Share

Non-Probability Sampling, Part 1

Brady T. West

Research Associate Professor, Survey Research Center,
Institute for Social Research

© 2018 The Regents of the University of Michigan
Except where otherwise noted, this work is licensed under
http://creativecommons.org/licenses/by-nc/3.0/

Share

Lecture Overview

- What defines a non-probability sample?
- **Common examples** of non-probability samples
- Intro to **two common population inference methods**
- Example of non-probability sampling: **Twitter data**

Logo from Twitter

Non-probability sampling namely drawing a sample from Twitter.

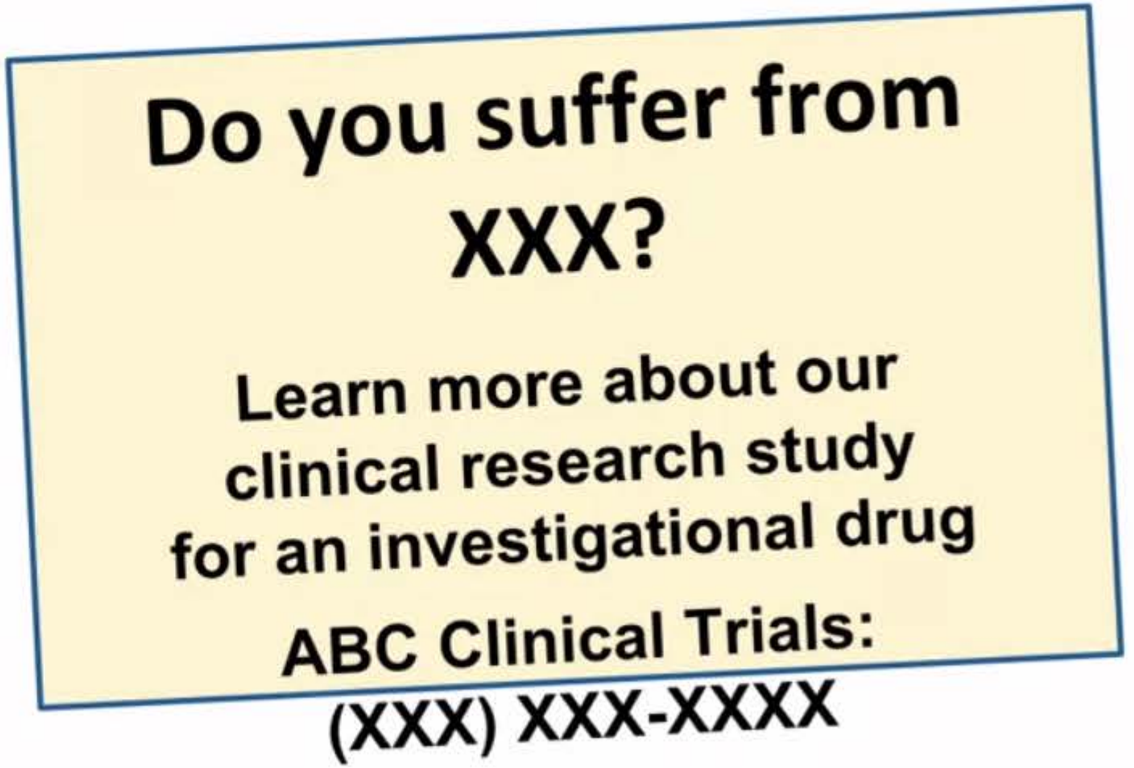
What Are Non-Probability Samples?

- **Features of Non-probability samples:**
  - Probabilities of selection **can't be determined** for sampled units
  - **No random selection** of individual units
  - Sample divided into groups (strata) or clusters, but **clusters not randomly sampled** in earlier stage
  - Data collection often very **cheap** ¢ relative to probability sampling

much less expensive than probability samples.

Examples of Non-Probability Samples

- Studies of **volunteers**



and then you're given a phone number or an email and then you call in and say,

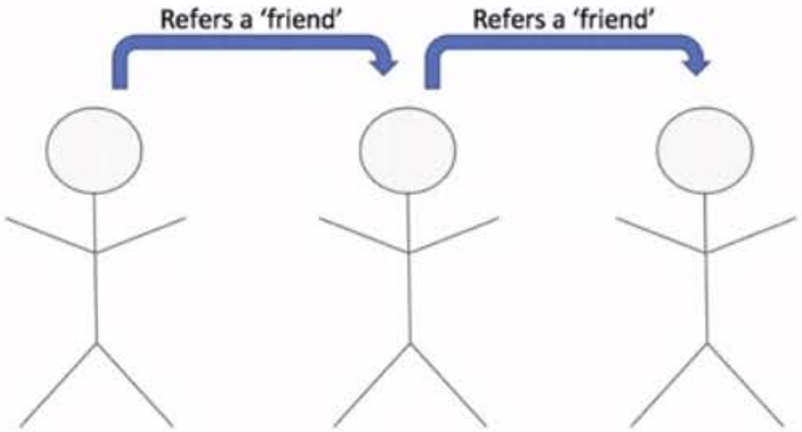
Examples of Non-Probability Samples

- Studies of **volunteers** (e.g., clinical trials)
- **Opt-in** / Intercept web surveys

So, when you're on a website and you see an invitation

Examples of Non-Probability Samples

- Studies of **volunteers** (e.g., clinical trials)
- **Opt-in** / Intercept web surveys
- **Snowball** samples



A third example is Snowball sampling and again we talked about this a little bit earlier,

Examples of Non-Probability Samples

- Studies of **volunteers** (e.g., clinical trials)
- **Opt-in** / Intercept web surveys
- **Snowball** samples
  (e.g., word-of-mouth data collection)
- **Convenience** samples

All Students in Psych 101

Classroom by velkr0 CC-BY 2.0

So, you see here a common example in academic settings is when

Examples of Non-Probability Samples

- Studies of **volunteers** (e.g., clinical trials)
- **Opt-in** / Intercept web surveys
- **Snowball** samples
  (e.g., word-of-mouth data collection)
- **Convenience** samples
  (e.g., all students in Psych 101)
- **Quota** samples

Recruit 1,000 males and 1,000 females in any way

certain targets that you wish to hit in terms of your sample size,

Share

Examples of Non-Probability Samples

Common Feature:

Probabilities of selection

cannot be determined

a priori!

probabilities of selection cannot be determined a

So What Is The Problem?

- Non-probability sample → **no statistical basis for making inference** about larger population from which sample selected
- **Knowing probabilities of selection**
  (in addition to population strata and randomly sampled clusters)
  → **can estimate features of sampling distribution**
  if were to take many random samples using same design

absolutely crucial for making inference about a larger population,

So What Is The Problem?

- Sampled units **not selected at random** → strong risk of **sampling bias**
  (e.g., people actually interested in visiting particular web site)

Across the board, there's no selection at random occurring here and what this

So What Is The Problem?

- Sampled units **not selected at random** → strong risk of **sampling bias**
  (e.g., people actually interested in visiting particular web site)
- Sampled units **not generally representative** of larger target population of interest

generally not representative of the larger target population of interests.

Share

So What Is The Problem?

- Sampled units **not selected at random** → strong risk of **sampling bias**
  (e.g., people actually interested in visiting particular web site)
- Sampled units **not generally representative** of larger target population of interest
- **"Big data"** (e.g., information from millions of tweets) **often from non-probability samples ~ be careful!**

Logo from Twitter

there's no probability sampling mechanism that gives

An undergraduate student in Psychology 101 is tasked with collecting data from a sample of students at that college on recent experiences with depression. To collect the data, the student stands in the middle of campus at Noon and asks 30 different people walking by whether they have recently had an experience with depression. What are the problems with the sample measured by this student? Please select all responses that apply.

- [x] There is no guarantee that all respondents are students (lack of representation).

Correct

- [x] The 30 people have not been selected at random (risk of sampling bias).

Correct

- [x] People walking on the campus at noon may be quite different from people walking on the campus at other times of day (lack of representation).

Correct

- [ ] The sample size is too large.

Un-selected is correct

- [x] The probabilities of selection for the sampled persons cannot be determined.

Correct

Share

So What Can We Do?

- Many data sets arise from non-probability samples
  ... can we say anything about a larger population?

- **Two possible approaches**:
  - Pseudo-Randomization
  - Calibration

For technical "deep dive" into estimation approaches: Elliott and Valliant (2017, Statistical Science)

So, you wait your Non-probability sample to