

# Estimating the Population Size for Women with Sexual Violence Related Pregnancies in South Kivu, Democratic Republic of Congo

Lisa Johnston, MA, MPH, PhD, Katherine R. McLaughlin, MSc, Shada A. Rouhani, MD, MPH; Susan Bartels, MD, MPH

On behalf of

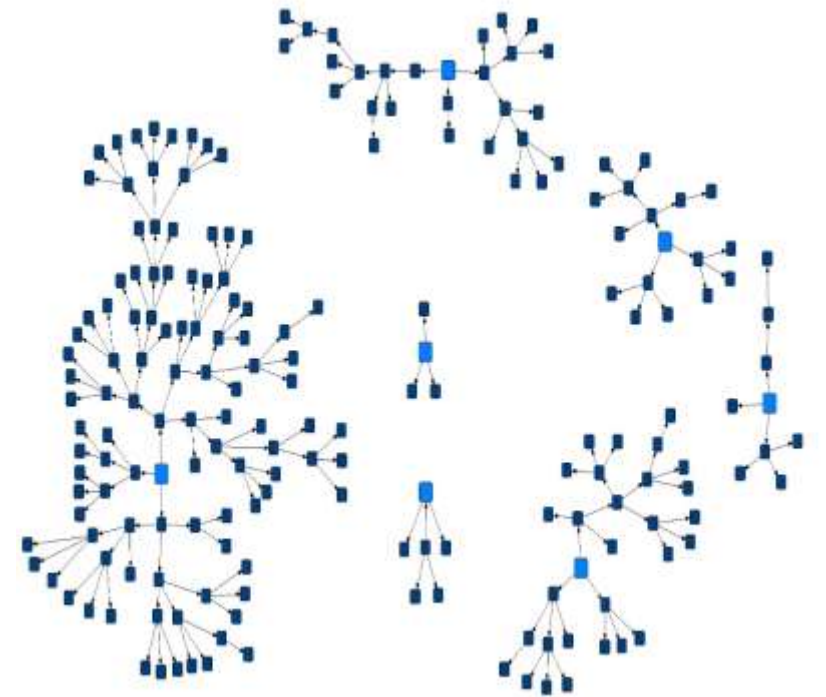
Harvard Humanitarian Initiative, Cambridge, Massachusetts, United States

# Background

- Females involved in conflict-related sexual violence in eastern Democratic Republic of Congo (DRC) are at risk of sexual violence-related pregnancies (SVRPs).
- Considered a hidden and hard-to-reach population.
- Few methods available to sample and estimate the sizes of hidden and hard-to-reach populations.
- Estimating the sizes of these populations helps inform the magnitude of the issue, guide resource allocation and advocate for needed programs.

# Respondent Driven Sampling (RDS)

- RDS used to sample 852 women with SVRPs for 'Self-study' in 2012.
- Peer to peer sampling method to sample hidden, socially networked populations.
- Begin with a set of 'seeds'.
- Seeds recruit peers, who recruit peers, etc.
- Recruits are linked by set number of coupons with unique identifying numbers.
- Recruitment is limited through a minimum set of coupons.
- Incentives provided for completed survey.
- Relies on analytical adjustments for differential recruitment/social network sizes.



# Social network size

Based on eligibility of 'Self-study':

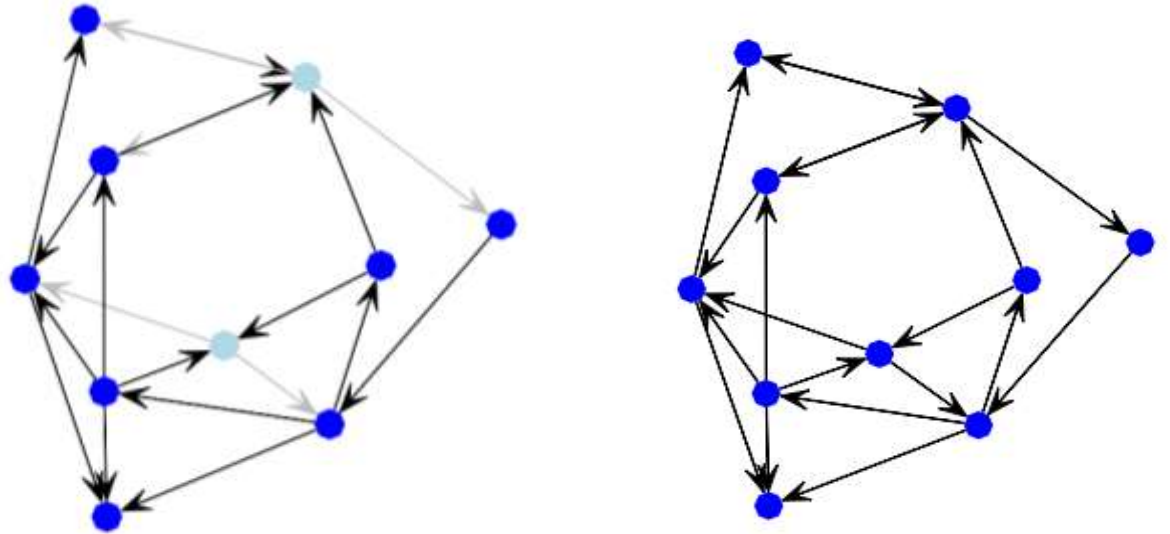
- females
- self-identified as survivors of sexual violence since the start of the war (in 1996) in eastern DRC
- became pregnant as a result of sexual violence
- were either raising the child or aborted the fetus (women who had a spontaneous miscarriage, still birth or were not raising the child were excluded)
- 18 years of age or older

# Successive Sampling-Population Size estimation SS-PSE

- Uses data from an RDS survey
  - Social network sizes (information about the network structure)
  - Enrollment date (order in which people were sampled)
  - Previous knowledge about population size
    - Expert belief (researcher, stakeholder)
    - PSE from other methods
    - Literature review
  - Maximum number of coupons distributed to each person
  - SS estimator assumptions (sampling proceeds according to a successive sampling procedure-each subsequent sample is selected from among the remaining units with probability proportional to size)

# SS-PSE Features UPDATE

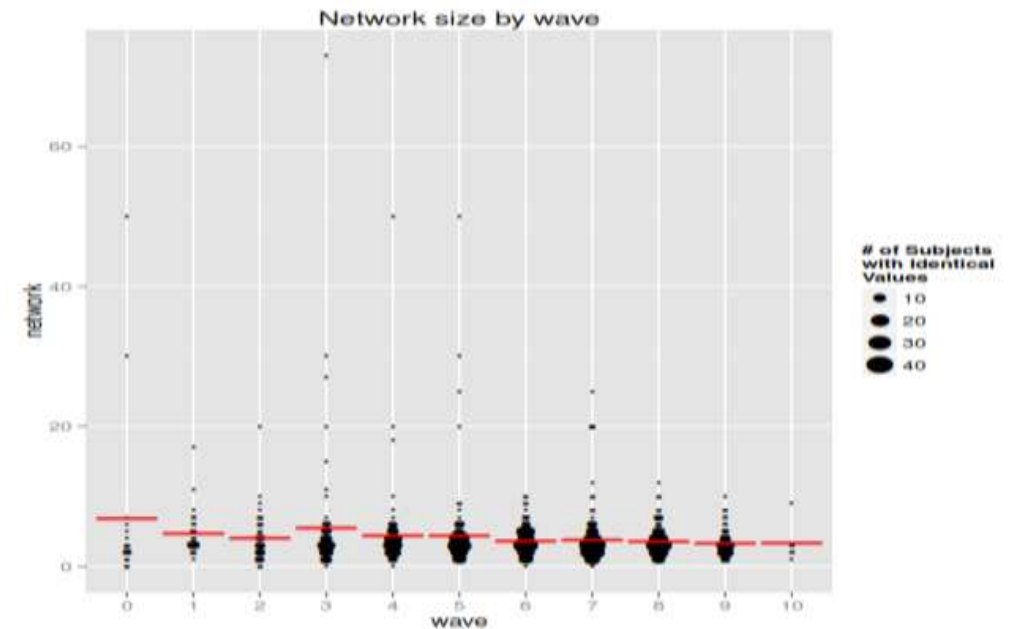
- Uses a Bayesian framework using information about prior knowledge and educated approximations of the population size.
- We assume that network structures in RDS follow a configuration model in which networks form completely at random among a population with fixed and observable degrees<sup>1</sup>.



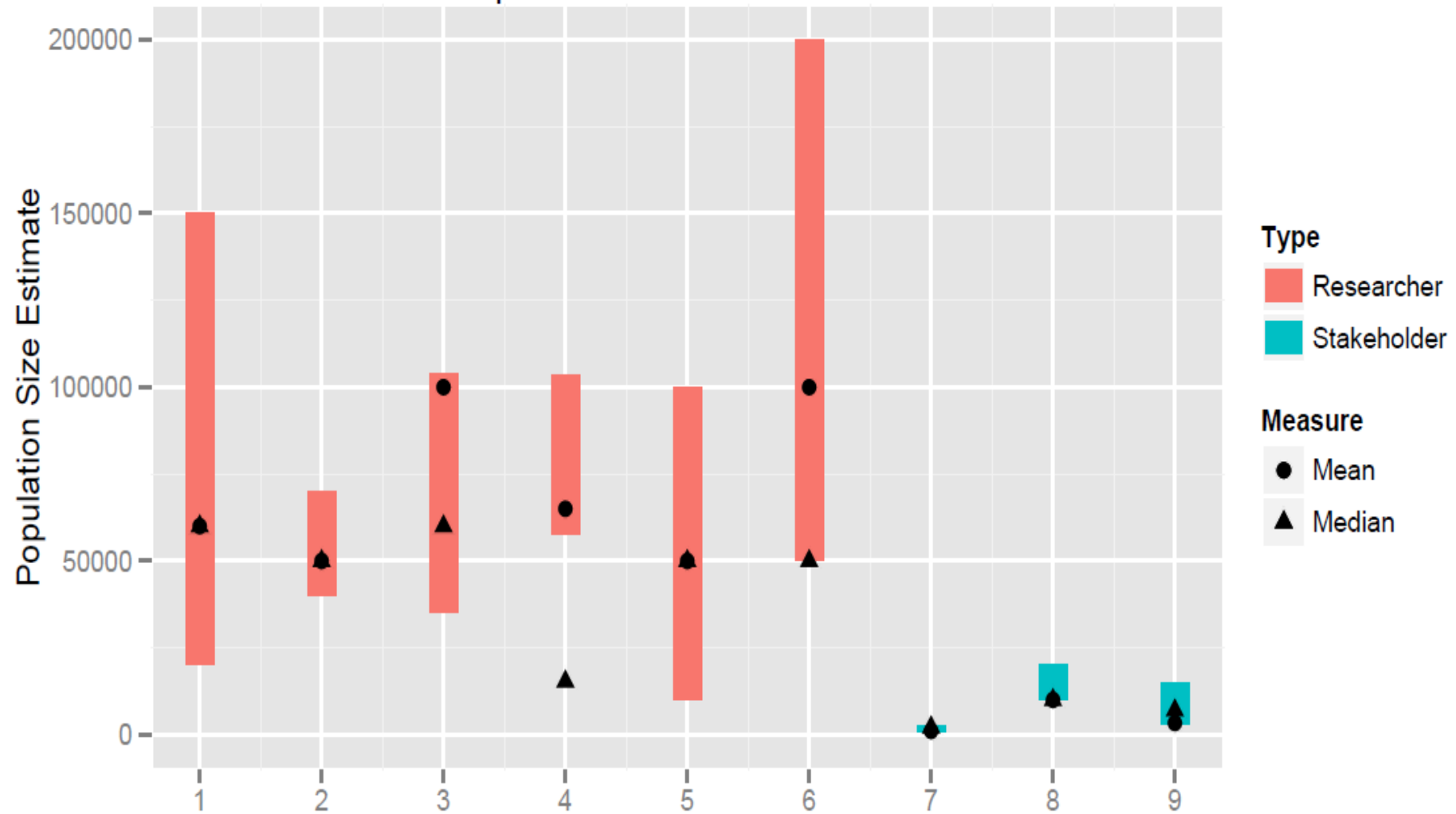
1. Standard RDS assumption but likely violated since one person does not know all member of the population.

# Conceptual Overview: SS-PSE

- More visible people (tend to have higher degree) are more likely to be sampled, and be sampled earlier in RDS
- If the frequency of high degrees decreases over RDS waves, population is being depleted, population size is likely smaller

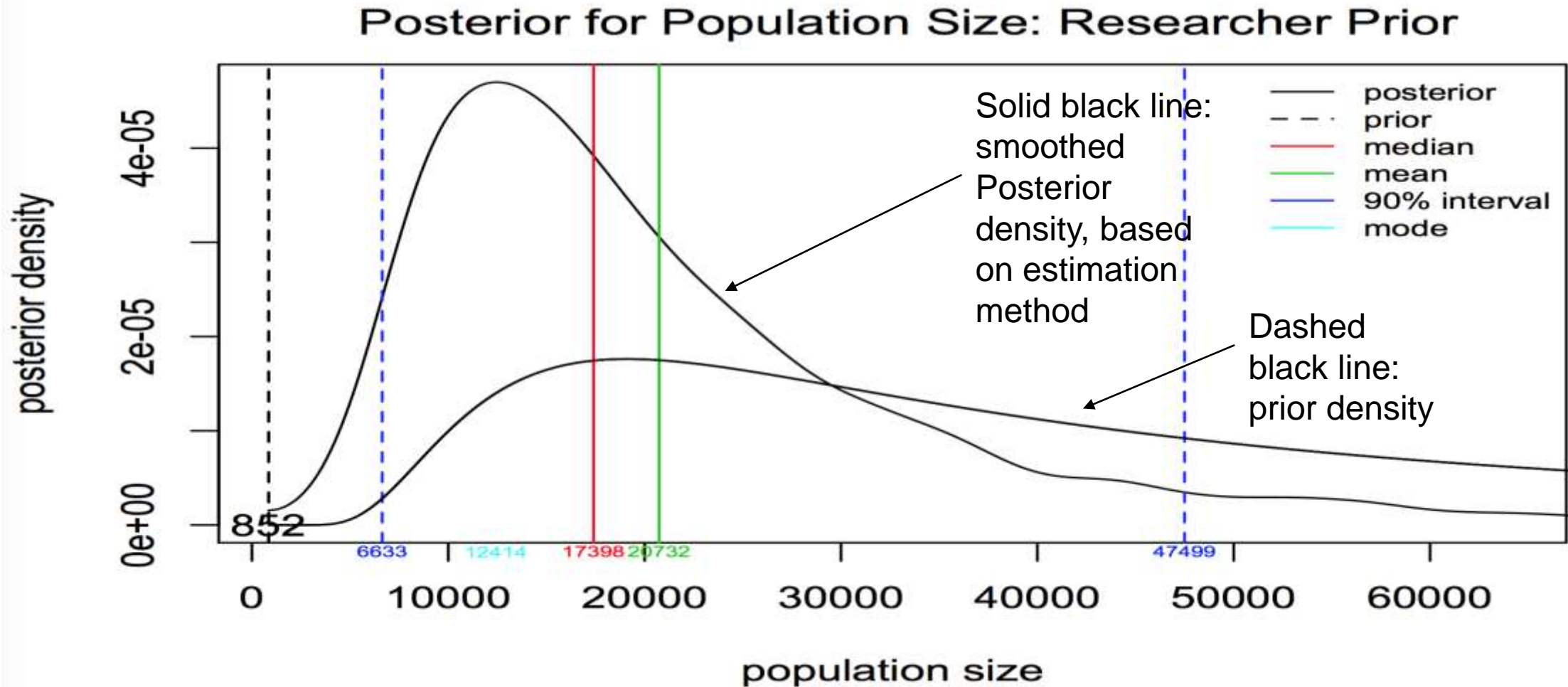


# SS-PSE: Prior Estimates(1)





# Plot of the posterior distribution using the average of the six researcher prior values



# Plot of the posterior distribution using the average of the six researcher prior values

---

Researchers (n=7)

---

	Mean	Median	Mode	25%	75%	90%	5%	95%
Prior	57645	42584	19100	24076	78490	127387	10074	175834
Posterior	20732	17398	12414	11816	26169	37731	6633	47499

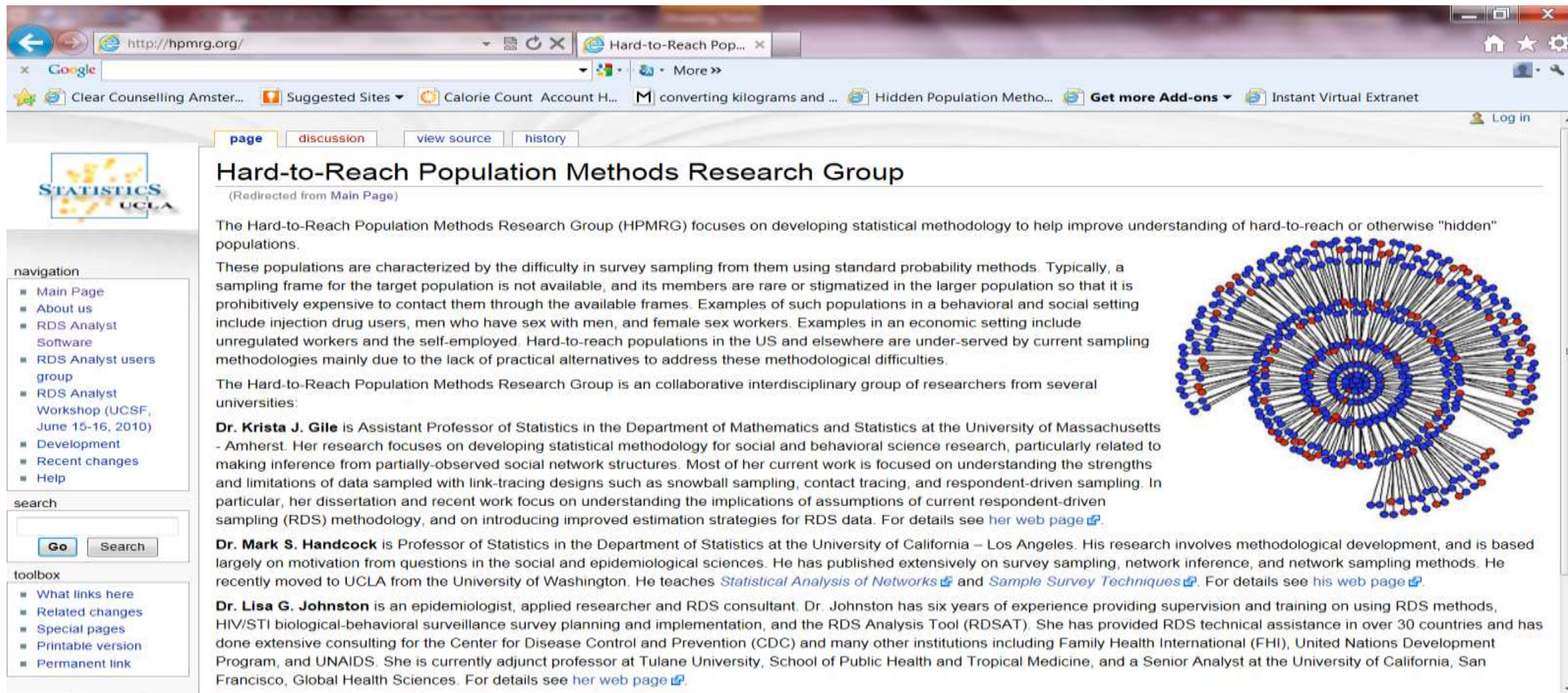
# Verifying our findings

- Median estimate of 17,398--2% of South Kivu women of reproductive age and 15% of South Kivu women who have experienced sexual violence
- Previous estimates:
  - National Institute of Statistical Sciences (2006) – approx. 20,000
  - Expanded Program on Immunization (2007) – approx. 20,000
  - Dr. Mukwege (2014) – 40,000

# Limitations

- Provides estimates (will never know the truth)
- Relies on numerous assumptions
- Relies on there being some prior knowledge
- May not do as well if the sampling fraction is small
- Relies on the quality of the survey (bottlenecks, sub populations, network size errors, etc.)

# RDS ANALYST at WWW.HPMRG.ORG



The screenshot shows a web browser window with the URL <http://hpmrg.org/>. The page title is "Hard-to-Reach Population Methods Research Group" and it is noted as "(Redirected from Main Page)".

**Navigation:**

- Main Page
- About us
- RDS Analyst Software
- RDS Analyst users group
- RDS Analyst Workshop (UCSF, June 15-16, 2010)
- Development
- Recent changes
- Help

**Search:**

Go Search

**Toolbox:**

- What links here
- Related changes
- Special pages
- Printable version
- Permanent link

**Hard-to-Reach Population Methods Research Group**  
(Redirected from Main Page)

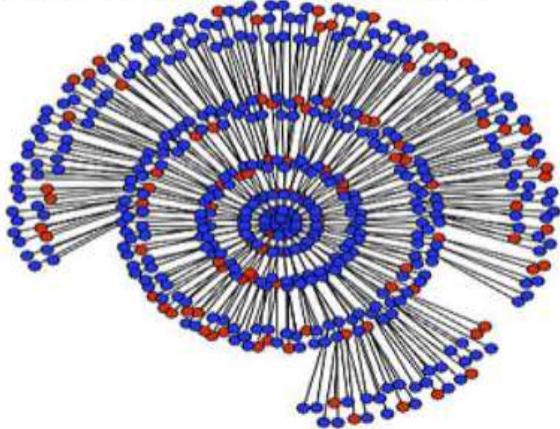
The Hard-to-Reach Population Methods Research Group (HPMRG) focuses on developing statistical methodology to help improve understanding of hard-to-reach or otherwise "hidden" populations. These populations are characterized by the difficulty in survey sampling from them using standard probability methods. Typically, a sampling frame for the target population is not available, and its members are rare or stigmatized in the larger population so that it is prohibitively expensive to contact them through the available frames. Examples of such populations in a behavioral and social setting include injection drug users, men who have sex with men, and female sex workers. Examples in an economic setting include unregulated workers and the self-employed. Hard-to-reach populations in the US and elsewhere are under-served by current sampling methodologies mainly due to the lack of practical alternatives to address these methodological difficulties.

The Hard-to-Reach Population Methods Research Group is an collaborative interdisciplinary group of researchers from several universities:

**Dr. Krista J. Gile** is Assistant Professor of Statistics in the Department of Mathematics and Statistics at the University of Massachusetts - Amherst. Her research focuses on developing statistical methodology for social and behavioral science research, particularly related to making inference from partially-observed social network structures. Most of her current work is focused on understanding the strengths and limitations of data sampled with link-tracing designs such as snowball sampling, contact tracing, and respondent-driven sampling. In particular, her dissertation and recent work focus on understanding the implications of assumptions of current respondent-driven sampling (RDS) methodology, and on introducing improved estimation strategies for RDS data. For details see [her web page](#).

**Dr. Mark S. Handcock** is Professor of Statistics in the Department of Statistics at the University of California – Los Angeles. His research involves methodological development, and is based largely on motivation from questions in the social and epidemiological sciences. He has published extensively on survey sampling, network inference, and network sampling methods. He recently moved to UCLA from the University of Washington. He teaches [Statistical Analysis of Networks](#) and [Sample Survey Techniques](#). For details see [his web page](#).

**Dr. Lisa G. Johnston** is an epidemiologist, applied researcher and RDS consultant. Dr. Johnston has six years of experience providing supervision and training on using RDS methods, HIV/STI biological-behavioral surveillance survey planning and implementation, and the RDS Analysis Tool (RDSAT). She has provided RDS technical assistance in over 30 countries and has done extensive consulting for the Center for Disease Control and Prevention (CDC) and many other institutions including Family Health International (FHI), United Nations Development Program, and UNAIDS. She is currently adjunct professor at Tulane University, School of Public Health and Tropical Medicine, and a Senior Analyst at the University of California, San Francisco, Global Health Sciences. For details see [her web page](#).



# Thank you

## Publications:

- Handcock MS, Gile KJ, and Mar CM. Estimating hidden population size using respondent-driven sampling data. *Electronic Journal of Statistics*. 2014a. 8(1), 1491-1521.
- Handcock MS, Gile KJ, and Mar CM. Estimating the size of populations at high risk for HIV using respondent-driven sampling data. *Biometrics*. 2014b. 71, 258-266.
- Johnston LG, McLaughlin K, et al. A novel method for estimating the size of hidden populations using respondent-driven sampling data: Case examples from Morocco. 2015. *Epidemiology*, In Press.
- McLaughlin K, Johnston LG, Handcock, MS. Inference for the visibility distribution of Respondent Driven Sampling data. (in progress)
- Johnston LG, Bartels SA, Rouhani SA, McLaughlin K. Estimating the population size for women with sexual violence related pregnancies in eastern Democratic Republic of Congo. (in progress)
- Gile KJ, Handcock MS. Respondent-Driven Sampling: An Assessment of Current Methodology. *Sociological Methods*, 2010. 40, 1: 285-327