

Lecture 22: Network scale-up method to study groups most at-risk for HIV

Matthew J. Salganik

Sociology 204: Social Networks
Princeton University

Pre-read



There are an estimated 38 million people [31.6 million–44.5 million] living with HIV in 2019. In most countries, the disease is concentrated in three high risk groups:

- ▶ drug users
- ▶ commercial sex workers
- ▶ men who have sex with men

Better information about these group can be used to understand and control the spread of HIV/AIDS: “know your epidemic”

Two main questions:

- ▶ prevalence of some trait within a hidden population (e.g., What proportion of sex workers in Moscow have HIV/AIDS?):

Two main questions:

- ▶ prevalence of some trait within a hidden population (e.g., What proportion of sex workers in Moscow have HIV/AIDS?): respondent-driven sampling (lecture 13)

Two main questions:

- ▶ prevalence of some trait within a hidden population (e.g., What proportion of sex workers in Moscow have HIV/AIDS?): respondent-driven sampling (lecture 13)
- ▶ size of hidden population (e.g., How many sex workers are there in Moscow?)

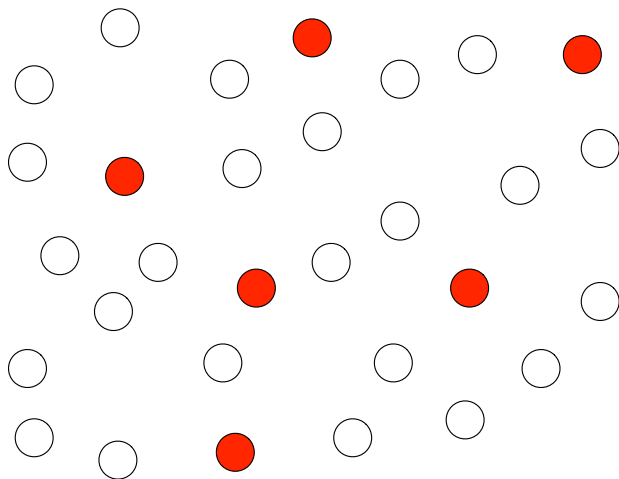
Two main questions:

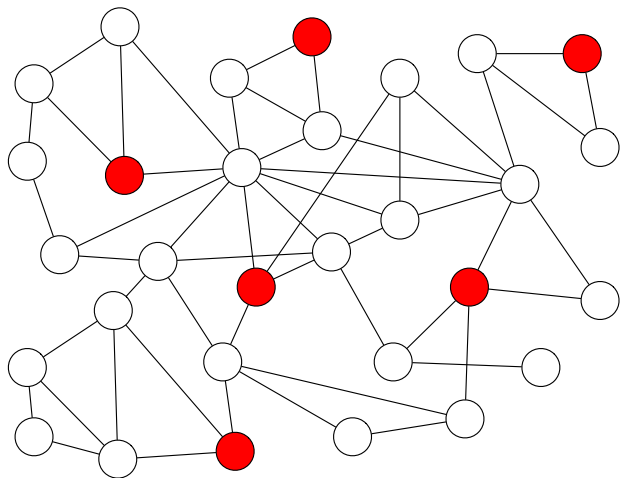
- ▶ prevalence of some trait within a hidden population (e.g., What proportion of sex workers in Moscow have HIV/AIDS?): respondent-driven sampling (lecture 13)
- ▶ size of hidden population (e.g., How many sex workers are there in Moscow?)
network scale-up method (this lecture)

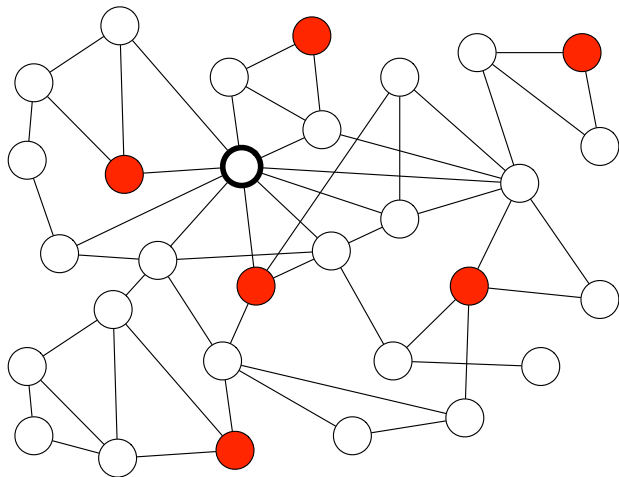
Network scale-up method

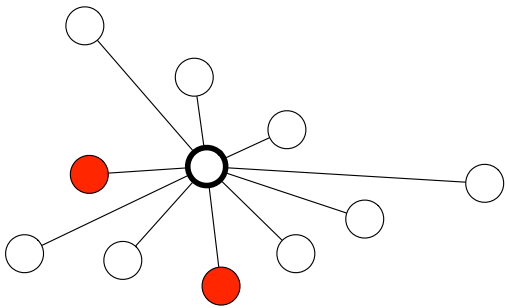


Basic insight from Bernard et al. (1989)



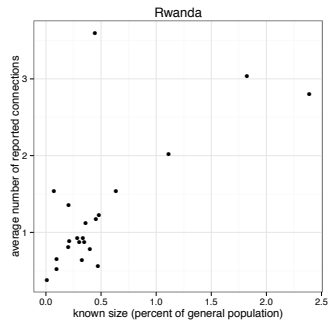
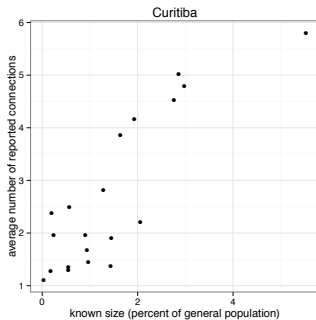
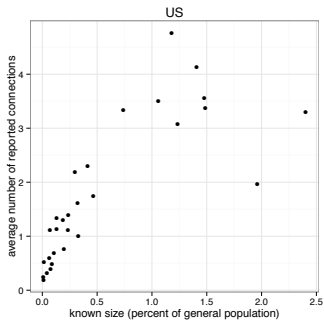






$$\hat{N}_H = \frac{2}{10} \times 30 = 6$$

- ▶ Requires a random sample from the entire population
- ▶ Respondents are asked:
 - ▶ How many people do you know who are drug injectors?
 - ▶ How many women do you know that have given birth in the last 12 months?
 - ▶ How many people do you know who are middle school teachers?
 - ▶ ...
 - ▶ How many people do you know named Michael?
- ▶ “Know” typically defined: you know them and they know you and have you been in contact with them over the past two years



On average, these answers are not crazy.

Other size estimation methods are problematic, and scale-up method has many nice properties:

- ▶ Requires a random sample of the general population, not specific contact with the hard-to-reach population

Other size estimation methods are problematic, and scale-up method has many nice properties:

- ▶ Requires a random sample of the general population, not specific contact with the hard-to-reach population
- ▶ Can be added as a module (5-10 minutes) in any existing survey

Other size estimation methods are problematic, and scale-up method has many nice properties:

- ▶ Requires a random sample of the general population, not specific contact with the hard-to-reach population
- ▶ Can be added as a module (5-10 minutes) in any existing survey
- ▶ Can estimate many target populations in a single survey

Other size estimation methods are problematic, and scale-up method has many nice properties:

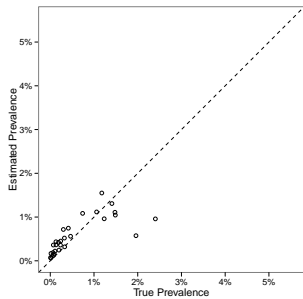
- ▶ Requires a random sample of the general population, not specific contact with the hard-to-reach population
- ▶ Can be added as a module (5-10 minutes) in any existing survey
- ▶ Can estimate many target populations in a single survey
- ▶ Can be applied at the city-level, sub-national-level, or national-level

Other size estimation methods are problematic, and scale-up method has many nice properties:

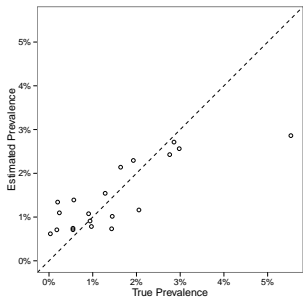
- ▶ Requires a random sample of the general population, not specific contact with the hard-to-reach population
- ▶ Can be added as a module (5-10 minutes) in any existing survey
- ▶ Can estimate many target populations in a single survey
- ▶ Can be applied at the city-level, sub-national-level, or national-level
- ▶ Statistical methods are improvable

Other size estimation methods are problematic, and scale-up method has many nice properties:

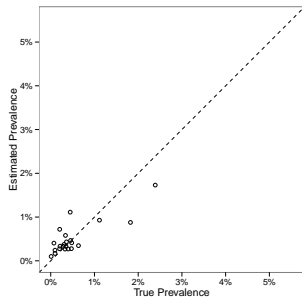
- ▶ Requires a random sample of the general population, not specific contact with the hard-to-reach population
- ▶ Can be added as a module (5-10 minutes) in any existing survey
- ▶ Can estimate many target populations in a single survey
- ▶ Can be applied at the city-level, sub-national-level, or national-level
- ▶ Statistical methods are improvable
- ▶ Partially self-validating because it uses groups of known size



(a) United States



(b) Curitiba



(c) Rwanda

Network scale-up method, basic estimator

$$\hat{N}_H = \frac{\sum_i y_{i,H}}{\sum_i \hat{d}_i} \times N$$

- ▶ \hat{N}_H : number of people in the hidden population
- ▶ $y_{i,H}$: number of people in hidden population known by person i
- ▶ \hat{d}_i : estimated number of people known by person i
- ▶ N : number of people in the population

$$\hat{d}_i = \frac{\sum_k y_{i,k}}{\sum_k N_k} \times N$$

- ▶ \hat{d}_i : estimated number of people known by person i
- ▶ $y_{i,k}$: number of people in group k known by person i
- ▶ N_k : number of people in group k
- ▶ N : number of people in the population

$$\hat{d}_i = \frac{\sum_k y_{i,k}}{\sum_k N_k} \times N$$

- ▶ \hat{d}_i : estimated number of people known by person i
- ▶ $y_{i,k}$: number of people in group k known by person i
- ▶ N_k : number of people in group k
- ▶ N : number of people in the population

There are 50,000 Nsabimanas in Rwanda and 10 million people in Rwanda. You know 2 Nsabimanas. We estimate you know:

$$\hat{d}_i = \frac{\sum_k y_{i,k}}{\sum_k N_k} \times N$$

- ▶ \hat{d}_i : estimated number of people known by person i
- ▶ $y_{i,k}$: number of people in group k known by person i
- ▶ N_k : number of people in group k
- ▶ N : number of people in the population

There are 50,000 Nsabimanas in Rwanda and 10 million people in Rwanda. You know 2 Nsabimanas. We estimate you know:

$$\hat{d}_i = \frac{2}{50,000} \times 10 \text{ million} = 400 \text{ people}$$

$$\hat{d}_i = \frac{\sum_k y_{i,k}}{\sum_k N_k} \times N$$

- ▶ \hat{d}_i : estimated number of people known by person i
- ▶ $y_{i,k}$: number of people in group k known by person i
- ▶ N_k : number of people in group k
- ▶ N : number of people in the population

There are 50,000 Nsabimanas in Rwanda; 1,000 Priests; and 10 million people in Rwanda. You know 2 Nsabimanas and 1 priest. We estimate you know:

$$\hat{d}_i = \frac{\sum_k y_{i,k}}{\sum_k N_k} \times N$$

- ▶ \hat{d}_i : estimated number of people known by person i
- ▶ $y_{i,k}$: number of people in group k known by person i
- ▶ N_k : number of people in group k
- ▶ N : number of people in the population

There are 50,000 Nsabimanas in Rwanda; 1,000 Priests; and 10 million people in Rwanda. You know 2 Nsabimanas and 1 priest. We estimate you know:

$$\hat{d}_i = \frac{2 + 1}{50,000 + 1,000} \times 10 \text{ million} \approx 600 \text{ people}$$

$$\hat{N}_H = \frac{\sum_i y_{i,H}}{\sum_i \hat{d}_i} \times N$$

- ▶ \hat{N}_H : number of people in the hidden population
- ▶ $y_{i,H}$: number of people in hidden population known by person i
- ▶ \hat{d}_i : estimated number of people known by person i
- ▶ N : number of people in the population

- ▶ Person 1 knows 2 female sex workers and 400 people
- ▶ Person 2 knows 4 female sex workers and 600 people

$$\hat{N}_H = \frac{\sum_i y_{i,H}}{\sum_i \hat{d}_i} \times N$$

- ▶ \hat{N}_H : number of people in the hidden population
- ▶ $y_{i,H}$: number of people in hidden population known by person i
- ▶ \hat{d}_i : estimated number of people known by person i
- ▶ N : number of people in the population

- ▶ Person 1 knows 2 female sex workers and 400 people
- ▶ Person 2 knows 4 female sex workers and 600 people

$$\hat{N}_H = \frac{2 + 4}{400 + 600} \times 10 \text{ million} = 60,000 \text{ people}$$

GENERALIZING THE NETWORK SCALE-UP METHOD: A NEW ESTIMATOR FOR THE SIZE OF HIDDEN POPULATIONS

*Dennis M. Feehan**
Matthew J. Salganik[†]

- ▶ We develop the generalized scale-up estimator and use it to point out possible biases in basic scale-up estimator, but requires two data collections, which is rare (but wait until lecture 23)

GENERALIZING THE NETWORK SCALE-UP METHOD: A NEW ESTIMATOR FOR THE SIZE OF HIDDEN POPULATIONS

*Dennis M. Feehan**
Matthew J. Salganik[†]

- ▶ We develop the generalized scale-up estimator and use it to point out possible biases in basic scale-up estimator, but requires two data collections, which is rare (but wait until lecture 23)
- ▶ For the purposes of this class, focus on basic scale-up estimator and the correction factors needed for it



Practice of Epidemiology

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

Dennis M. Feehan*, Aline Umubyeyi, Mary Mahy, Wolfgang Hladik, and Matthew J. Salganik

* Correspondence to Dr. Dennis M. Feehan, Department of Demography, College of Letters and Science, University of California, Berkeley, 2232 Piedmont Avenue, Berkeley, CA 94720 (e-mail: feehan@berkeley.edu).

- Are there tie definitions other than “to know” that would lead to better estimates? We did a survey experiment to compare different definitions and then combined them to produce a single estimate



Practice of Epidemiology

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

Dennis M. Feehan*, Aline Umubyeyi, Mary Mahy, Wolfgang Hladik, and Matthew J. Salganik

* Correspondence to Dr. Dennis M. Feehan, Department of Demography, College of Letters and Science, University of California, Berkeley, 2232 Piedmont Avenue, Berkeley, CA 94720 (e-mail: feehan@berkeley.edu).

- ▶ Are there tie definitions other than “to know” that would lead to better estimates? We did a survey experiment to compare different definitions and then combined them to produce a single estimate
- ▶ Notice how we decided which is better and how we combined estimates



Practice of Epidemiology

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

Dennis M. Feehan*, Aline Umubyeyi, Mary Mahy, Wolfgang Hladik, and Matthew J. Salganik

* Correspondence to Dr. Dennis M. Feehan, Department of Demography, College of Letters and Science, University of California, Berkeley, 2232 Piedmont Avenue, Berkeley, CA 94720 (e-mail: feehan@berkeley.edu).

- ▶ Are there tie definitions other than “to know” that would lead to better estimates? We did a survey experiment to compare different definitions and then combined them to produce a single estimate
- ▶ Notice how we decided which is better and how we combined estimates
- ▶ Notice how we dealt with the many things we did not know with sensitivity analysis.

Enjoy the reading