Surveying the Homeless: A Statistical Perspective

Mark S. Handcock

Department of Statistics
University of California - Los Angeles

Conference on Transdisciplinary Homelessness Research, UCLA, May 24, 2018

Conventional Survey Sampling

Stylized description

- Choose a *population* of interest and a population characteristic of interest μ
- Determine the *sampling frame*: $i = 1, \ldots, N$ sample units.
- Choose variables to measure on them: *outcome variables* $y_i, i = 1, \ldots, N$.

Population

Conventional Survey Sampling

Stylized description

- Choose a *population* of interest and a population characteristic of interest μ
- Determine the *sampling frame*: $i = 1, \ldots, N$ sample units.
- Choose variables to measure on them: *outcome variables* $y_i, i = 1, \ldots, N$.
- Choose a *sampling design*:
  e.g., simple random sampling, stratified sampling on covariates,
  stratified sampling on y
- Choose a sample of units $i = 1, \ldots, n$ and collect data on the sampled units

Sampled people (green)

Conventional Survey Sampling

Stylized description

- Choose a *population* of interest and a population characteristic of interest μ
- Determine the *sampling frame*: $i = 1, \ldots, N$ sample units.
- Choose variables to measure on them: *outcome variables* $y_i, i = 1, \ldots, N$.
- Choose a *sampling design*:
  e.g., simple random sampling, stratified sampling on covariates,
  stratified sampling on y
- Choose a sample of units $i = 1, \ldots, n$ and collect data on the sampled units
- Estimate the population characteristics of interest based on the sample

In the context of the homeless

- The sampling frame could be based on:
  - service providers
  - drop-in centers
  - sheltered populations
  - permanent supportive housing (PSH)
- These can be dealt with using conventional methods
- The usual issues apply:
  - person-level non-response
  - item-level missing data
  - drop out in longitudinal studies

What about everything and everybody else?

- unsheltered
- vehicle dwelling

Limitation: No practical conventional sampling frame.

Counting the Homeless: Point in Time Survey

- Each Continuum of Care (COC) is required by HUD to submit a point in time (PIT) count
- Methodology varies by COC
- Most often done using a spatial sampling frame (e.g., census tracts)
  - *Street Count* (all census tracts): Estimate of the unsheltered population.
  - *Shelter Count* (from HMIS frame): Estimates the population in emergency shelters, transitional housing, safe havens and vouchered motels/hotels.
  - *Youth Count* (sample census tracts): Estimate of homeless youth.
  - *Demographic Survey* (sample census tracts): Estimate the demographics of the unsheltered

Assessment of the Point in Time Survey

Daniel Flaming and Patrick Burns
Who Counts? Assessing Accuracy of the Homeless Count
(November 21, 2017).
`https://economicrt.org/publication/who-counts/`

- The uncertainty in the estimates from year-to-year make comparisons across years unreliable.
- The methodology does not adjust for known sources of *measurement error*.
  - The Count likely under estimates in important and measurable ways.
- Demographic survey based on quasi-random selection of unsheltered
- Count is inconsistent with patterns in General Relief caseload and School Records of Homeless Students
- Fundamentally flawed statements about the accuracy of the estimates

Recommendations for the Point in Time Survey

Daniel Flaming and Patrick Burns
Who Counts? Assessing Accuracy of the Homeless Count
(November 21, 2017).
`https://economicrt.org/publication/who-counts/`

- Make year-to-year comparability in population estimates a primary goal
- Enhance the sampling frame: hotspot and other census tracts.
  - "homelessness causes placelessness"
- Where possible, integrate the demographic survey into the street count
- Use "decoy" quality assurance and adjustment
- Use follow-up survey at providers and "capture-recapture" method
- Use model-based statistical methods to improve the estimates
- Use supplementary data source to "post-stratify" estimates

Other methods?

Suppose:

- The population is joined by an informal social network of relationships.
- Cell phone app based surveys: becoming practical
- Adaptive Network Sampling
- Combining spatial sampling and link-tracing sampling

Networked Population

Adaptive Network Sampling

Suppose:

- The population is joined by an informal social network of relationships.
- Researchers can access some members of the population.

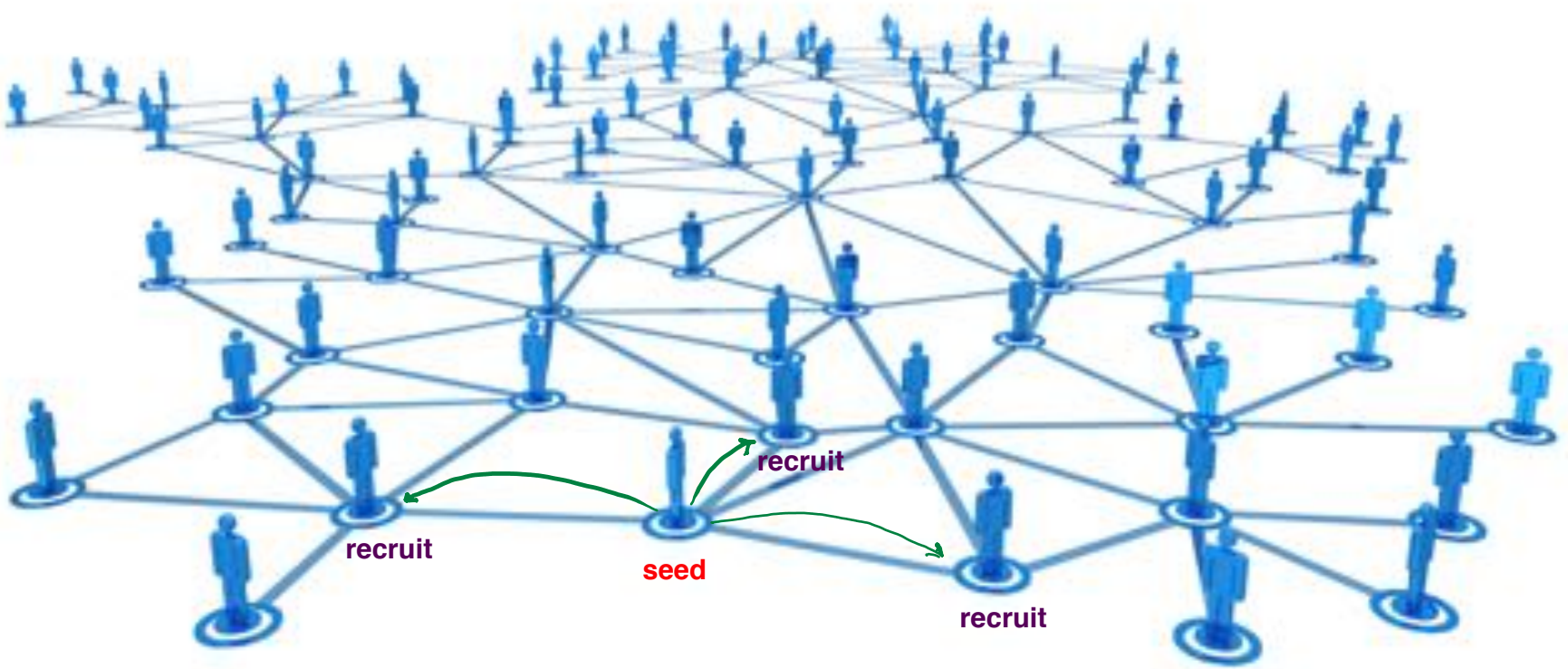
Sampling design:

- Begin with an initial sample (the *seeds*)
- Expand sample by the researchers sampling those tied to those already in the sample.
  a process called *link tracing*.

Start with a *seed* person

Contact other people via the seed's social network

Contact other people via the seed's social network



Contact other people via the seed's social network

Link-Tracing Sampling:

An effective way to collect data, but estimation is problematic

Challenges:

- Sampling depends on (typically) partially-observed network data
- Seed Dependence: final sample depends on sampling mechanism of seeds
- Privacy/Confidentiality: some homeless prefer to stay "hidden"
- Estimation: The sample and sampling probabilities depend on the unknown network

Adaptive Network Sampling

Strengths:

- Exploits information in the network of relationships
- Network structure used to improve the design
- Increases the range of possible designs
- Adjusts for discovered features in the population
- Leads to increased efficiency of sampling

A peculiar case: Respondent-Driven Sampling

- *Sampling design*: Require respondents to choose from among their social circle rather than the researcher chooses.
- *Seed Dependence*: follow only a few links from each sampled
- *Privacy*: *respondent-driven:* respondents distribute uniquely identified coupons. no names.
- *Link-tracing*: none by researchers, done by respondents.
- *Estimation*: Challenging to get valid estimates

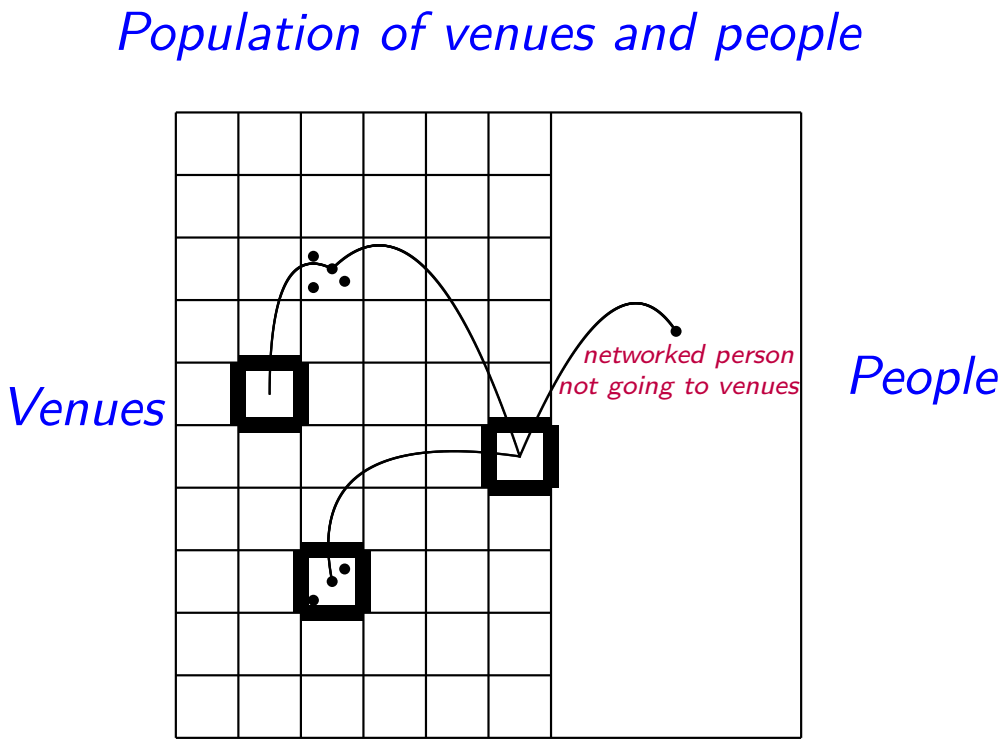
- Effective at obtaining large varied samples in many populations.
- Widely used: over 100+ studies, in over 30 countries. Often HIV-risk populations.

Other Designs: Combining spatial sampling and link-tracing sampling

Martin Félix-Medina and Steve Thompson (2004)

- Identify a list of sites/venues/buildings that people can be found
- Sample these venues
- Link-trace out from these to the broader population

Combining spatial sampling and link-tracing sampling



Adapted, by permission, from Martin H. Félix-Medina and Steve K. Thompson (2004). Combining cluster sampling and link-tracing sampling to estimate the size of hidden populations. *Journal of Official Statistics*, 20, 19-38.

Discussion: Surveying the homeless

- Traditional approaches are useful but challenged
- Modern survey methods should be applied to improve accuracy
- Social network based ideas are promising
- Typically, RDS not advisable if alternatives available.

- Improving surveying
  - privatized network sampling: RDS, but collects more information on the network while preserving the privacy
  - Surveys using a natural byproduct of digital technology
  - Using call-back surveys to collect more information