

Studies of vulnerable populations get a 'bootstrapped' boost from statisticians

December 13 2016, by James Urton



Credit: University of Washington

A hallmark of good government is policies which lift up vulnerable or neglected populations. But crafting effective policy requires sound knowledge of vulnerable groups. And that is a daunting task since these

populations—which include undocumented immigrants, homeless people or drug users—are usually hidden in the margins thanks to cultural taboos, murky legal status or simple neglect from society.

"These are not groups where there's a directory you can go to and look up a random sample," said Adrian Raftery, a professor of statistics and sociology at the University of Washington. "That makes it very difficult to make inferences or draw conclusions about these 'hidden' groups."

Since these groups are hard to identify and reach, researchers like Raftery can struggle to make accurate inferences about them, determine their needs and find effective ways to reach them. And government policies to help vulnerable groups run a high risk of failing.

Sociologists once hoped that an approach called respondent-driven sampling—or RDS—would help them make reliable inferences about hard-to-reach groups. But subsequent analyses cast doubt on the efficacy of RDS studies.

In a paper published online Dec. 7 in the *Proceedings of the National Academy of Sciences*, Raftery and his team report how a [statistical approach](#) called "tree bootstrapping" can accurately assess uncertainty in RDS studies. That would put RDS on firm ground as one of the few methods to study vulnerable groups.

First described in 1997, respondent-driven sampling in studies works around the "problem" of recruitment. Normally, social scientists try to recruit study subjects at random from their target population. But this is not possible when social or legal issues act as barriers between researchers and subjects.

"This is an underlying problem when you're trying to access and make inferences about populations that are hard to access, like [drug users](#),"

said Raftery.

With the RDS method, researchers can start with a handful of participants, and use them to recruit additional participants using existing social connections.

"You can set up a storefront and find a few people in the hard-to-reach population: You interview them, collect data and give them vouchers to give to their friends—who can come in as well," said Raftery. "It was immediately useful for accessing these populations."

To date, over 460 RDS studies of vulnerable populations have been conducted. But researchers have shown that the standard estimates of uncertainty are wrong, making it hard to use RDS in a valid way. It turns out that the inferences that researchers drew about these populations were biased by the fact that their study subjects weren't chosen at random.

"RDS is kind of like trying to describe an elephant when you're blindfolded and only get to touch one part of the elephant," said Raftery. "You can get a lot of data about that one part of the elephant, but we—the researchers—didn't have the proper methods to draw firm, scientifically sound conclusions about the elephant as a whole."

Raftery and his team started looking for methods to assess the uncertainty in RDS studies. They quickly settled on bootstrapping, a statistical approach used to assess uncertainty in estimates based on a random sample. In traditional bootstrapping, researchers take an existing dataset—for example, condom use among 1,000 HIV-positive men—and randomly resample a new dataset, calculating condom use in the new dataset. They then do this many times, yielding a distribution of values of condom use that reflects the uncertainty in the original sample.

The team modified bootstrapping for RDS datasets. But instead of bootstrapping data on individuals, they bootstrapped data about the connections among individuals.

To see if this "tree bootstrapping" could attach certainty to conclusions from RDS datasets, they turned to two large, publicly available datasets. One was a multiyear survey of health and achievement among more than 90,000 adolescents, while the other was a survey of social contacts and sexual and drug habits among about 5,400 heterosexual adults. Neither dataset was collected using the RDS method. But since both datasets included information about the social contacts among subjects, the researchers could modify them to "simulate" data from a RDS study.

By tree bootstrapping, Raftery's team found that they could get much better statements of scientific certainty about their conclusions from these RDS-like studies. They then applied their method to a third dataset—a RDS study of intravenous drug users in Ukraine. Again, Raftery's team found that they could draw firm conclusions.

"Previously, RDS might give an estimate of 20 percent of drug users in an area being HIV positive, but little idea how accurate this would be. Now you can say with confidence that at least 10 percent are," said Raftery. "That's something firm you can say. And that can form the basis of a policy to respond, as well as additional studies of these groups."

With tree bootstrapping, Raftery believes researchers can draw more certain, less variable conclusions from RDS studies. He wants other groups to examine and use tree bootstrapping on both existing RDS datasets and future RDS studies.

"I hope this paper will help put RDS on a firm basis, and tell us what we can and can't conclude from RDS studies," said Raftery.

More information: Aaron J. Baraff et al, Estimating uncertainty in respondent-driven sampling using a tree bootstrap method, *Proceedings of the National Academy of Sciences* (2016). [DOI: 10.1073/pnas.1617258113](https://doi.org/10.1073/pnas.1617258113)

Provided by University of Washington

Citation: Studies of vulnerable populations get a 'bootstrapped' boost from statisticians (2016, December 13) retrieved 8 May 2024 from <https://phys.org/news/2016-12-vulnerable-populations-bootstrapped-boost-statisticians.html>

This document is subject to copyright. Apart from any fair dealing for the purpose of private study or research, no part may be reproduced without the written permission. The content is provided for information purposes only.