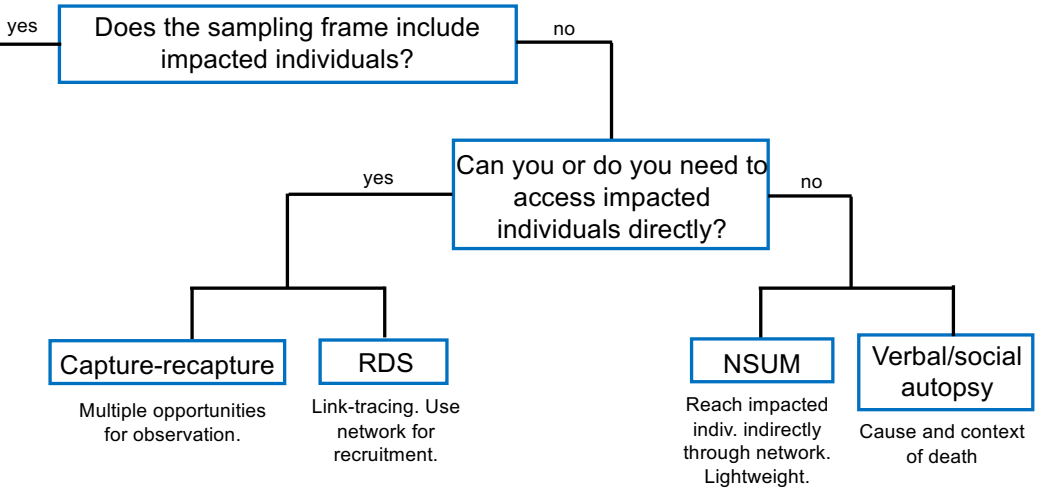
Network Scale-up and indirect methods for group size estimation

Tyler H. McCormick

Best Practices in Assessing Mortality and
Significant Morbidity Following Large-Scale Disasters

February 11, 2020

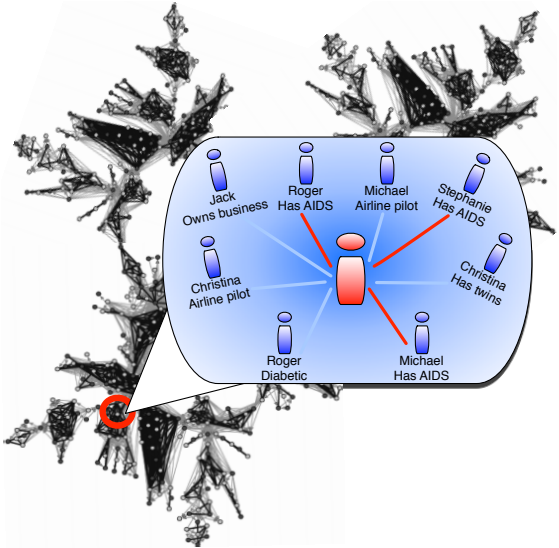
Overview/summary



Network scale-up method

- Uses respondents' network to access hard-to-reach groups
- Key distinctions:
  - Respondents come from a probability sample (*not* the hard-to-reach group)
  - Reach members of hard-to-reach (HTR) groups *indirectly* through respondents
- Assumption that peoples' social networks are, on average, representative of the general population in which you live and move

How many people do you know who —— ?



Network scale-up method

Three Steps:

1. Estimate the personal network size of the members of a probability sample of a total population.

2. Determine how many people in each personal network are in the population of interest.

Network scale-up method

Three Steps:

1. Estimate the personal network size of the members of a probability sample of a total population.

2. Determine how many people in each personal network are in the population of interest.

   Example: If a sample of respondents report knowing 300 people on average, two of whom were injured in an earthquake, we estimate that 2/300th of the general population was injured in the earthquake.

Network scale-up method

Three Steps:

1. Estimate the personal network size of the members of a probability sample of a total population.

2. Determine how many people in each personal network are in the population of interest.

   Example: If a sample of respondents report knowing 300 people on average, two of whom were injured in an earthquake, we estimate that 2/300th of the general population was injured in the earthquake.

3. Averaging proportion in group of interest known by a respondent over many respondents.

Network scale-up method

- 2 types of questions:
  1. “Reference” questions to estimate degree
  2. Hard-to-reach (HTR) populations
- Improve accuracy by combining responses across respondents

$$\widehat{HTR} = \frac{\sum_i \#known\_in\_HTR_i}{\sum_i estimated\_degree_i} \times N$$

Network scale-up features

- Does not access HTR group directly
- Sampling only first-order network dependent
- Estimate size of multiple groups with the same sample
- Large “indirect” sample size
  - Sample 1000 respondents who know on average 750 alters

Three issues

1. Recall problems
2. Respondent systematically knows more (or fewer) members of a specific group than expected under random mixing (barrier effects, nonrandom mixing)
3. Respondent knows someone in a specific group but is not aware that the person is actually in that group (transmission errors)

Lessons from my experiences

- Preparation: Methods require high quality input data
  - NSUM uses data on demographics of reference groups
  - Symptoms of Ebola in children?

Lessons from my experiences

- Preparation: Methods require high quality input data
  - NSUM uses data on demographics of reference groups
  - Symptoms of Ebola in children?
- It's a data science problem
  - Prioritize **both** innovations in analytical methods **and** data collection
  - OpenVA

Lessons from my experiences

- Preparation: Methods require high quality input data
  - NSUM uses data on demographics of reference groups
  - Symptoms of Ebola in children?
- It's a data science problem
  - Prioritize **both** innovations in analytical methods **and** data collection
  - OpenVA
- Measured use of link-tracing designs

Uncertainty: a persistent challenge

Uncertainty: a persistent challenge

- Your estimate is probably too small.

Uncertainty: a persistent challenge

- Your estimate is probably too small.
- What if an accurate estimate is too large to be useful?

Uncertainty: a persistent challenge

- Your estimate is probably too small.
- What if an accurate estimate is too large to be useful?
- Communicating estimated uncertainty effectively

Uncertainty: a persistent challenge

- Your estimate is probably too small.
- What if an accurate estimate is too large to be useful?
- Communicating estimated uncertainty effectively
- Engagement; context and policy decisions at stake

Uncertainty: a persistent challenge

- Your estimate is probably too small.
- What if an accurate estimate is too large to be useful?
- Communicating estimated uncertainty effectively
- Engagement; context and policy decisions at stake

More sophisticated methods exacerbate this challenge.

References

Bernard, H. R., Johnsen, E. C., Killworth, P. D., McCarty, C., Shelley, G. A. and Robinson, S. (1990), 'Comparing four different methods for measuring personal social networks', *Social Networks* **12**, 179–215.

Bernard, R. H., Johnsen, E., Killworth, P. and Robinson, S. (1989), Estimating the size of an average personal network and of an event subpopulation, *in* M. Kochen, ed., 'The Small World', Ablex Press, New Jersey, pp. 159–175.

Bernard, R. H., Johnsen, E., Killworth, P. 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.

References

Feehan, D. M., Mahy, M. and Salganik, M. J. (2017), 'The network survival method for estimating adult mortality: Evidence from a survey experiment in rwanda', *Demography* **54**(4), 1503–1528.

Feehan, D. M. and Salganik, M. J. (2016), 'Generalizing the network scale-up method', *Sociological Methodology* **46**(1), 153–186.
URL: *http://dx.doi.org/10.1177/0081175016665425*

Feehan, D. M., Umubyeyi, A., Mahy, M., Hladik, W. and Salganik, M. J. (2016), 'Quantity versus quality: A survey experiment to improve the network scale-up method', *American journal of epidemiology* **183**(8), 747–757.

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.

References

Killworth, P. D. and Bernard, H. R. (1978), 'The reverse small-world experiment', *Social Networks* **1**(2), 159–192.

Killworth, P. D., Bernard, H. R. and McCarty, C. (1984), 'Measuring patterns of acquaintanceship', *Current Anthropology* **23**, 318–397.

Killworth, P. D., McCarty, C., Bernard, H. R., Johnsen, E. C., Domini, J. and Shelley, G. A. (2003), 'Two interpretations of reports of knowledge of subpopulation sizes', *Social Networks* **25**, 141–160.

Killworth, P., Johnsen, E., McCarty, C., Shelley, G. and Bernard, H. (1998), 'A social network approach to estimating seroprevalence in the united states', *Social Networks* **20**(1), 23–50.

References

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**, 28–39.

McCormick, T. H. and Zheng, T. (2007), Adjusting for recall bias in "How many X's do you know?" surveys, *in* 'Proceedings of the Joint Statistical Meetings'.

McCormick, T. H. and Zheng, T. (2012), 'Latent demographic profile estimation in hard-to-reach groups', *The annals of applied statistics* **6**(4), 1795.

McCormick, T. H. and Zheng, T. (2015), 'Latent surface models for networks using aggregated relational data', *Journal of the American Statistical Association* **110**(512), 1684–1695.

References

Salganik, M., Fazito, D., Bertoni, N., Abdo, A., Mello, M. and Bastos, F. (2011), '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. (2011), 'The game of contacts: Estimating the social visibility of groups', *Social Networks* **33**(1), 70–78.

Scutelniciuc, O. (2012), 'Network scale-up method experiences: Republic of kazakhstan', *Consultation on estimating population sizes through household surveys: Successes and challenges (New York, NY)* .

Issue 1: Recall issues

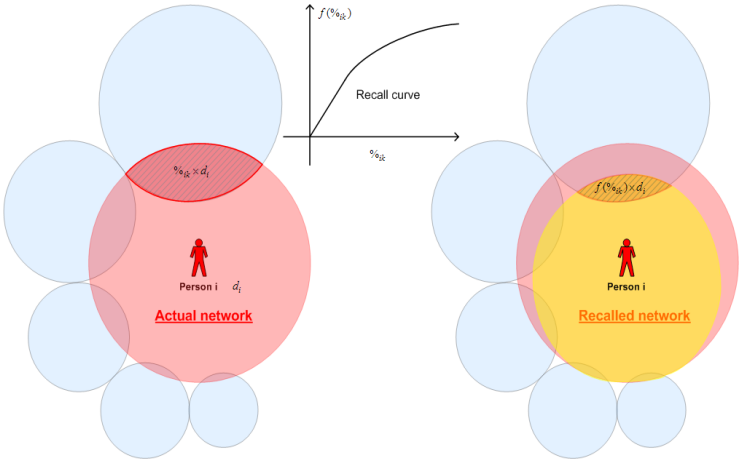


Figure: The Calibration Curve.

Solution(s)

- Use small groups
- Calibrate with known sizes

Issue 2: Barrier effects

How many people named Christina do you know?

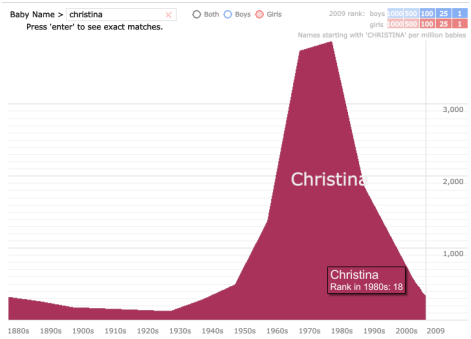
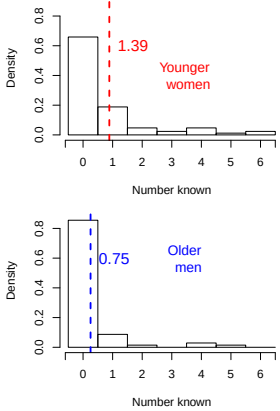
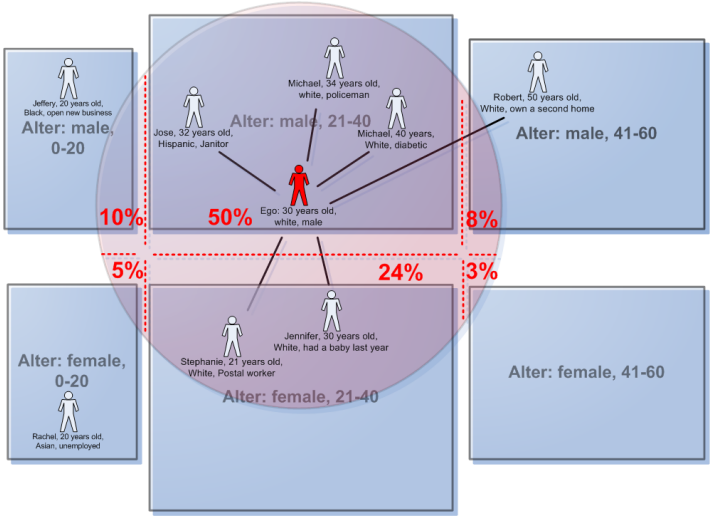


Figure: NameVoyager from babynamewizard.com.



Issue 2: Barrier effects

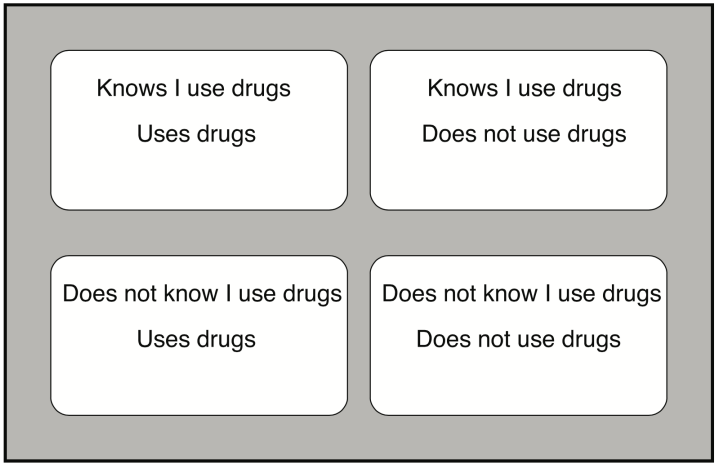


Solution(s)

- Calibrate using group demographics
- Scaled-down condition

Issue 3: Transmission errors

M.J. Salganik et al. / Social Networks 33 (2011) 70–78



Solution(s)

- External data