\pi_i \tag{D.10}$$

is consistent and unbiased for y_{F,O_α}.

Proof: Equation D.10 is a standard Horvitz-Thompson estimator (see, eg Sarndal et al., 2003, chap. 2), so it is consistent and unbiased for the total $\sum_{i \in F} y_{i,O_\alpha} = y_{F,O_\alpha}$. ■

The next result, Result D.2.3, is about estimating $\bar{v}_{F,O_\alpha}$. The result shows that it is possible to use information about the survey respondents' personal networks to estimate the visibility of deaths if two additional conditions are satisfied: the visible deaths condition and the decedent network condition.

The *visible deaths condition* says that the average number of times a death could be reported (the average visibility of deaths), is the same as the average number of network connections people who died have to the frame population. Substantively, we would expect this condition to hold when everyone who is connected to a person who died is aware of that fact and reports it on a survey. Mathematically, the visible deaths condition can be written $\bar{v}_{O_\alpha,F} = \bar{d}_{O_\alpha,F}$.

The *decedent network condition* says that the average size of personal networks is the same for dead people and for the people who respond to the survey. For example, we assume that if women aged 50-54 who are eligible to be sampled by our survey have an average personal network size of 100, then the defaults network condition is satisfied if women aged 50-54 who died also have an average personal network size of 100. Mathematically, the decedent network condition can be written $\bar{d}_{O_\alpha,F} = \bar{d}_{F_\alpha,F}$, where F_α denotes the subset of the frame population that is in cell α.

Result D.2.3 *Suppose that $\widehat{\bar{d}}_{F_\alpha,F}$ is a consistent and unbiased estimator for $\bar{d}_{F_\alpha,F}$ (such as the one in Result D.1.2). Furthermore, suppose that the following conditions hold:*

- $\bar{v}_{O_\alpha,F} = \bar{d}_{O_\alpha,F}$ *(visible deaths condition)*
- $\bar{d}_{O_\alpha,F} = \bar{d}_{F_\alpha,F}$ *(decedent network condition)*

Then $\widehat{\bar{d}}_{F_\alpha,F}$ is a consistent and unbiased estimator for $\bar{v}_{O_\alpha,F}$.

Proof: By assumption, $\widehat{\bar{d}}_{F_\alpha,F}$ is consistent and unbiased for $\bar{d}_{F_\alpha,F}$. By the decedent network condition, $\bar{d}_{F_\alpha,F} = \bar{d}_{O_\alpha,F}$. And, by the visible deaths condition, $\bar{d}_{O_\alpha,F} = \bar{v}_{O_\alpha,F}$. ■

The visible networks assumption and the decedent network assumption could both be violated in practice. Therefore, Appendix D.3 proposes a framework that can help researchers to understand the impact that violations of these two assumptions will have on the accuracy of estimated death rates.

The next result (Result D.2.4) shows how the network survival method estimates the number of deaths (O_α). Note that, in practice, it turns out not to be necessary to explicitly compute this estimated number of deaths if the frame population is complete, so that $N_{F_\alpha} = N_\alpha$ (see Result D.2.6). However, it is helpful to produce estimates for the number of deaths so that researchers can construct figures like Figure 5.3, which illustrate the components of death rate estimates. Furthermore, in some situations, researchers may be interested in the number of deaths as well as the death rate.

Result D.2.4 *Suppose $\widehat{y}_{F,O_\alpha}$ is a consistent and unbiased estimator for y_{F,O_α}, and that $\widehat{\bar{v}}_{O_\alpha,F}$ is a consistent and unbiased estimator for $\bar{v}_{O_\alpha,F}$. Suppose also that there are no false positive reports, so that $v_{i,F} = 0$ for all $i \notin O_\alpha$. Then*

$$\widehat{O}_\alpha = \frac{\widehat{y}_{F,O_\alpha}}{\widehat{\bar{v}}_{O_\alpha,F}} \tag{D.11}$$

is consistent and essentially unbiased for O_α.

Proof: With consistent and unbiased estimators for y_{F,O_α} and for $\bar{v}_{O_\alpha,F}$, we can form a consistent and essentially unbiased estimator for $y_{F,O_\alpha}/\bar{v}_{O_\alpha,F}$ using a standard ratio approach (Sarndal et al., 2003, chap. 5). So it remains to show that $y_{F,O_\alpha}/\bar{v}_{O_\alpha,F} = O_\alpha$. Since in-reports must equal out-reports (see Feehan and Salganik (2015) and Chapter 2), $y_{F,O_\alpha} = v_{U,F}$, where U is the set of all of the people who could be

reported about, living or dead (note that $O_\alpha \subset U$ and $F \subset U$). By the no false positives assumption, $v_{i,F} = 0$ for all $i \notin O_\alpha$, which means that

$$v_{U,F} = \sum_{i \in U} v_{i,F} = \sum_{i \in O_\alpha} v_{i,F} = v_{O_\alpha,F}. \tag{D.12}$$

So we conclude that $y_{F,O_\alpha} = v_{O_\alpha,F}$. Dividing both sides of this identity by O_α and re-arranging produces

$$O_\alpha = \frac{y_{F,O_\alpha}}{v_{O_\alpha,F}/O_\alpha} = \frac{y_{F,O_\alpha}}{\bar{v}_{O_\alpha,F}}. \tag{D.13}$$

■

D.2.3 Combined estimator for M_α

We now bring together the previous results to discuss estimating the death rate M_α. The first result (Result D.2.5) shows that, if we have consistent and unbiased estimators for the components of O_α (ie, y_{F,O_α} and $\bar{v}_{O_\alpha,F}$), and if we have a consistent and unbiased estimator for the population size N_α, then we can form a consistent and essentially unbiased estimator for the death rate O_α/N_α. The next result (Result D.2.6) shows that in the special case where the decedent network and visible deaths conditions hold, and if the frame population in α is complete (ie, $N_{F_\alpha} = N_\alpha$), then a simplified estimator for M_α can be used.

Note that, as is typical in demographic research, we approximate the exposure, E_α, with the size of the population; our approach is thus to form an estimator for O_α/N_α, where N_α is the size of the population in demographic group α. This approximation should not be problematic unless (i) the time period over which death rates are computed is long; or (ii) death rates are extremely high (much higher than populations typically experience). For the 12-month death rates we study in Rwanda, we do not expect this approximation to pose a problem. We will occasionally refer to O_α/N_α

as a death rate, leaving the fact that N_α is an approximation of the exposure E_α implicit.

The first result (Result D.2.5) shows that we can put together several of the estimators we have discussed so far to form a consistent and essentially unbiased estimator of M_α.

Result D.2.5 *Suppose we have a probability sample s taken from the frame population with known probabilities of inclusion π_i. Suppose also that we have a consistent and unbiased estimator $\widehat{y}_{F,O_\alpha}$ (eg, Result D.2.2); a consistent and unbiased estimator $\widehat{\bar{v}}_{O_\alpha,F}$ (eg, Result D.2.3); and a consistent and unbiased estimator $\widehat{N}_\alpha$. Then*

$$\widehat{M}_\alpha = \frac{\widehat{y}_{F,O_\alpha}}{\widehat{\bar{v}}_{O_\alpha,F}} \frac{1}{\widehat{N}_\alpha} \tag{D.14}$$

is consistent and essentially unbiased for $M_\alpha = O_\alpha/N_\alpha$.

Proof: Equation D.14 is a compound ratio estimator; Rao and Pereira (1968) and Feehan and Salganik (2015, Appendix E) give proofs that compound ratio estimators are consistent and essentially unbiased. ■

Simplified estimator for M_α when $N_{F_\alpha} = N_\alpha$

The network survival estimator in Result D.14 can be simplified when the frame population F is *complete*, meaning that $N_{F_\alpha} = N_\alpha$. The estimator in Result D.2.5 is

$$\widehat{M}_\alpha = \frac{\widehat{y}_{F,O_\alpha}}{\widehat{\bar{v}}_{O_\alpha,F}} \frac{1}{\widehat{N}_\alpha}. \tag{D.15}$$

Substituting $\widehat{\bar{d}}_{F_\alpha,F}$ for $\widehat{\bar{v}}_{O_\alpha,F}$, we have

$$\widehat{M}_\alpha = \frac{\widehat{y}_{F,O_\alpha}}{\widehat{\bar{d}}_{F_\alpha,F}} \frac{1}{\widehat{N}_\alpha}. \tag{D.16}$$

Now, substituting the adapted known population estimator for $\bar{d}_{F_\alpha,F}$ (Appendix D.1) into the expression produces

$$\frac{\widehat{y}_{F,O_\alpha}}{\widehat{\bar{d}}_{F_\alpha,F}} \frac{1}{\widehat{N}_\alpha} = \frac{\widehat{y}_{F,O_\alpha}}{\widehat{y}_{F_\alpha,\mathcal{A}}} \frac{N_\mathcal{A}}{N_F} \frac{N_{F_\alpha}}{\widehat{N}_\alpha}. \tag{D.17}$$

Now we focus on the constant factor $N_{F_\alpha}/\widehat{N}_\alpha$ in Equation D.17. It is important to distinguish between the frame population in demographic group α (a group whose size is N_α) and the entire population in demographic group α (a group whose size is N_{F_α}); in general, these two quantities will not be the same. However, if the frame population is *complete* in demographic group α, meaning that $N_{F_\alpha} = N_\alpha$, then the denominator of the constant factor $N_{F_\alpha}/\widehat{N}_\alpha$ in Equation D.17 is estimating N_{F_α}. We can therefore plug in N_{F_α} for $\widehat{N}_\alpha$, leading the constant factors to cancel out. Then,

$$\widehat{M}_\alpha = \frac{\widehat{y}_{F,O_\alpha}}{\widehat{y}_{F_\alpha,\mathcal{A}}} \frac{N_\mathcal{A}}{N_F} \quad \text{when } N_\alpha = N_{F_\alpha}. \tag{D.18}$$

Equation D.19 is the estimator we use to produce the $\widehat{M}_\alpha$ in Rwanda. This argument is formalized in Result D.2.6.

Result D.2.6 *Suppose we have a probability sample s taken from the frame population with known probabilities of inclusion π_i. Suppose that we have a set of probe alters $\mathcal{A}$ (also called known populations) that satisfy the reporting condition ($y_{F_\alpha,\mathcal{A}} = d_{F_\alpha,\mathcal{A}}$) and the probe alter condition ($\bar{d}_{\mathcal{A},F_\alpha} = \bar{d}_{F,F_\alpha}$) from Result D.1.2. Suppose that the visible deaths condition ($\bar{v}_{O_\alpha,F} = \bar{d}_{O_\alpha,F}$) and the decedent network condition ($\bar{d}_{O_\alpha,F} = \bar{d}_{F_\alpha,F}$) from Result D.2.3 are satisfied. Finally, suppose that the frame population is complete, so that $N_{F_\alpha} = N_\alpha$, and that there are no false positive reports about deaths, so that $v_{i,F} = 0$ for all $i \notin O_\alpha$. Then*

$$\widehat{M}_\alpha = \frac{\sum_{i\in s} y_{i,O_\alpha}/\pi_i}{\sum_{i\in s_\alpha} y_{i,\mathcal{A}}} \frac{N_\mathcal{A}}{N_F} = \frac{\widehat{y}_{F,O_\alpha}}{\widehat{y}_{F_\alpha,\mathcal{A}}} \frac{N_\mathcal{A}}{N_F} = \frac{\widehat{\bar{y}}_{F,O_\alpha}}{\widehat{\bar{d}}_{F_\alpha,F}} \tag{D.19}$$

is consistent and essentially unbiased for $M_\alpha = O_\alpha / N_\alpha$.

Proof: Equation D.19 is a standard ratio estimator, so it is consistent and essentially unbiased for the quantity

$$Q_\alpha = \frac{y_{F,O_\alpha}}{y_{F_\alpha,\mathcal{A}}} \frac{N_\mathcal{A}}{N_F} \tag{D.20}$$

(see, e.g. Sarndal et al., 2003, chap. 5). So it remains to show that $Q_\alpha = O_\alpha / N_\alpha = M_\alpha$. We will do this by working backwards through the discussion above. First, multiply Q_α by N_{F_α} / N_α (which equals 1, by the completeness of the frame population), to obtain

$$Q_\alpha = \frac{y_{F,O_\alpha}}{y_{F_\alpha,\mathcal{A}}} \frac{N_\mathcal{A}}{N_F} \frac{N_{F_\alpha}}{N_\alpha}. \tag{D.21}$$

Now we can use the reporting condition ($y_{F_\alpha,\mathcal{A}} = d_{F_\alpha,\mathcal{A}}$) followed by the probe alter condition ($\bar{d}_{\mathcal{A},F_\alpha} = \bar{d}_{F,F_\alpha}$) to rewrite the expression as

$$Q_\alpha = \frac{y_{F,O_\alpha}}{\bar{d}_{F,F_\alpha}} \frac{1}{N_F} \frac{N_{F_\alpha}}{N_\alpha}. \tag{D.22}$$

Now, recall that $\bar{d}_{F,F_\alpha} \; N_F / N_{F_\alpha} = \bar{d}_{F_\alpha,F}$. Applying this relationship to simplify the denominator of Equation D.22 produces

$$Q_\alpha = \frac{y_{F,O_\alpha}}{\bar{d}_{F_\alpha,F}} \frac{1}{N_\alpha}. \tag{D.23}$$

Finally, applying the decedent network condition ($\bar{d}_{F_\alpha,F} = \bar{d}_{O_\alpha,F}$) and the visible deaths condition ($\bar{d}_{O_\alpha,F} = \bar{v}_{O_\alpha,F}$), we have

$$Q_\alpha = \frac{y_{F,O_\alpha}}{\bar{v}_{O_\alpha,F}} \frac{1}{N_\alpha}. \tag{D.24}$$

Now, since there are no false positive reports, we can apply the argument in Result D.2.4 to conclude that $y_{F,O_\alpha}/\bar{v}_{O_\alpha,F} = O_\alpha$. Therefore,

$$Q_\alpha = \frac{O_\alpha}{N_\alpha} = M_\alpha. \tag{D.25}$$

■

D.3 Robustness framework

In this section, we develop a framework to help researchers understand how death rates estimated using network survival can be affected by two potential sources of inaccuracy: (i) inaccuracies in respondents' reports about their personal network members who have died; and, (ii) violations of the conditions needed for survey respondents' personal networks to provide good estimates of the visibility of deaths. This framework is useful for several reasons. First, the framework identifies precise and well-defined quantities that future studies may be able to measure; with measurements for the factors in our decomposition, network survival estimates could be adjusted and potentially improved. Second, the framework permits researchers to conduct robustness analyses which can reveal how their death rate estimates are affected by different types of reporting error and violations of the visible deaths and decedent network conditions. This framework is an adapted version of the one introduced for the scale-up estimator in Feehan and Salganik (2015).

D.3.1 Inaccuracies in reports about deaths, y_{F,O_α}

We begin by analyzing the impact that inaccurate reports will have on death rate estimates. To do so, we will revisit the network reporting framework. Figure 5.1(b) shows an example of a reporting network that has been rearranged into what we call a *bipartite reporting graph*. The edges in this bipartite reporting graph contribute

two types of quantities to the vertices in the graph: each edge adds an *out-report* to the people who report, on the left-hand side of the graph; and each edge adds an *in-report* to the people who get reported about, on the right-hand side of the graph. We call the sum of all of the out-reports y_{F,O_α}, and the sum of all of the in-reports $v_{U,F}$, where U is the set of all people who could be reported about (the right-hand side of the bipartite reporting graph). This example highlights an important identity which is true in *any* bipartite reporting graph, no matter how accurate or inaccurate respondents' reports are: the total number of out-reports must equal the total number of in-reports ($y_{F,O_\alpha} = v_{U,F}$). (This is equivalent to the trivial fact that, in any directed graph, the sum of the out-degrees must equal the sum of the in-degrees.)

We can separate the out-reports into two groups: (i), the *true positives*, which are reports that are correctly lead to people who died; and (ii) the *false positives*, which are reports that incorrectly lead to people who did not die. We write the true positives as y^+_{F,O_α}, and the false positives as y^-_{F,O_α}. By definition, all of the true positive reports lead to O_α, meaning that $y^+_{F,O_\alpha} = v_{O_\alpha,F}$. This identity is true in any bipartite reporting graph, no matter how accurate or inaccurate respondents' reports are. Starting from $y^+_{F,O_\alpha} = v_{O_\alpha,F}$, multiplying both sides by O_α, and then rearranging the terms yields

$$O_\alpha = \frac{y^+_{F,O_\alpha}}{\bar{v}_{O_\alpha,F}}. \tag{D.26}$$

Equation D.26 relates the number of deaths (O_α) to the true positive reports (y^+_{F,O_α}). In the main text, we derived a different identity that based on the assumption that there were no false positive reports. Now we can see that this no false positives identity in Equation 5.4 is a special case of the general identity in Equation D.26: when there are no false positives, $y_{F,O_\alpha} = y^+_{F,O_\alpha}$; substituting this into Equation D.26 yields the identity in the main text (Equation 5.4).

D.3.2 Decomposition framework

The results above show that the network survival estimator can be used to estimate death rates when $y_{F,O_\alpha} = y^+_{F,O_\alpha}$, and $\bar{v}_{O_\alpha,F} = \bar{d}_{F_\alpha,F}$. When those conditions are not met, the network survival estimates will be inconsistent and biased. Additionally, the presence of inaccuracies in reporting (the false positives and false negatives described above) can also lead to inconsistent or biased estimates. Therefore, following the approach of Feehan and Salganik (2015), we propose a decomposition of the difference between network survival estimates for O_α and the true value of O_α. Our decomposition separates the different sources of inaccuracy into separate, precisely defined quantities; it is:

$$O_\alpha = \underbrace{\left(\frac{y_{F,O_\alpha}}{\bar{d}_{F_\alpha,F}}\right)}_{\substack{\text{network}\\ \text{survival}}} \times \underbrace{\underbrace{\frac{1}{\bar{d}_{O_\alpha,F}/\bar{d}_{F_\alpha,F}}}_{\substack{\text{degree ratio}\\ \delta_{F,\alpha}}} \times \underbrace{\frac{1}{\bar{v}_{O_\alpha,F}/\bar{d}_{O_\alpha,F}}}_{\substack{\text{true positive rate}\\ \tau_{F,\alpha}}} \times \underbrace{\frac{y^+_{F,O_\alpha}}{y_{F,O_\alpha}}}_{\substack{\text{precision}\\ \eta_{F,\alpha}}}}_{\text{adjustment factors}} \tag{D.27}$$

The decomposition in Equation D.27 is based on three adjustment factors: $\delta_{F,\alpha}$, $\tau_{F,\alpha}$, and $\eta_{F,\alpha}$. The decomposition in Equation D.27, shows that the network survival estimates will be consistent (ie, exactly correct in a census) if the the three adjustment factors satisfy $\eta_{F,\alpha}/(\delta_{F,\alpha} \times \tau_{F,\alpha}) = 1$. In all other cases, the network survival estimator will be inconsistent.

We now describe each of the three adjustment factors in turn. The first adjustment factor is the *degree ratio*:

$$\delta_{F,\alpha} = \frac{\text{avg \# edges from a death in } \alpha \text{ to the frame population}}{\text{avg \# edges from a frame pop member in } \alpha \text{ to the entire frame pop}} = \frac{\bar{d}_{O_\alpha,F}}{\bar{d}_{F_\alpha,F}}. \tag{D.28}$$

$\delta_{F,\alpha}$ will range from 0 to infinity. When it is less than one, people who die in demographic group α tend to have fewer connections to the frame population than frame population members in demographic group α; when it is greater than one, people who die in demographic group α tend to have more connections to the frame population than frame population members in demographic group α. Values of $\delta_{F,\alpha}$ other than 1 mean that the decedent network condition is violated.

The second adjustment factor is the *true positive rate*:

$$\tau_{F,\alpha} = \frac{\text{avg \# of in-reports from the frame to each death}}{\text{avg \# of network connections from a death to the frame population}} = \frac{\bar{v}_{O_\alpha,F}}{\bar{d}_{O_\alpha,F}}. \tag{D.29}$$

$\tau_{F,\alpha}$ relates network degree to network reports; it will range from 1, when reporting is perfectly accurate, to 0, when no network edges leading to deaths are reported. Values of $\tau_{F,\alpha}$ other than 1 mean that the visible deaths condition is violated.

The third adjustment factor is the *precision*:

$$\eta_{F,\alpha} = \frac{\text{total \# of out-reports from frame popn that correctly lead to deaths}}{\text{total \# of out-reports from frame popn}} = \frac{y^{+}_{F,O_\alpha}}{y_{F,O_\alpha}}. \tag{D.30}$$

$\eta_{F,\alpha}$ relates accurate network reports to all network reports; it will range from 1, when reporting is perfectly accurate, to 0, when none of the out-reports correctly leads to a death. Values of $\eta_{F,\alpha}$ other than 1 mean that the no false positives assumption is violated.

Appendix D.4 applies the framework in Equation D.27 to the case of the death rate estimates from Rwanda.

D.4 Results for both sexes and tie definitions

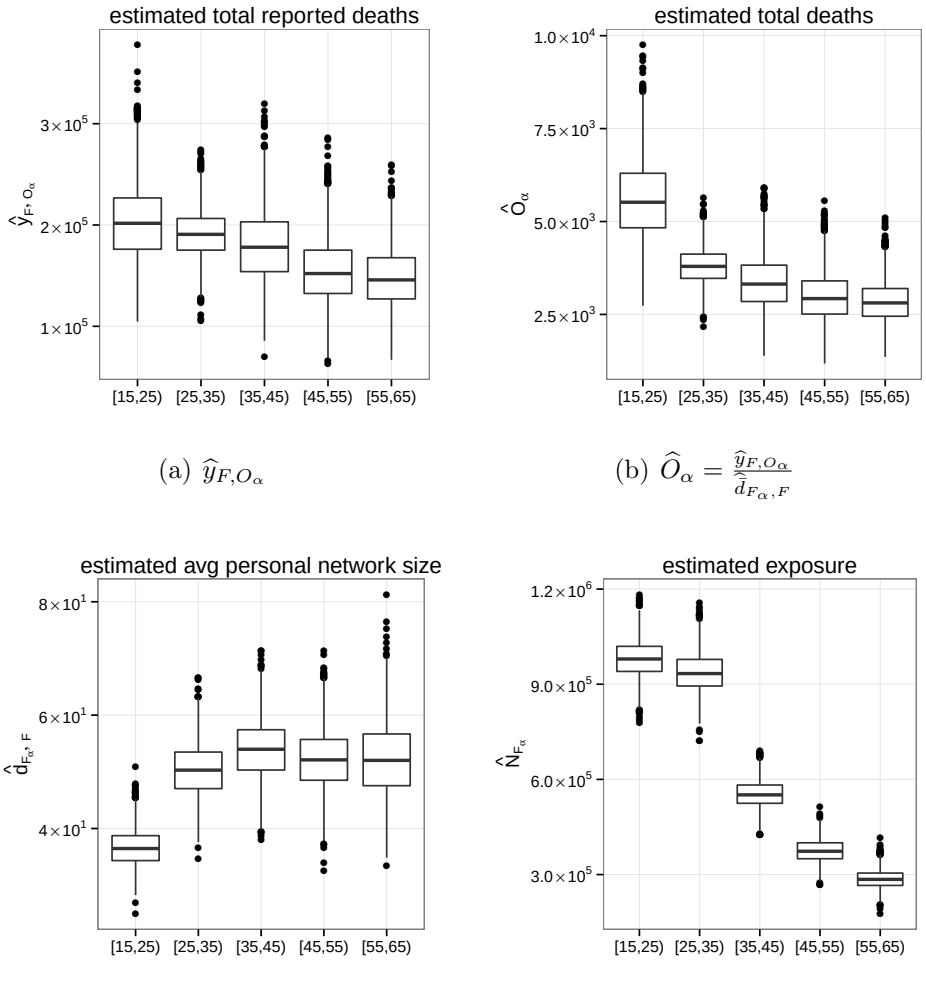
In the main text, Figure 5.3(a) illustrates all of the separate steps that go into constructing age-specific death rate estimates for Rwandan males using survey reports about the meal tie definition. This appendix has figures that illustrate the detailed steps for the other network survival death rates we constructed in Rwanda: female death rates from the meal network (Figures D.1(a)); male death rates from the acquaintance network (Figure D.2(a)); and female death rates from the acquaintance network (Figure D.3(a)).

D.5 Sibling survival estimates

This appendix focuses on the sibling survival method. First, in Section D.5.1, we extend the discussion in Section 5.2.2 by providing a more detailed review of the direct sibling survival method, focusing on recent methodological debates. Second, in Section D.5.2, we provide a detailed description of how we computed estimated age-specific death rates from the 2010 Rwanda DHS sibling histories.

D.5.1 Background

Researchers have studied and improved several sibling survival methods over many years. The initial idea, inspired by research on indirect estimation techniques pioneered by William Brass and Ansley Coale, was to ask a survey or census respondent how many siblings she had, and how many were still alive. Hill and Trussell (1977) first proposed using this information about the proportion of siblings that had died to form an indirect estimator of adult mortality, and this approach was later adapted by Graham et al. (1989) to create an indirect estimator for maternal mortality. Both of these early, indirect approaches were based on formal demographic relationships and some analytical approximations. These indirect estimators had several limitations:



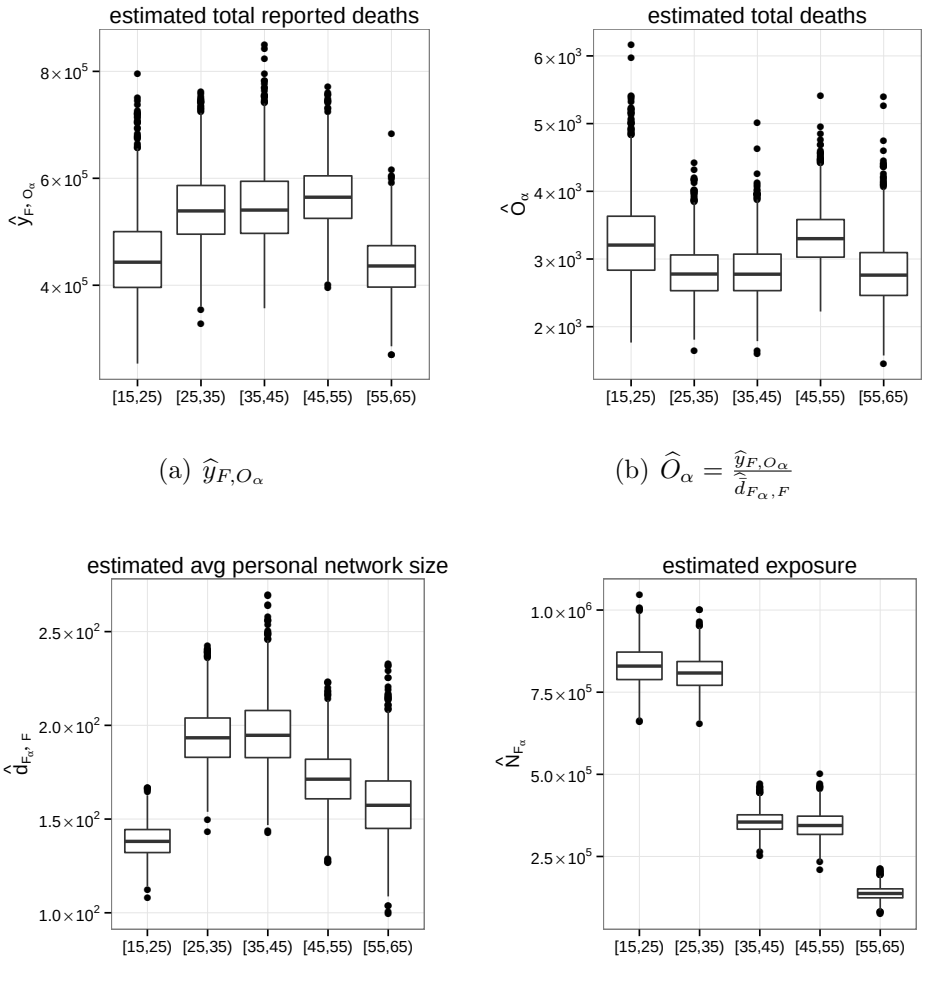
(a) $\widehat{y}_{F,O_\alpha}$

(b) $\widehat{O}_\alpha = \frac{\widehat{y}_{F,O_\alpha}}{\widehat{\bar{d}}_{F_\alpha,F}}$

(c) $\widehat{\bar{d}}_{F_\alpha,F}$ (an estimate of $\bar{v}_{O_\alpha,F}$)

(d) $\widehat{N}_{F_\alpha}$ (an estimate of E_α)

Figure D.1: Estimating components of age-specific death rates for Rwandan females for 12 months prior to our survey using responses from the meal tie definition. The interpretation of this figure is analogous to Figure 5.3. The death rates estimated with these components are shown in the middle column of the bottom row in Figure 5.6.



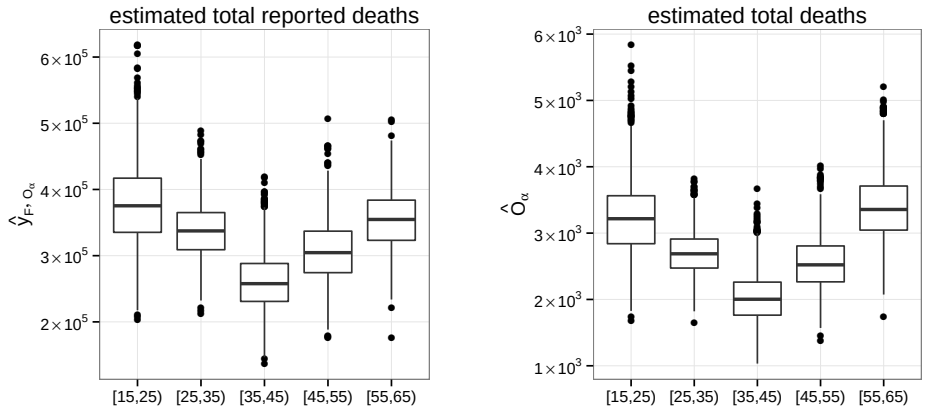
(a) $\widehat{y}_{F,O_\alpha}$

(b) $\widehat{O}_\alpha = \frac{\widehat{y}_{F,O_\alpha}}{\widehat{\bar{d}}_{F_\alpha,F}}$

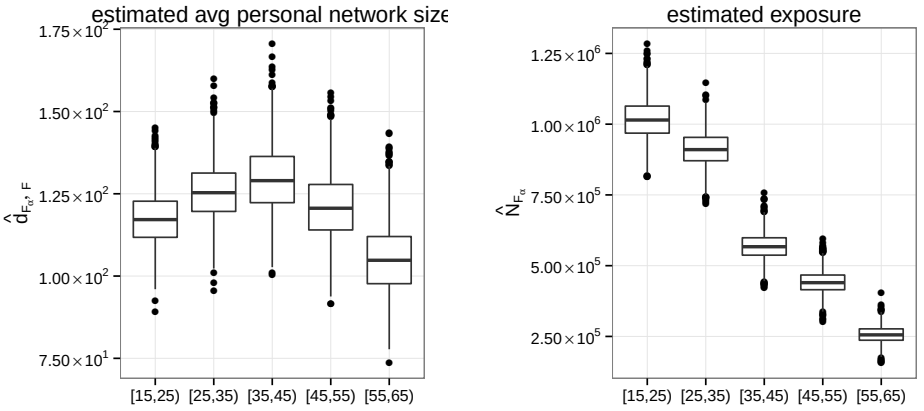
(c) $\widehat{\bar{d}}_{F_\alpha,F}$ (an estimate of $\bar{v}_{O_\alpha,F}$)

(d) $\widehat{N}_{F_\alpha}$ (an estimate of E_α)

Figure D.2: Estimating components of age-specific death rates for Rwandan males for 12 months prior to our survey using responses from the acquaintance tie definition. The interpretation of this figure is analogous to Figure 5.3. The death rates estimated with these components are shown in the left-hand column of the top row in Figure 5.6.



(a) $\widehat{y}_{F,O_\alpha}$ (b) $\widehat{O}_\alpha = \frac{\widehat{y}_{F,O_\alpha}}{\widehat{\bar{d}}_{F_\alpha,F}}$



(c) $\widehat{\bar{d}}_{F_\alpha,F}$ (an estimate of $\bar{v}_{O_\alpha,F}$) (d) $\widehat{N}_{F_\alpha}$ (an estimate of E_α)

Figure D.3: Estimating components of age-specific death rates for Rwandan females for 12 months prior to our survey using responses from the acquaintance tie definition. The interpretation of this figure is analogous to Figure 5.3. The death rates estimated with these components are shown in the left-hand column of the bottom row in Figure 5.6.

first, they required strong assumptions about population stability to be valid; second, they were sensitive to reporting problems, especially omissions of siblings who died before the respondent was born; and, third, they relied upon model schedules of mortality and fertility.

Rutenberg and Sullivan (1991) introduced the direct sibling survival method, which has become the standard approach used today; sibling histories have been collected from over 150 surveys in dozens of countries around the developing world. Section 5.2.2 argues that this vast amount of sibling history data has been relatively under-used, and that there are two reasons why this under-use may have happened: first, for sample sizes typical of the DHS, the direct sibling survival method does not produce estimates for age-, sex-, and year-specific death rates that are precise enough to be useful (Stanton et al., 2000; Timaeus and Jasseh, 2004; Hill et al., 2006). And, second, there is a considerable methodological debate about how sibling histories should be analyzed.

We now turn to a review of the recent methodological work on the analysis of sibling history data. Several studies have evaluated the quality of DHS sibling histories. Stanton et al. (2000) found that the sibling history reports were generally complete, but that adult mortality estimates from sibling histories tended to be lower than estimates from other sources. Masquelier and Dutreuilh (2014), found that that quality varied considerably across surveys and that sibship sizes are typically under-reported; however, the omissions appear to be focussed at young ages, which will not affect adult mortality rates estimated using sibling history data. Merdad et al. (2013) compared sibling history estimates of death rates for men and for women, and found no difference in quality by sex. Finally, Hill et al. (2006) conducted a very large survey of about 99,000 households in Bangladesh. By comparing estimates from direct sibling survival to estimated death rates from a large cohort study, the authors

conclude that direct sibling survival performed well in estimating adult death rates, but that sampling variation was very large.

Another group of studies specifically investigated the individual-level quality of sibling history reports, especially in the context of estimating maternal mortality. Helleringer et al. (2013) started with a set of known maternal deaths in a demographic surveillance site (DSS) in rural Senegal, and then interviewed the surviving adult sisters of the deceased to collect sibling histories from them. The authors found some evidence that sisters were omitted, and that survival status was misreported. More recently, Helleringer et al. (2014a) compared a DHS-like survey in the same Senegalese site to individual-level data tracked by the DSS. They found that under-reporting of siblings was common, and that there were additional errors in reporting ages and dates. Together, these reporting errors led to an underestimated ${}_{45}q_{15}$ (see Appendix D.6) from the direct sibling survival approach. Helleringer et al. (2014b) tests a new survey instrument for collecting sibling histories, based on the event history calendar; the authors found some evidence that the new instrument improves the quality of reports about sisters, but not about brothers.

Other studies have focused on developing models for pooling, smoothing, and otherwise adjusting sibling history data. Timaeus and Jasseh (2004) was the first study to pool several surveys together to try and estimate trends in adult mortality from several countries in Sub-Saharan Africa. The authors found that sibling reports about years long before the interview were implausible, so they focused only on information about the 7 years prior to the survey. In order to circumvent the sampling noise inherent in direct estimates, the authors proposed a Poisson regression framework, combined with predictors from model life tables and special adjustments for countries known to have high HIV prevalence. Gakidou and King (2006) proposed a set of weights that adjust for the selection bias that is introduced by the fact that sibships with more survivors are more likely to be sampled, an issue first addressed in the con-

text of the indirect sibling survival estimator (Trussell and Rodriguez, 1990; Graham et al., 1989). The authors also propose a model-based extrapolation to account for siblings that have no survivors left to be sampled at the time of the survey. These selection bias corrections were then used in Obermeyer et al. (2010), which pooled many surveys together using a logistic regression framework and further introduced a clever approach to identifying recall bias by comparing estimates for the same age-sex-time period produced by two different surveys. Masquelier (2013) used microsimulations to argue that the weights proposed by Gakidou and King (2006) make little difference in practice, and further argued that the corrections used by Obermeyer et al. (2010) had not adequately accounted for the sampling design. Moreover, Masquelier (2013) concluded that the model-based adjustment for sibships with no survivors was very sensitive to the specification of the extrapolation model. Subsequent studies have therefore been divided: one group has applied the Gakidou-King selection bias adjustments (Kassebaum et al., 2014; Wang et al., 2013; Rajaratnam et al., 2010) while another has not (Reniers et al., 2011; Moultrie et al., 2013; Masquelier et al., 2014).

To conclude, the direct sibling survival method has been the subject of considerable methodological debate. Unfortunately, no consensus has yet emerged about how to address the potential structural and reporting problems that may affect sibling history data, and this methodological uncertainty likely contributes to some researchers' skepticism about mortality estimates produced from reports about siblings (Reniers et al., 2011).

D.5.2 Direct sibling survival estimates

In this section, we describe how we computed estimated adult death rates from the sibling histories in the 2010 Rwanda DHS. NISR et al. (2012) contains detailed infor-

mation about the survey, and all of the data are freely available online through the DHS website[1].

Sections 5.2 and Appendix D.5.1 describe the considerable methodological debate over how to produce estimated death rates from DHS sibling histories. Our goal here was to construct the simplest direct sibling survival estimates possible; although we believe that contributing to the sibling survival debate is a valuable direction for future work, that is not our intent here. We therefore follow the recommendation of the offical *Guide to DHS Statistics* (Rutstein and Rojas, 2006) and the International Union for the Scientific Study of Population's *Tools for Demographic Estimation* (Moultrie et al., 2013) by using the original direct sibling survival estimator proposed by Rutenberg and Sullivan (1991). The estimator can be written

$$\widehat{M}_\alpha = \frac{\sum_{i \in s} \frac{1}{\pi_i} \sum_{k \in \sigma(i)} O_{k,\alpha}}{\sum_{i \in s} \frac{1}{\pi_i} \sum_{k \in \sigma(i)} E_{k,\alpha}}, \tag{D.31}$$

where $\widehat{M}_\alpha$ is the estimated death rate in demographic group α; s is the sample of survey respondents, π_i is respondent i's probability of inclusion from the sampling design; $\sigma(i)$ is the set of siblings that respondent i reports about; $O_{k,\alpha}$ is an indicator variable for whether or not k died when in demographic group α, and $E_{k,\alpha}$ is the amount of time k spent alive in demographic group α. We used the networkreporting R package to compute these estimated death rates from the DHS sibling histories (Feehan and Salganik, 2014).

We wanted to compare the network survival results (based on 12 months prior to the survey) to the sibling survival estimates. Therefore, our preference would be to compute sibling survival estimates for the 12 months prior to the survey. However, the left-hand panel of Figure D.4 shows that estimates for this time frame have too much sampling variation to be practically useful (and this is consistent with the sib-

[1] `http://dhsprogram.com/what-we-do/survey/survey-display-364.cfm`

ling history literature; see Section 5.2). Since samples are not typically large enough to permit estimating yearly age-specific death rates using the estimator in Equation D.31, Rutstein and Rojas (2006) and Rutenberg and Sullivan (1991) recommend producing estimates for the 84 months (ie, 7 years) prior to the survey. However, we found that, in the special case of the 2010 Rwanda DHS, the sibling history estimates for 60 months (ie, 5 years) prior to the survey appeared to be stable enough to use for comparison (Figure D.4). Therefore, in the main text, we present sibling estimates based on 60 months (5 years) prior to the survey instead of the recommended 84 months (7 years).

D.6 ${}_{45}q_{15}$ comparisons

In this appendix, we review the definition of ${}_{45}q_{15}$, and we then describe how we constructed the survey-based estimates of ${}_{45}q_{15}$ using the age-specific death rate estimates from the network survival and from the sibling survival methods. Throughout, we only consider death rates by age, to keep the notation simple. The analysis here would work for any other demographic grouping beyond age, including age-sex, age-sex-education, and so on.

In a cohort, ${}_{45}q_{15}$ is the probability of death between the exact ages of 15 and 60, conditional on surviving to exact age 15; the cohort quantity can be written

$$ {}_{45}q_{15}^{\text{c}} = \frac{\text{\# deaths to cohort members between ages 15 and 59}}{\text{\# cohort members who survive from birth to age 15}}, \tag{D.32} $$

where we use the notation ${}_{45}q_{15}^{\text{c}}$ to denote that this is a cohort probability. When working with period data, as we do in Rwanda, we adopt the conventional approach in demography: we imagine a synthetic cohort whose members go through their lives facing the estimated cross-sectional (period) death rates; the period ${}_{45}q_{15}$ is the ${}_{45}q_{15}^{\text{c}}$ for this synthetic cohort. Period ${}_{45}q_{15}$ can therefore be interpreted as the probability

of dying before age 60 among members of a synthetic cohort who survive to age 15. It is interpretable as a probability whose value ranges from 0 to 1; the closer ${}_{45}q_{15}$ is to 1, the more adult mortality there is in the population.

In mathematical terms, ${}_{45}q_{15}$ is defined in terms of the continuous mortality hazard function, $\mu(x)$. This function $\mu(x)$ is the theoretical, continuous analogue of the age-specific death rates. In terms of $\mu(x)$, ${}_{45}q_{15}$ is defined as:

$$ {}_{45}q_{15} = 1 - \exp\left(-\int_{15}^{60} \mu(x)dx\right). \tag{D.33} $$

In practice, we never observe the continuous mortality hazard $\mu(x)$; instead, we use death rates by age groups $\widehat{M_\alpha}$. We then use the estimated age-specific death rates to approximate the continuous expression in Equation D.33:

$$ {}_{45}q_{15} \approx 1 - \sum_{\alpha \in A} \exp(-w\ M_\alpha), \tag{D.34} $$

where A is the set of all age groups α between 0 and 60, and w is the width of the age groups. Equation D.34 makes the simplifying assumption that the hazard of death is constant within each age group α; standard demography texts contain detailed explanations (Preston et al., 2001; Wachter, 2014). Many other approximations of Equation D.33 are possible; for example, we also tried approximating ${}_{45}q_{15}$ from estimated death rates by assuming that deaths were evenly distributed in the time interval. For our estimates in Rwanda, we found that the choice of approximation made no appreciable difference in the results (Figure D.5). Therefore, we use the constant hazard assumption in the results shown in the main text.

D.7 Data collection and network survival survey instrument

In this appendix, we discuss some details of the Rwanda data collection. We reproduce an excerpt of the English translation of the survey instrument that we used for the meal tie definition in Rwanda, and we comment on its design. All of the survey materials, including the original Kinyarwanda instruments for both the meal and tie definition, as well as their English translations, are freely available from the DHS website[2].

We had to pay careful attention to constructing the wording of the question that asked respondents to report about deaths. Both tie definitions used in the Rwanda study were based on interactions (Table 4.1) – either contact, for the acquaintance definition, or sharing a meal or drink, for the meal definition. Of course, people who have died cannot continue to interact with others. We therefore expect people who have died in the 12 months before a survey to have had fewer total interactions than people who did not. This expected systematic difference in the amount of interaction for people who died in the past 12 months is problematic for network survival estimates, which are based on the assumption that the visibility of deaths can be estimated by the personal network size of survey respondents (the *decedent network assumption* in Result D.2.4). Thus, we do not want the personal networks of people who died to be smaller, on average, than people who lived.

Question 226 in the sample instrument, which collects information about deaths, shows how we attempted to circumvent this potential problem in the Rwanda study. The question asks respondents to report people who satisfy two conditions: (i) the person died in the 12 months before the interview; and (ii) the person shared a meal with the respondent *in the 12 months before death*. As long as the rate of sharing

[2] `http://dhsprogram.com/what-we-do/survey/survey-display-422.cfm`

meals has not changed over the 24 months preceding the survey, this assumption should not cause a problem. However, if the rate of sharing a meal has changed over the 24 months prior to the survey, this assumption would not hold.

Note that this need to make assumptions about the rate at which meals are shared or contact is maintained is a function of the tie definition, and not the network survival estimator. If we had used a tie definition that was fixed over time – for example, ties based on a kinship relation like siblings or cousins, or ties based on mutual attendance at some fixed event – then this problem would be avoided. We therefore consider this need to assume that interaction patterns have not changed in the 24 months before the survey to be one of the trade-offs researchers will need to make when considering different tie definitions.

SECTION 2. KNOWN POPULATION

NO.	QUESTIONS AND FILTERS	CODING CATEGORIES	SKIP
200	Now I am going to ask you some questions about people that you know. These questions will help us count the number of people who may be in need of certain health services These people should be: - people you know by sight AND name, and who also know you by sight and name. In other words, you should not consider famous people that you know about, but who do not know about you. - people you have shared a meal or drink with in the past 12 months. These could be family members, friends, co-workers, or neighbors. You should include meals or drinks taken at any location, such as at home, at work, or in a restaurant. - people of all ages who live in Rwanda.		
201	How many men have you shared a meal or drink with whose wife has died and they have not remarried? IF DOES NOT KNOW ANY, RECORD '00' IF KNOWS 95 OR MORE, RECORD '95	NUMBER OF MEN WHOSE WIFE HAS DIED	
202	How many people have you shared a meal or drink with who are currently nurses or doctors? IF DOES NOT KNOW ANY, RECORD '00' IF KNOWS 95 OR MORE, RECORD '95	NUMBER OF NURSES/DOCTORS	
203	How many people have you shared a meal or drink with who are currently male community health workers in 2010? IF DOES NOT KNOW ANY, RECORD '00' IF KNOWS 95 OR MORE, RECORD '95	NUMBER OF MALE COM. HEALTH WORKERS . . .	
204	How many people have you shared a meal or drink with who are currently primary or secondary teachers? IF DOES NOT KNOW ANY, RECORD '00' IF KNOWS 95 OR MORE, RECORD '95	NUMBER OF TEACHERS	
205	How many women have you shared a meal or drink with who currently smoke a pipe or cigarettes? IF DOES NOT KNOW ANY, RECORD '00' IF KNOWS 95 OR MORE, RECORD '95	NUMBER OF WOMEN WHO SMOKE	
206	How many men have you shared a meal or drink with who are currently catholic priests? IF DOES NOT KNOW ANY, RECORD '00' IF KNOWS 95 OR MORE, RECORD '95	NUMBER OF PRIEST	
207	How many people have you shared a meal or drink with who are currently civil servants? IF DOES NOT KNOW ANY, RECORD '00' IF KNOWS 95 OR MORE, RECORD '95	NUMBER OF CIVIL SERVANTS	
208	How many women have you shared a meal or drink with who gave birth in the last 12 months? IF DOES NOT KNOW ANY, RECORD '00' IF KNOWS 95 OR MORE, RECORD '95	NUMBER OF WOMEN WHO GAVE BIRTH	
209	How many people have you shared a meal or drink with who are Muslims? IF DOES NOT KNOW ANY, RECORD '00' IF KNOWS 95 OR MORE, RECORD '95	NUMBER OF MUSLIMS	
210	How many people have you shared a meal or drink with who are currently incarcerated? IF DOES NOT KNOW ANY, RECORD '00' IF KNOWS 95 OR MORE, RECORD '95	NUMBER OF PEOPLE INCARCERATED	
211	How many people have you shared a meal or drink with who were Gacaca judges in 2010? IF DOES NOT KNOW ANY, RECORD '00' IF KNOWS 95 OR MORE, RECORD '95	NUMBER OF GACACA JUDGES	
212	How many men have you shared a meal or drink with who are divorced or separated and not remarried? IF DOES NOT KNOW ANY, RECORD '00' IF KNOWS 95 OR MORE, RECORD '95	NUMBER OF MEN DIVORCED/SEPARATED . . .	

NO.	QUESTIONS AND FILTERS	CODING CATEGORIES	SKIP
213	How many people have you shared a meal or drink with who are being treated for TB? IF DOES NOT KNOW ANY, RECORD '00' IF KNOWS 95 OR MORE, RECORD '95	NUMBER OF PEOPLE TREATED FOR TB	
	Just as a reminder I am only interested in - people you shared a meal or drink with in the past 12 months - People of all ages who live in Rwanda.		
214	How many people have you shared a meal or drink with are named NSENGIMANA? IF DOES NOT KNOW ANY, RECORD '00' IF KNOWS 95 OR MORE, RECORD '95	NUMBER OF NSENGIMANA	
215	How many people have you shared a meal or drink with are named MUREKATETE? IF DOES NOT KNOW ANY, RECORD '00' IF KNOWS 95 OR MORE, RECORD '95	NUMBER OF MUREKATETE	
216	How many people have you shared a meal or drink with are named TWAHIRWA? IF DOES NOT KNOW ANY, RECORD '00' IF KNOWS 95 OR MORE, RECORD '95	NUMBER OF TWAHIRWA	
217	How many people have you shared a meal or drink with are named MUKANDEKEZI? IF DOES NOT KNOW ANY, RECORD '00' IF KNOWS 95 OR MORE, RECORD '95	NUMBER OF MUKANDEKEZI	
218	How many people have you shared a meal or drink with are named NSABIMANA? IF DOES NOT KNOW ANY, RECORD '00' IF KNOWS 95 OR MORE, RECORD '95	NUMBER OF NSABIMANA	
219	How many people have you shared a meal or drink with are named MUKAMANA? IF DOES NOT KNOW ANY, RECORD '00' IF KNOWS 95 OR MORE, RECORD '95	NUMBER OF MUKAMANA	
220	How many people have you shared a meal or drink with are named NDAYAMBAJE? IF DOES NOT KNOW ANY, RECORD '00' IF KNOWS 95 OR MORE, RECORD '95	NUMBER OF NDAYAMBAJE	
221	How many people have you shared a meal or drink with are named NYIRANEZA? IF DOES NOT KNOW ANY, RECORD '00' IF KNOWS 95 OR MORE, RECORD '95	NUMBER OF NYIRANEZA	
222	How many people have you shared a meal or drink with are named BIZIMANA? IF DOES NOT KNOW ANY, RECORD '00' IF KNOWS 95 OR MORE, RECORD '95	NUMBER OF BIZIMANA	
223	How many people have you shared a meal or drink with are named NYIRAHABIMANA? IF DOES NOT KNOW ANY, RECORD '00' IF KNOWS 95 OR MORE, RECORD '95	NUMBER OF NYIRAHABIMANA	
224	How many people have you shared a meal or drink with are named NDAGIJIMANA? IF DOES NOT KNOW ANY, RECORD '00' IF KNOWS 95 OR MORE, RECORD '95	NUMBER OF NDAGIJIMANA	
225	How many people have you shared a meal or drink with are named MUKANDAYISENGA? IF DOES NOT KNOW ANY, RECORD '00' IF KNOWS 95 OR MORE, RECORD '95	NUMBER OF MUKANDAYISENGA	

NO.	QUESTIONS AND FILTERS	CODING CATEGORIES	SKIP
226	Now I would like to ask you a few questions about people who have died. Similar to the previous questions only tell me about - people you shared a meal or drink with in the past 12 months before they died. - These should be people of all ages living in Rwanda. How many people have you shared a meal or drink with who have died in the past 12 months?	NUMBER OF DEATHS . . . . . ☐☐ NONE . . . . . 00 → 301	

227	I would like to ask a couple of questions about each of these people who died. To keep track of the different people we are discussing, could you tell me the first name of each person you know who died in the past 12 months? RECORD THE FIRST NAME OF EACH PERSON WHO HAS DIED AND ASK Q.228 AND 229 IF AGE IS NOT KNOWN, GET THE BEST POSSIBLE ESTIMATE IF AGE 95 OR MORE, RECORD '95'	228 Was (NAME) male or female?	229 How old was (NAME)?	
	NAME 1 _____	MALE . . . . . 1 FEMALE 2	☐☐	
	NAME 2 _____	MALE . . . . . 1 FEMALE 2	☐☐	
	NAME 3 _____	MALE . . . . . 1 FEMALE 2	☐☐	
	NAME 4 _____	MALE . . . . . 1 FEMALE 2	☐☐	
	NAME 5 _____	MALE . . . . . 1 FEMALE 2	☐☐	
	NAME 6 _____	MALE . . . . . 1 FEMALE 2	☐☐	
	NAME 7 _____	MALE . . . . . 1 FEMALE 2	☐☐	
	NAME 8 _____	MALE . . . . . 1 FEMALE 2	☐☐	
	NAME 9 _____	MALE . . . . . 1 FEMALE 2	☐☐	
	NAME 10 _____	MALE . . . . . 1 FEMALE 2	☐☐	
	NAME 11 _____	MALE . . . . . 1 FEMALE 2	☐☐	
	NAME 12 _____	MALE . . . . . 1 FEMALE 2	☐☐	

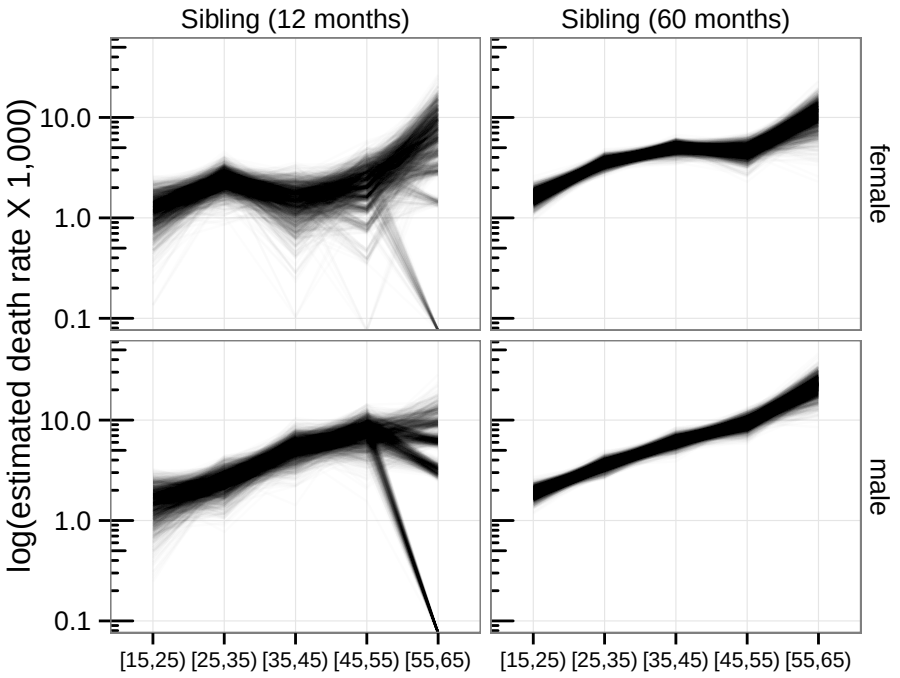


Figure D.4: Comparison between sibling estimates based on deaths reported 12 months and 60 months before the interview. The estimates from 12 months before the interview are very imprecise, while the estimates from 60 months before the survey are much more stable. Therefore, we use the 60-month estimates when we compare to the network survival results in the main text.

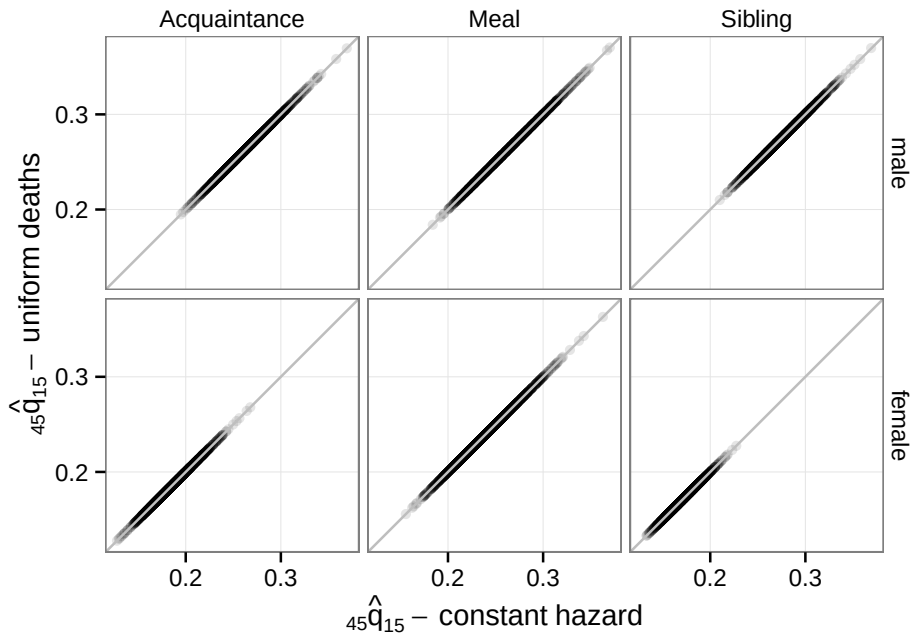


Figure D.5: Comparison between ${}_{45}q_{15}$ approximated from estimated age-specific death rates under two different assumptions: on the x axis, the approximations make the assumption that the estimated age-specific death rate is constant throughout the age interval; on the y axis, the approximations make the assumption that deaths are evenly spaced across the age interval. Results are shown for both network tie definitions, and also for sibling survival. All points are all very close to the $y = x$ line, indicating that the choice of approximation makes essentially no difference to the resulting estimates of ${}_{45}q_{15}$.

Bibliography

AbouZahr, C., de Savigny, D., Mikkelsen, L., Setel, P. W., Lozano, R., and Lopez, A. D. (2015a). Towards universal civil registration and vital statistics systems: the time is now. *The Lancet*.

AbouZahr, C., de Savigny, D., Mikkelsen, L., Setel, P. W., Lozano, R., Nichols, E., Notzon, F., and Lopez, A. D. (2015b). Civil registration and vital statistics: progress in the data revolution for counting and accountability. *The Lancet*.

Aramrattan, A. and Kanato, M. (2012). Network scale-up method: Application in Thailand. Presented at Consultation on estimating population sizes through household surveys: Successes and challenges (New York, NY).

Auguie, B. (2012). *gridExtra: functions in Grid graphics*. R package version 0.9.1.

Berchenko, Y., Rosenblatt, J., and Frost, S. D. W. (2013). Modeling and Analysing Respondent Driven Sampling as a Counting Process. *arXiv:1304.3505 [stat]*.

Bernard, H. R., Hallett, T., Iovita, A., Johnsen, E. C., Lyerla, R., McCarty, C., Mahy, M., Salganik, M. J., Saliuk, T., Scutelniciuc, O., Shelley, G. A., Sirinirund, P., Weir, S., and Stroup, D. F. (2010). Counting hard-to-count populations: the network scale-up method for public health. *Sexually Transmitted Infections*, 86(Suppl 2):ii11–ii15.

Bernard, H. R., Johnsen, E. C., Killworth, P. D., and Robinson, S. (1989). Estimating the Size of an Average Personal Network and of an Event Subpopulation. In Kochen, M., editor, *The Small World*, pages 159–175. Ablex Publishing.

Bernard, H. R., Johnsen, E. C., Killworth, P. D., and Robinson, S. (1991). Estimating the size of an average personal network and of an event subpopulation: Some empirical results. *Social Science Research*, 20(2):109–121.

Bowers, J., Fredrickson, M., and Hansen, B. (2010). RItools: randomization inference tools. *R package version 0.1-11*.

Bradshaw, D. and Timaeus, I. M. (2006). Levels and trends of adult mortality.

Brewer, D. D. (2000). Forgetting in the recall-based elicitation of person and social networks. *Social Networks*, 22:29–43.

Brewer, D. D. and Webster, C. M. (1999). Forgetting of friends and its effects on measuring friendship networks. *Social Networks*, 21(4):361–373.

Checchi, F. and Roberts, L. (2008). Documenting Mortality in Crises: What Keeps Us from Doing Better? *PLoS Med*, 5(7):e146.

Corsi, D. J., Neuman, M., Finlay, J. E., and Subramanian, S. V. (2012). Demographic and health surveys: a profile. *International journal of epidemiology*, 41(6):1602–1613.

Cowan, S. (2014). Secrets and Misperceptions: The Creation of Self-Fulfilling Illusions. *Sociological Science*, 1:466–492.

Crawford, F. W., Wu, J., and Heimer, R. (2015). Hidden population size estimation from respondent-driven sampling: a network approach. *arXiv:1504.08349 [stat]*.

Csárdi, G. and Nepusz, T. (2006). The igraph software package for complex network research. *InterJournal Complex Systems*, page 1695.

Danenberg, P. (2013). *functional: Curry, Compose, and other higher-order functions*. R package version 0.4.

DiPrete, T. A., Gelman, A., McCormick, T., Teitler, J., and Zheng, T. (2011). Segregation in social networks based on acquaintanceship and trust. *American Journal of Sociology*, 116(4):1234–1283.

Efron, B. and Tibshirani, R. J. (1993). *An Introduction to the Bootstrap*. Chapman and Hall/CRC.

Epicentre, W. G. f. M. E. i. E. (2007). Wanted: studies on mortality estimation methods for humanitarian emergencies, suggestions for future research. *Emerging Themes in Epidemiology*, 4(1):9.

Ezoe, S., Morooka, T., Noda, T., Sabin, M. L., and Koike, S. (2012). Population size estimation of men who have sex with men through the network scale-up method in Japan. *PLoS ONE*, 7(1):e31184.

Fabic, M. S., Choi, Y., and Bird, S. (2012). A systematic review of Demographic and Health Surveys: data availability and utilization for research. *Bulletin of the World Health Organization*, 90(8):604–612.

Feehan, D. M. and Salganik, M. J. (2014). *The networkreporting package*.

Feehan, D. M. and Salganik, M. J. (2015). Generalizing the Network Scale-Up Method: A New Estimator for the Size of Hidden Populations. *arXiv preprint arXiv:1404.4009*.

Feehan, D. M., Umubyeyi, A., Mahy, M., Hladik, W., and Salganik, M. J. (2015). Quantity vs quality: A survey experiment to improve the network scale-up method. *Under review*.

Fox, J. and Weisberg, S. (2011). *An R Companion to Applied Regression*. Sage, Thousand Oaks CA, second edition.

Gakidou, E., Hogan, M., and Lopez, A. D. (2004). Adult mortality: time for a reappraisal. *International Journal of Epidemiology*, 33(4):710–717.

Gakidou, E. and King, G. (2006). Death by Survey: Estimating Adult Mortality without Selection Bias from Sibling Survival Data. *Demography*, 43:569–585.

Gilovich, T., Savitsky, K., and Medvec, V. H. (1998). The illusion of transparency: Biased assessments of others' ability to read one's emotional states. *Journal of Personality and Social Psychology*, 75(2):332–346.

Goel, S., Mason, W., and Watts, D. J. (2010). Real and perceived attitude agreement in social networks. *Journal of Personality and Social Psychology*, 99(4):611–621.

Goel, S. and Salganik, M. J. (2009). Respondent-driven sampling as Markov chain Monte Carlo. *Statistics in Medicine*, 28(17):2202–2229.

Goel, S. and Salganik, M. J. (2010). Assessing respondent-driven sampling. *Proceedings of the National Academy of Science, USA*, 107(15):6743–6747.

Graham, W., Brass, W., and Snow, R. W. (1989). Estimating Maternal Mortality: The Sisterhood Method. *Studies in Family Planning*, 20(3):125–135.

Granovetter, M. (1973). The Strength of Weak Ties. *American Journal of Sociology*, 78(6):1360–1380.

Greenland, S. (1996). Basic methods for sensitivity analysis of biases. *International Journal of Epidemiology*, 25(6):1107–1116.

Greenland, S. (2003). The impact of prior distributions for uncontrolled confounding and response bias: a case study of the relation of wire codes and magnetic fields to childhood leukemia. *Journal of the American Statistical Association*, 98(461):47–54.

Greenland, S. (2005). Multiple-bias modelling for analysis of observational data. *Journal of the Royal Statistical Society: Series A (Statistics in Society)*, 168(2):267–306.

Groves, R. M. and Lyberg, L. (2010). Total Survey Error: Past, Present, and Future. *Public Opinion Quarterly*, 74(5):849–879.

Guo, W., Bao, S., Lin, W., Wu, G., Zhang, W., Hladik, W., Abdul-Quader, A., Bulterys, M., Fuller, S., and Wang, L. (2013). Estimating the size of HIV key affected populations in Chongqing, China using the network scale-up method. *PLoS ONE*, 8(8):e71796.

Handcock, M. S., Gile, K. J., and Mar, C. M. (2014). Estimating hidden population size using Respondent-Driven Sampling data. *Electronic Journal of Statistics*, 8(1):1491–1521.

Handcock, M. S., Gile, K. J., and Mar, C. M. (2015). Estimating the size of populations at high risk for HIV using respondent-driven sampling data. *Biometrics*, 71(1):258–266.

Hansen, B. B. and Bowers, J. (2008). Covariate balance in simple, stratified and clustered comparative studies. *Statistical Science*, pages 219–236.

Heckathorn, D. D. (1997). Respondent-driven sampling: A new approach to the study of hidden populations. *Social Problems*, 44(2):174–199.

Hedges, L. V. and Olkin, I. (1985). *Statistical Methods for Meta-Analysis*. Academic Press, Orlando, FL, 1 edition edition.

Helleringer, S., Duthé, G., Kanté, A. M., Andro, A., Sokhna, C., Trape, J.-F., and Pison, G. (2013). Misclassification of pregnancy-related deaths in adult mortality surveys: case study in Senegal. *Tropical Medicine & International Health*, 18(1):27–34.

Helleringer, S., Pison, G., Kanté, A. M., Duthé, G., and Andro, A. (2014a). Reporting errors in siblings' survival histories and their impact on adult mortality estimates: results from a record linkage study in Senegal. *Demography*, 51(2):387–411.

Helleringer, S., Pison, G., Masquelier, B., Kanté, A. M., Douillot, L., Duthé, G., Sokhna, C., and Delaunay, V. (2014b). Improving the quality of adult mortality data collected in demographic surveys: validation study of a new siblings' survival questionnaire in Niakhar, Senegal. *PLoS medicine*, 11(5):e1001652.

Hill, K. (2000). Methods for measuring adult mortality in developing countries: a comparative review. *The Global Burden of Disease in Aging Populations–Research Paper*, (01.13).

Hill, K. (2003). Adult mortality in the developing world: what we know and how we know it. Technical report, UN.

Hill, K. and Choi, Y. (2004). The adult mortality in developing countries project: substantive findings. In *Adult Mortality in Developing Countries Workshop*.

Hill, K., Choi, Y., and Timaeus, I. (2005). Unconventional approaches to mortality estimation. *Demographic Research*, 13(12):281–300.

Hill, K., El Arifeen, S., Koenig, M., Al-Sabir, A., Jamil, K., and Raggers, H. (2006). How should we measure maternal mortality in the developing world? A comparison of household deaths and sibling history approaches. *Bulletin of the World Health Organization*, 84(3):173–180.

Hill, K., Lopez, A. D., Shibuya, K., Jha, P., Monitoring of Vital Events (MoVE) writing group, and others (2007). Interim measures for meeting needs for health sector data: births, deaths, and causes of death. *The Lancet*, 370(9600):1726–1735.

Hill, K. and Trussell, J. (1977). Further developments in indirect mortality estimation. *Population Studies*, 31(2):313–334.

Hill, K., Zlotnik, H., Trussell, J., United Nations Department of International Economic, Social Affairs Population Division, and others (1983). *Manual X: Indirect techniques for demographic estimation*. UN.

Hlavac, M. (2014). *stargazer: LaTeX code and ASCII text for well-formatted regression and summary statistics tables.* R package version 5.1.

ICF International (2012). *Demographic and Health Survey Sampling and Household Listing Manual*. MEASURE DHS. ICF International, Calverton, Maryland, U.S.A.

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

Jing, L., Qu, C., Yu, H., Wang, T., and Cui, Y. (2014). Estimating the sizes of populations at high risk for HIV: A comparison study. *PLoS ONE*, 9(4):e95601.

Johnsen, E., Bernard, R., Killworth, P., Shelley, G., and McCarty, C. (1995). A social network approach to corroborating the number of AIDS/HIV+ victims in the US. *Social Networks*, 17(3-4):167–187.

Kadushin, C., Killworth, P. D., Bernard, H. R., and Beveridge, A. A. (2006). Scale-up methods as applied to estimates of heroin use. *Journal of Drug Issues*, 36(2):417–440.

Kalton, G. and Anderson, D. W. (1986). Sampling Rare Populations. *Journal of the Royal Statistical Society, Series A*, 149(1):65–82.

Karon, J. and Wejnert, C. (2012). Statistical methods for the analysis of time–location sampling data. *Journal of Urban Health*, 89(3):565–586.

Kassebaum, N. J., Bertozzi-Villa, A., Coggeshall, M. S., Shackelford, K. A., Steiner, C., Heuton, K. R., Gonzalez-Medina, D., Barber, R., Huynh, C., Dicker, D., and others (2014). Global, regional, and national levels and causes of maternal mortality during 1990–2013: a systematic analysis for the Global Burden of Disease Study 2013. *The Lancet*, 384(9947):980–1004.

Khounigh, A. J., Haghdoost, A. A., Lak, S. S., Zeinalzadeh, A. H., Yousefi-Farkhad, R., Mohammadzadeh, M., and Holakouie-Naieni, K. (2014). Size estimation of most-at-risk groups of HIV/AIDS using network scale-up in Tabriz, Iran. *Journal of Clinical Research & Governance*, 3(1):21–26.

Killworth, P. D., Johnsen, E. C., McCarty, C., Shelley, G. A., and Bernard, H. R. (1998a). A social network approach to estimating seroprevalence in the United States. *Social Networks*, 20(1):23–50.

Killworth, P. D., McCarty, C., Bernard, H. R., Johnsen, E. C., Domini, J., and Shelly, G. A. (2003). Two interpretations of reports of knowledge of subpopulation sizes. *Social Networks*, 25(2):141–160.

Killworth, P. D., McCarty, C., Bernard, H. R., Shelley, G. A., and Johnsen, E. C. (1998b). Estimation of seroprevalence, rape, and homelessness in the United States using a social network approach. *Evaluation Review*, 22(2):289–308.

Killworth, P. D., McCarty, C., Johnsen, E. C., Bernard, H. R., and Shelley, G. A. (2006). Investigating the variation of personal network size under unknown error conditions. *Sociological Methods & Research*, 35(1):84–112.

Laumann, E. O. (1969). Friends of urban men: An assessment of accuracy in reporting their socioeconomic attributes, mutual choice, and attitude agreement. *Sociometry*, 32(1):54–69.

Lavallée, P. (2007). *Indirect Sampling*. Springer.

Maghsoudi, A., Baneshi, M. R., Neydavoodi, M., and Haghdoost, A. (2014). Network scale-up correction factors for population size estimation of people who inject drugs and female sex workers in Iran. *PLoS ONE*, 9(11):e110917.

Maltiel, R., Raftery, A. E., and McCormick, T. H. (2015). Estimating Population Size Using the Network Scale Up Method. *Annals of Applied Statistics*, (Forthcoming).

Marin, A. (2004). Are respondents more likely to list alters with certain characteristics?: Implications for name generator data. *Social Networks*, 26(4):289–307.

Marsden, P. V. (1990). Network Data and Measurement. *Annual Review of Sociology*, 16:435–463.

Mas-Colell, A., Whinston, M. D., Green, J. R., and others (1995). *Microeconomic theory*, volume 1. Oxford university press New York.

Masquelier, B. (2013). Adult mortality from sibling survival data: a reappraisal of selection biases. *Demography*, 50(1):207–228.

Masquelier, B. and Dutreuilh, C. (2014). Sibship Sizes and Family Sizes in Survey Data Used to Estimate Mortality. *Population, English edition*, 69(2):221–238.

Masquelier, B., Reniers, G., and Pison, G. (2014). Divergences in trends in child and adult mortality in sub-Saharan Africa: survey evidence on the survival of children and siblings. *Population studies*, 68(2):161–177.

McCarty, C., Killworth, P. D., Bernard, H. R., Johnsen, E. C., and Shelley, G. A. (2001). Comparing two methods for estimating network size. *Human Organization*, 60(1):28–39.

McCormick, T., He, R., Kolaczyk, E., and Zheng, T. (2012a). Surveying Hard-to-Reach Groups Through Sampled Respondents in a Social Network. *Statistics in Biosciences*, pages 1–19.

McCormick, T., He, R., Kolaczyk, E., and Zheng, T. (2012b). Surveying hard-to-reach groups through sampled respondents in a social network. *Statistics in Biosciences*, pages 1–19.

McCormick, T. H., Salganik, M. J., and Zheng, T. (2010). How many people do you know?: Efficiently estimating personal network size. *Journal of the American Statistical Association*, 105(489):59–70.

McCormick, T. H. and Zheng, T. (2007). Adjusting for recall bias in "how many X's do you know?" surveys. Conference Procedings of the Joint Statistical meetings 2007: Vancover, British Columbia.

McGill, R., Tukey, J. W., and Larsen, W. A. (1978). Variations of box plots. *The American Statistician*, 32(1):12–16.

Merdad, L., Hill, K., and Graham, W. (2013). Improving the measurement of maternal mortality: the sisterhood method revisited. *PloS one*, 8(4):e59834.

Mikkelsen, L., Phillips, D. E., AbouZahr, C., Setel, P. W., de Savigny, D., Lozano, R., and Lopez, A. D. (2015). A global assessment of civil registration and vital statistics systems: monitoring data quality and progress. *The Lancet*.

Moultrie, T., Dorrington, R., Hill, A., Hill, K., Timaeus, I., and Zaba, B., editors (2013). *Tools for demographic estimation*. International Union for the Scientific Study of Population, Paris.

NISR, MOH, and ICF (2012). Rwanda Demographic and Health Survey 2010. Technical report, National Institute of Statistics of Rwanda (NISR), Ministry of Health (MOH) [Rwanda], and ICF International, Calverton, Maryland, USA.

Obermeyer, Z., Rajaratnam, J. K., Park, C. H., Gakidou, E., Hogan, M. C., Lopez, A. D., and Murray, C. J. L. (2010). Measuring Adult Mortality Using Sibling Survival: A New Analytical Method and New Results for 44 Countries, 1974–2006. *PLoS Med*, 7(4):e1000260.

Paniotto, V., Petrenko, T., Kupriyanov, V., and Pakhok, O. (2009). Estimating the size of populations with high risk for HIV using the network scale-up method. Technical report, Kiev Internation Institute of Sociology.

Preston, S. H., Heuveline, P., and Guillot, M. (2001). *Demography: measuring and modeling population processes*.

R Core Team (2014). *R: A Language and Environment for Statistical Computing*. R Foundation for Statistical Computing, Vienna, Austria.

Rajaratnam, J. K., Marcus, J. R., Levin-Rector, A., Chalupka, A. N., Wang, H., Dwyer, L., Costa, M., Lopez, A. D., and Murray, C. J. (2010). Worldwide mortality in men and women aged 15–59 years from 1970 to 2010: a systematic analysis. *The Lancet*, 375(9727):1704–1720.

Rao, J., Wu, C., and Yue, K. (1992). Some recent work on resampling methods for complex surveys. *Survey Methodology*, 18(2):209–217.

Rao, J. N. and Wu, C. F. J. (1988a). Resampling inference with complex survey data. *Journal of the American Statistical Association*, 83(401):231–241.

Rao, J. N. K. and Pereira, N. P. (1968). On double ratio estimators. *Sankhyā: The Indian Journal of Statistics, Series A (1961-2002)*, 30(1):83–90.

Rao, J. N. K. and Wu, C. F. J. (1988b). Resampling inference with complex survey data. *Journal of the American Statistical Association*, 83(401):231–241.

RBC/IHDPC, SPH, UNAIDS, and ICF (2012). *Estimating the Size of Key Populations at Higher Risk of HIV through a Household Survey*. Rwanda Biomedical Center/Institute of HIV/AIDS, Disease Prevention and Control Department (RBC/IHDPC), School of Public Health (SPH), UNAIDS, and ICF International, Calverton, Maryland, USA.

Reniers, G., Masquelier, B., and Gerland, P. (2011). Adult mortality in Africa. *International Handbook of Adult Mortality*, pages 151–170.

Rust, K. and Rao, J. (1996). Variance estimation for complex surveys using replication techniques. *Statistical Methods in Medical Research*, 5(3):283–310.

Rutenberg, N. and Sullivan, J. (1991). Direct and indirect estimates of maternal mortality from the sisterhood method. [Unpublished] 1991. Presented at the Demographic and Health Surveys World Conference Washington DC August 5-7 1991.

Rutstein, S. O. and Rojas, G. (2006). Guide to DHS statistics. *Calverton, Maryland: ORC Macro*.

Rwanda Biomedical Center (2010). Behavioral and biological surveillance survey among female sex workers.

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

Salganik, M. J. (2006). Variance estimation, design effects, and sample size calculations for respondent-driven sampling. *Journal of Urban Health*, 83(7):98–112.

Salganik, M. J., Fazito, D., Bertoni, N., Abdo, A. H., Mello, M. B., and Bastos, F. I. (2011a). 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. *American Journal of Epidemiology*, 174(10):1190–1196.

Salganik, M. J., Mello, M. B., Abdo, A. H., Bertoni, N., Fazito, D., and Bastos, F. I. (2011b). The game of contacts: Estimating the social visibility of groups. *Social Networks*, 33(1):70–78.

Sarndal, C. E., Swensson, B., and Wretman, J. (2003). *Model assisted survey sampling*. Springer Verlag.

Scutelniciuc, O. (2012a). Network scale-up method experiences: Republic of Kazakhstan. Presented at Consultation on estimating population sizes through household surveys: Successes and challenges (New York, NY).

Scutelniciuc, O. (2012b). Network scale-up method experiences: Republic of Moldova. Presented at Consultation on estimating population sizes through household surveys: Successes and challenges (New York, NY).

Setel, P. W., Macfarlane, S. B., Szreter, S., Mikkelsen, L., Jha, P., Stout, S., and AbouZahr, C. (2007). A scandal of invisibility: making everyone count by counting everyone. *The Lancet*, 370(9598):1569–1577.

Shao, J. (2003). Impact of the bootstrap on sample surveys. *Statistical Science*, 18(2):191–198.

Sheikhzadeh, K., Baneshi, M. R., Afshari, M., and Haghdoost, A. A. (2014). Comparing direct, network scale-up, and proxy respondent methods in estimating risky behaviors among collegians. *Journal of Substance Use*, pages 1–5.

Shelley, G. A., Bernard, H. R., Killworth, P., Johnsen, E., and McCarty, C. (1995). Who knows your HIV status? What HIV+ patients and their network members know about each other. *Social Networks*, 17(3-4):189–217.

Shelley, G. A., Killworth, P. D., Bernard, H. R., McCarty, C., Johnsen, E. C., and Rice, R. E. (2006). Who Knows Your HIV status II?: Information propagation within social networks of seropositive people. *Human Organization*, 65(4):430–444.

Shokoohi, M., Baneshi, M. R., and Haghdoost, A.-A. (2012). Size estimation of groups at high risk of HIV/AIDS using network scale up in Kerman, Iran. *International Journal of Preventive Medicine*, 3(7):471–476.

Sirken, M. G. (1970). Household surveys with multiplicity. *Journal of the American Statistical Association*, 65(329):257–266.

Snidero, S., Morra, B., Corradetti, R., and Gregori, D. (2007). Use of the scale-up methods in injury prevention research: An empirical assessment to the case of choking in children. *Social Networks*, 29(4):527–538.

Snidero, S., Soriani, N., Baldi, I., Zobec, F., Berchialla, P., and Gregori, D. (2012). Scale-up approach in CATI surveys for estimating the number of foreign body injuries in the aero-digestive tract in children. *International Journal of Environmental Research and Public Health*, 9(11):4056–4067.

Snidero, S., Zobec, F., Berchialla, P., Corradetti, R., and Gregori, D. (2009). Question order and interviewer effects in CATI scale-up surveys. *Sociological Methods & Research*, 38(2):287–305.

Stanton, C., Abderrahim, N., and Hill, K. (2000). An assessment of DHS maternal mortality indicators. *Studies in family planning*, 31(2):111–123.

Sudman, S. (1988). Experiments in measuring neighbor and relative social networks. *Social Networks*, (10):93–108.

Thompson, S. K. (2002). *Sampling*. Wiley, 2nd edition.

Timaeus, I. M. (1991). Measurement of adult mortality in less developed countries: a comparative review. *Population Index*, pages 552–568.

Timaeus, I. M. and Jasseh, M. (2004). Adult mortality in sub-Saharan Africa: evidence from Demographic and Health Surveys. *Demography*, 41(4):757–772.

Trussell, J. and Rodriguez, G. (1990). A Note on the Sisterhood Estimator of Maternal Mortality. *Studies in Family Planning*, 21(6):344–346.

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

UNAIDS (2010). Guidelines on Estimating the Size of Populations Most at Risk to HIV. Technical report, UNAIDS/WHO Working Group on Global HIV/AIDS and STI Surveillance, Geneva, Switzerland.

UNAIDS (2013). Global report: UNAIDS report on the global AIDS epidemic 2013. Technical report, Joint United Nations Programme on HIV/AIDS (UNAIDS, Geneva, Switzerland.

UNAIDS (2014). *Quick Start for Spectrum*. Joint United Nations Programme on HIV/AIDS, Geneva, Switzerland.

UNAIDS and WHO (2011). Guidelines on surveillance among populations most at risk for HIV. Technical report, UNAIDS and WHO Working Group on Global HIV/AIDS and STI Surveillance, World Health Organization, Geneva, Switzerland.

United Nations (2013). World population prospects: The 2012 revision. *Population Division, Department of Economic and Social Affairs, United Nations, New York*.

Varian, H. R. (1992). Microeconomic analysis.

Verdery, A. M., Mouw, T., Bauldry, S., and Mucha, P. J. (2013). Network structure and biased variance estimation in respondent driven sampling. *arXiv:1309.5109 [cs, stat]*.

Wachter, K. W. (2014). *Essential demographic methods*. Harvard University Press.

Wang, H., Dwyer-Lindgren, L., Lofgren, K. T., Rajaratnam, J. K., Marcus, J. R., Levin-Rector, A., Levitz, C. E., Lopez, A. D., and Murray, C. J. (2013). Age-specific and sex-specific mortality in 187 countries, 1970–2010: a systematic analysis for the Global Burden of Disease Study 2010. *The Lancet*, 380(9859):2071–2094.

Wasserman, S. and Faust, K. (1994). *Social Network Analysis*. Cambridge University Press.

White, H. C., Boorman, S. A., and Breiger, R. L. (1976). Social structure from multiple networks. I. Blockmodels of roles and positions. *American Journal of Sociology*, 81(4):730–780.

WHO (2015). Global Health Observatory Data Repository.

Wickham, H. (2009). *ggplot2: elegant graphics for data analysis*. Springer New York.

Wickham, H. (2011). The split-apply-combine strategy for data analysis. *Journal of Statistical Software*, 40(1):1–29.

Wickham, H. (2012). *stringr: Make it easier to work with strings.* R package version 0.6.2.

Wickham, H. and Chang, W. (2013). *devtools: Tools to make developing R code easier*. R package version 1.4.1.

Wickham, H. and Francois, R. (2015). *dplyr: A Grammar of Data Manipulation*. R package version 0.4.1.

Yamanis, T. J., Merli, M. G., Neely, W. W., Tian, F. F., Moody, J., Tu, X., and Gao, E. (2013). An empirical analysis of the impact of recruitment patterns on RDS estimates among a socially ordered population of female sex workers in China. *Sociological Methods & Research*, 42(3):392–425.

Zheng, T., Salganik, M. J., and Gelman, A. (2006). How many people do you know in prison?: Using overdispersion in count data to estimate social structure in networks. *Journal of the American Statistical Association*, 101(474):409–423.