American Journal of Epidemiology, 2025, **00**, 1–10
https://doi.org/10.1093/aje/kwae342
Advance access publication date September 4, 2024
Original Contribution

Innovating a community-driven enumeration and needs assessment of people experiencing homelessness: a network sampling approach for the HUD-mandated point-in-time count

Zack W. Almquist*[,1], Ihsan Kahveci[2], Mary Ashley Hazel[3], Owen Kajfasz[4], Janelle Rothfolk[4], Claire Guilmette[4], M. C. Anderson[5], Larisa Ozeryansky[6], Amy Hagopian[7]

[1]Departments of Sociology and Statistics, University of Washington, Seattle, WA 98195, United States
[2]Department of Sociology, University of Washington, Seattle, WA 98195, United States
[3]Francis I. Proctor Foundation, University of California, San Francisco, San Francisco, CA 94143, United States
[4]King County Regional Homelessness Authority, Seattle, WA 98104, United States
[5]Earth System Science, Stanford University, Palo Alto, CA 94305, United States
[6]Interdisciplinary Studies, University of Washington, Seattle, WA 98195, United States
[7]Departments of Health Systems and Population Health and Global Health University of Washington, Seattle, WA 98195, United States
*Corresponding author: Dr. Zack W. Almquist, Department of Sociology, University of Washington, 211 Savery Hall Mail Code 353340, Seattle, WA 98195 (zalmquist@uw.edu)

Abstract

To enumerate people experiencing homelessness in the United States, the federal Department of Housing and Urban Development (HUD) mandates its designated local jurisdictions regularly conduct a crude census of this population. This Point-in-Time (PIT) body count, typically conducted on a January night by volunteers with flashlights and clipboards, is often followed by interviews with a separate convenience sample. Here, we propose employing a network-based (peer-referral) respondent-driven sampling (RDS) method to generate a representative sample of unsheltered people, accompanied by a novel method to generate a statistical estimate of the number of unsheltered people in the jurisdiction. First, we develop a power analysis for the sample size of our RDS survey to count unsheltered people experiencing homelessness. Then, we conducted 3 large-scale population-representative samples in King County, WA (Seattle metro) in 2022, 2023, and 2024. We describe the data collection and the application of our new method, comparing the 2020 PIT count (the last visual PIT count performed in King County) to the new method of 2022 and 2024 PIT counts. We conclude with a discussion and future directions.

This article is part of a Special Collection on Methods in Social Epidemiology.

Key words: homelessness; people experiencing homelessness; social networks; hard-to-reach populations; sample data; networks; social networks; respondent-driven sampling.

Introduction

The US Department of Housing and Urban Development (HUD)'s 2023 Annual Homeless Assessment Report suggests 653 100 people in the United States are without homes, a 12% increase from 2022. The national count requires each HUD "Continuum of Care" (COC) jurisdiction to regularly conduct a local "Point-in-Time" (PIT) count. Two tallies are made: (1) emergency shelter report from administrative records and (2) count of unsheltered people living in tents, vehicles, and other arrangements unsuitable for human habitation. This HUD-mandated, biennial unsheltered count is generally followed by a demographic and needs survey of unhoused individuals over 2 to 4 weeks. Volunteers with flashlights and clipboards typically conduct unsheltered PIT counts on a given night in January, and the demographic and needs survey is often directed by a large survey firm (see for example https://www.appliedsurveyresearch.org/).

The PIT count has dominated how people experiencing homelessness in the United States are enumerated, even though there is a long history in the literature of proposed methods for counting the unhoused population beyond what is currently done for the Congressional report.[1,2] The PIT count is also one of few studies to collect demographic and needs-based assessments of people who are disconnected from social services systems. Thus, unsheltered PIT counts fill 2 important roles: It is (1) a systematic count of those not currently using shelter services on a single night and (2) provides insight into characteristics and needs of this population. Importantly, because the unsheltered PIT count characterizes the population size and needs of people experiencing homelessness who may not be accessing services (eg, emergency shelters, soup kitchens/food pantries, emergency departments), these people are likely to be systematically different from those tracked using services in the homeless care system.

Policymakers and policy-change advocates for the human right to housing (eg, the UN, see David et al[3] for a history) rely on accurate counts of homelessness to both decide where to focus their energies and measure their progress. Recently, Tsai and

Received: May 10, 2024. **Accepted:** August 30, 2024

Alarcón in the *American Journal of Public Health*[4] and others have critiqued the traditional 1-night PIT methodology used by HUD-designated administrative regions on both methodological and cost grounds. All this leads to a fundamental research question: *how do we provide a (quasi)-probability sample for unsheltered populations to provide both a count and a demographic and needs-based survey acceptable to the studied population, care providers, and the US government?*

Here, we describe an innovative strategy for conducting a quasi-probability sample of unsheltered people in King County, WA, to obtain a cost-effective and accurate count. This strategy was combined with a rigorous population-representative sample to produce a high-quality understanding of demographics and service needs.

In 2019, we worked in Davidson County (Nashville metro), TN—Davidson is a large county (800 000 people) including Nashville, a large urban center—to test an approach outside the bounds of the traditional 1-night count used by HUD Communities of Care (CoCs) everywhere. There, we collected a novel network-based sample of people experiencing homelessness that would allow for the simulation of a complete social network of people experiencing homelessness. Specifically, this allowed us to demonstrate the efficacy of a new approach to counting unsheltered people experiencing homelessness through a network-based (peer-referral) respondent-driven sampling (RDS) approach. Details are in the online appendix (Section A). This provided a proof of concept and power analysis (sample-size selection) that we used in the full study conducted in King County, WA.

In 2022, the University of Washington was invited to work with the King County Regional Homelessness Authority, the CoC for one of the largest homeless populations in the United States, to co-create a new enumeration method. In King County, we employed a network-based (peer-referral) respondent-driven sampling method to generate our representative sample, which we combined with a novel statistical estimator to count unsheltered people. After that successful collaboration, we piloted additional innovations in King County in 2023 and conducted an official PIT count again in early 2024. Respondent-driven sampling should be appealing to homelessness enumerators, as it has been endorsed by the CDC, NIH, and World Health Organization. The method also enjoys endorsement from ethical review boards because participants opt in.[5]

Here, we describe the sampling design and statistical methods we used to construct a representative sample, demographic and needs survey, and count of unsheltered people in King County, built jointly between the University of Washington and the King County Regional Homelessness Authority. We are aware of only one other US city, San Francisco, where a team has also begun experimenting with representative sampling methods for people experiencing homelessness,[6] although that work was conducted by a third party external to the CoC infrastructure.

Methods

Respondent-driven sampling

To generate our representative sample of King County's unsheltered population, we employed respondent-driven sampling (RDS)—a peer recruitment and social-network-based approach. This method uses multiple peer-to-peer recruitment waves to approximate random sampling in hard-to-reach populations.[7] A sociologist introduced RDS in the late 1990s,[8] and it has become a popular technique to survey hard-to-reach populations in public health.[9]

The theory supporting RDS requires that (1) respondents be selected from a social network of existing target population members; (2) network degree (sometimes referred to, imprecisely, as personal network, connectivity, or a number of connections that ego (focal node) has to other nodes) is measured accurately; and (3) the target population is well connected within their community.[10,11] While Fellows[12] found that network degrees often have significant measurement errors in RDS studies, he also found that most RDS estimators remain consistent under an imperfect measurement model (although with increases in estimator variance).

We ran 3 RDS surveys to generate a representative sample of unsheltered people experiencing homelessness. The first survey was fielded from March 9 to April 6, 2022 (24 days); the second was fielded from April 24 to June 1, 2023 (38 days); and the third was fielded from January 22 to February 2 (11 days). There were large gains to the data collection process over time, allowing us to collect more responses in a shorter time period by the third round of data collection, largely due to increased hubs, volunteers, and improved software. See Figure 1 to demonstrate the longest tree observed in our 2024 RDS study, along with a count of the wave number observed (waves are defined as the length of the referral chain).

Respondent-driven sampling typically employs incentives for surveying and referrals. For our RDS implementation in 2022, we provided a $25 Visa gift card (physical card) to each survey respondent, but we did not provide an incentive for those who successfully distributed their 3 recruitment coupons. In 2023 and 2024, we provided a smaller $20 Visa gift card (physical card) for each survey respondent, along with 3 $5 Visa coupons (digital cards) sent by email or phone for those who successfully referred their similarly situated friends to the study. Again, we provided 3 coupons per respondent. In all 3 years, we provided bus tickets with coupons for hubs in difficult-to-reach locations.

Similar to the visual PIT strategy, we employed volunteers to conduct the survey. This is advantageous as one of the important aspects of the classic unsheltered PIT is that it functions as a large community-building event for local homelessness care organizations. In 2022, our survey was conducted by a combination of King County Regional Homelessness Authority personnel, local volunteers, and paid volunteers from the local Lived Experience Coalition (LEC), which describes themselves as a "largely BIPOC led" activist organization (see https://wearelec.org/) with strong ties to the community of people experiencing homelessness. In 2023, we employed largely volunteers from the University of Washington community, and then in 2024, we again employed King County Regional Homelessness Authority (KCRHA) employees, community volunteers, and paid volunteers from the LEC. All volunteers were trained in sensitivity, harm reduction, software, and survey methods. In 2022, we had approximately 50 volunteers (this first time included 3-4 surveyors from the Lived Experience Collision who were paid and one KCRHA staff at each hub); in 2023, we had approximately 110 volunteers from the UW community (90 undergraduates and 20 graduate students and faculty); and in 2024, we had approximately 111 community volunteers. Note that in 2020, the classic visual census took about 1000 volunteers to complete. In 2023 and 2024, each hub had 3 people working 4-hour shifts, 4-6 days a week, depending on hub operating hours, 2 survey takers, and 1 hub manager who fielded issues and managed gift cards and coupon printing.

Software and coupon management

In 2022, all data were collected on paper with handwritten coupon codes (tracking referrals). The KCRHA team then transcribed the

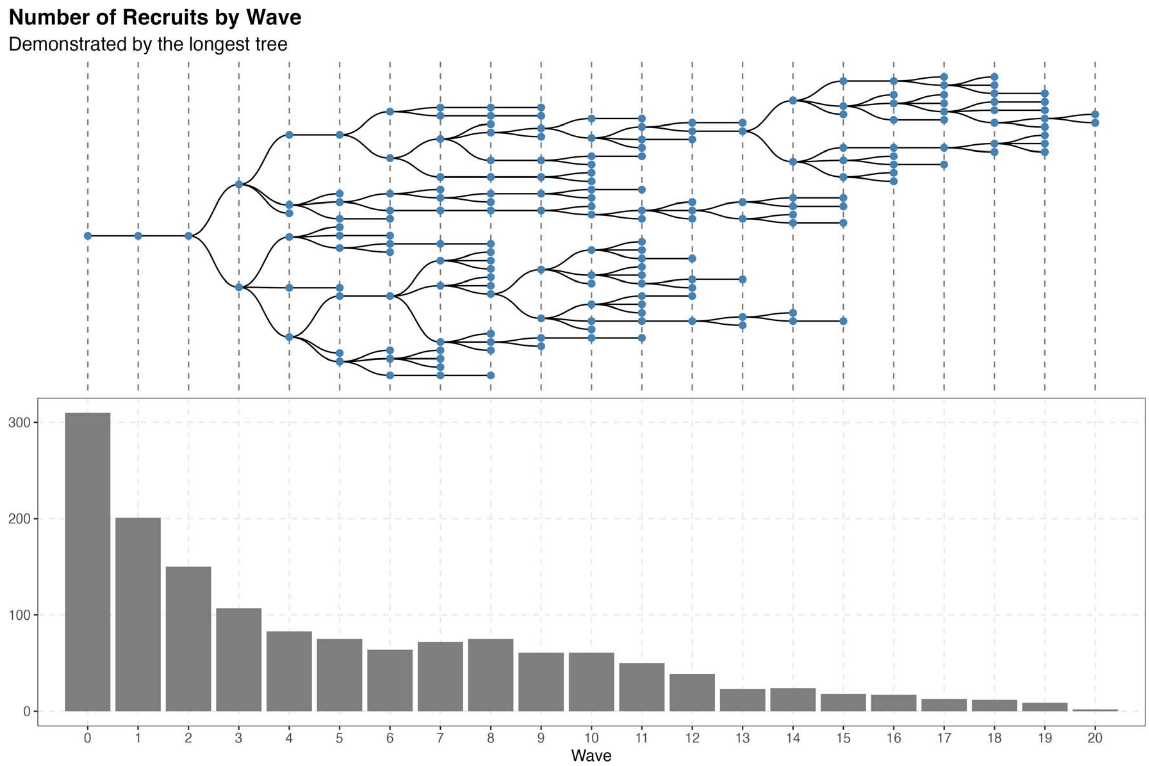


Figure 1. Plot of the longest chain observed in our 2024 RDS study mapped on top of the histogram of the count of chains per wave.

data. In 2023, we pioneered using the custom app based on Google Power App software for managing referral coupons, where we would print out a QR code on the fly for connecting the network data, and all surveys were administered with an iPad using a connected Qualtrics survey. In 2024, we adapted the 2023 methodology for the Microsoft 365 Power App. The coupon for our RDS study from 2024 can be seen in Figure 3 and a picture from the 2023 study of a student administering the study to a respondent can be seen in Figure 2.

Sample size

A common problem in RDS is ascertaining how large the sample size should be for a given target population, which a general research question of interest in the field. In this case, we want a sample that accurately estimates the total count of unsheltered people in a given jurisdiction. Our preliminary 2019 work in Davidson County, TN (Nashville metro),[13] prepared us to design the King County study despite the COVID-19 interruptions. Our power analysis suggested that a minimum 5% sample would be sufficient, but as high as a 20% sample would be ideal. The King County, WA, January 2020 PIT count reported around 12 000 people were experiencing homelessness,[14] suggesting from our power analysis that a sample size of 600 to 2400 people would be the most efficient (low bias and variance) and thus the most cost-effective. Details are available in the online appendix (Section A). In 2022, we surveyed 671 respondents in King County; we then did a methods follow-up in 2023, focusing on improving data collection where we surveyed 1106 people experiencing homelessness and performed an official 2024 PIT count with 1464 respondents.

Figure 2. University of Washington unofficial PIT count 2023, South Park Library, Seattle, WA. On the left is a volunteer survey taker, and on the right is the respondent.

Seed selection

For seed selection, we engaged with outreach workers, service providers, and community liaisons who knew the various communities (eg, tents, tiny homes, etc.) and established trust with people experiencing homelessness.

Hub selection

The spatial distribution of data collection is important for RDS studies.[15,16] To provide sufficient geographic coverage within our time and budget constraints, we chose physical locations for surveying, known as "hubs." These were selected based on 10 official subregions (see Figure 4d), transit maps, previous PIT counts, and feedback from people with lived experience to optimize the location of each hub. We varied the hub location a bit for a combination of practical reasons and coverage as we scaled up our sampling strategy. All hub locations can be seen in Figure 4.

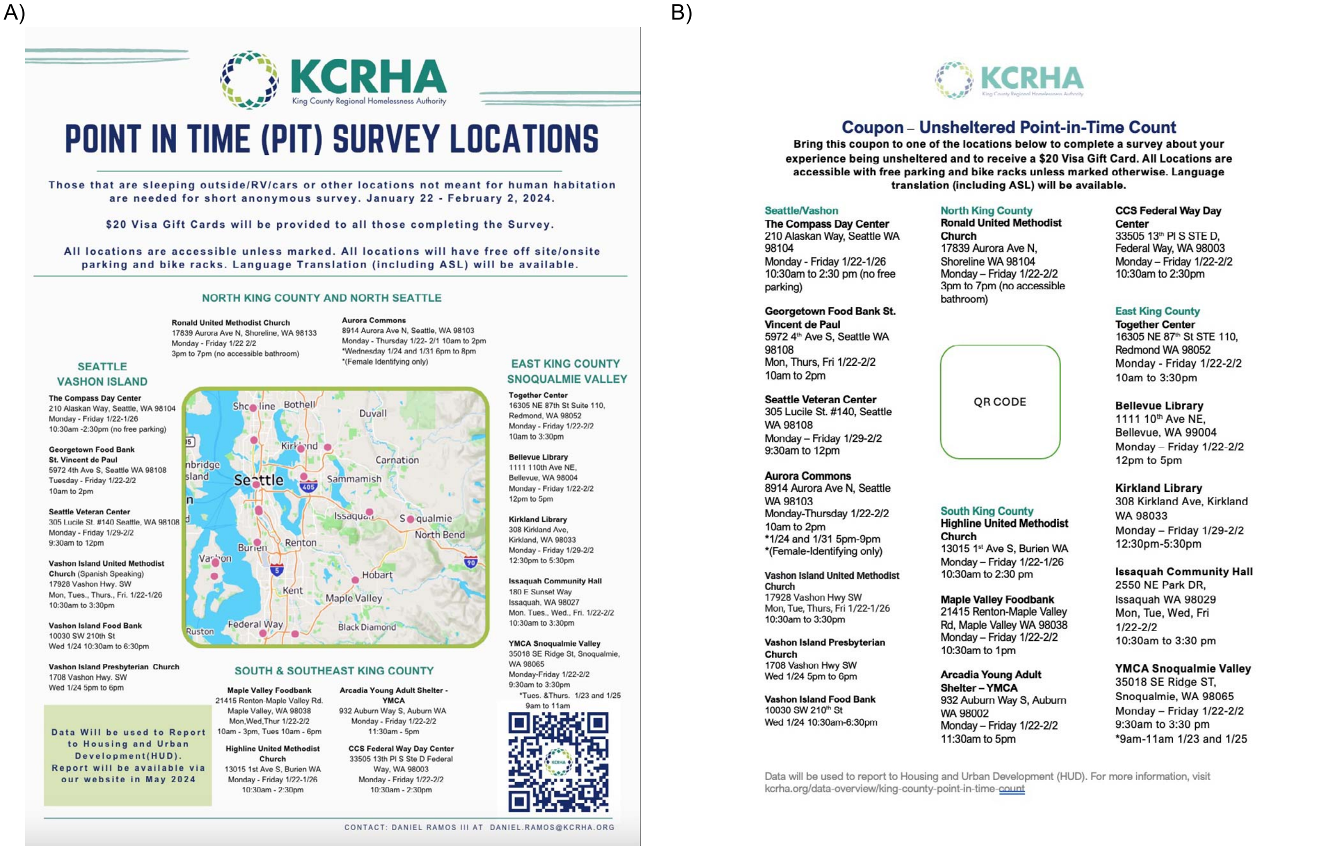


Figure 3. Coupons were preprinted for each hub location with space reserved for the QR code printed on the fly from our custom RDS software.

Eligibility criteria

We interviewed individuals who (1) were 18 years or older; (2) were able to give informed consent; and (3) identified as unsheltered, using the HUD definition of "an individual or family who lacks a fixed, regular, and adequate nighttime residence," including people who have a primary nighttime residence that is a public or private place not meant for human habitation. Interviewers conducted interviews in English and Spanish; we used a phone system to provide real-time translations for the few participants who spoke neither English nor Spanish (in 2024).

Ethics

The University's institutional review board approved all study procedures. Respondents were assigned a unique ID that allowed linkage to individuals they referred to be surveyed. Otherwise, interviewers collected only limited personal identification information: (1) first two letters of first and last name, and (2) year and month of birth. We collected items (1) and (2) as one of our methods of limiting duplication in the survey sample. We collected phone numbers or email data to provide digital gift cards for the peer referral incentive. However, these data were held separately and were not connected to the survey data respondents provided.

Estimation strategy

The estimation strategy leverages the dichotomous nature of how HUD counts homelessness: (1) sleeping in an emergency shelter or (2) sleeping unsheltered. The population of unhoused people is the sum of (1) and (2). We employ administrative data to enumerate sheltered people and based our estimate of unsheltered individuals by the estimate of the proportion of unsheltered people from our RDS procedure. The following section details our approach.

Estimating percentages

A robust literature offers estimators for population percentages using RDS samples. Details on the classic estimators can be found in Salganik et al,[7] Fellows,[12] and Gile et al.[17] with assessments in Sharad,[18] and a simulation-based approach developed by Gile and Handcock.[19] All the estimators rely on a reweighting by the degree distribution to unbias the resulting sample estimator and generally follow the basic logic of the classic Horwitz-Thompson estimator. These estimators are readily available in R via the *RDS* package.[20] In this article, we use the Salganik-Heckathorn estimator[7] (see details in the online appendix: Section B).

Estimating totals

To estimate the total number of unsheltered people, we employed the Salganik-Heckathorn estimator.[7] We leveraged the fact that people experiencing homelessness are in one of 2 categories for the PIT: (1) the shelter count (those tabulated as using emergency shelter) and (2) the unsheltered population (all other people who meet the HUD definition of homeless and are not on the night of question using an emergency shelter). This allows us to create a ratio between unsheltered and sheltered counts over the years; while this varies over time, it is a rough test of the validity of our findings. Thus, we can describe the population of people experiencing homelessness as,

$$N = N_U + N_S. \tag{1}$$

Where, N is the total number of people experiencing homelessness on a given night, N_U is the total number of people living unsheltered on a given night, and N_S is the total number of people living in emergency shelters on a given night.

Combining this observation with the Salganik-Heckathorn proportion estimator allows us to derive the following estimator for the total number of people living unsheltered:

$$\hat{N}_U = N_S \frac{\hat{\mu}_U}{1 - \hat{\mu}_U}. \tag{2}$$

Where, $\hat{N}_U$ is our statistical estimator for the total number of people living unsheltered on a given night, N_S is the total number living in an emergency shelter on a given night (as reported to HUD from the HMIS database and other service providers), and $\hat{\mu}_U$ is obtained from the Salganik-Heckathorn estimator (see the online appendix: Section B). Standard errors can be computed via the δ-method[21] or through bootstrap methods.[22]

Results

RDS results

In 2022, we interviewed 671 people in 9 hubs over 24 days (see Figure 4a), with the longest chain being 8; see Figure 5a and Figure 5b to visualize the hub and network results. (It is also important to note that in 2022 the demographic, needs, and network survey was paired with a 90-minute qualitative interview.) This greatly affected the number of interviews we could do in the time allotted for the PIT count. In 2023, we interviewed 1107 people in 11 hubs (see Figure 4b) over 38 days, with the longest chain being 19; see Figure 5c and Figure 5d to visualize the hub and network results. In 2024, we interviewed 1446 people in 17 hubs (see Figure 4c) over 11 days, with the longest chain being 20; see Figure 5e and Figure 5f to visualize the hub and network results. We observed large improvements in the number of recorder referrals, length of chains, and minimization of recorded isolates with the shift from pen and paper recording to our custom software and survey system (this also resulted in reduced cost as we did not have to pay for transcription in 2023 and 2024). Convergence plots of race/ethnicity and gender can be found in the online appendix (Section B) following the basic guidance of.[11]

Comparison between 2020 PIT and 2022 PIT in King County

HUD waived the requirement to conduct an unsheltered PIT count in 2021 in King County due to COVID-19. The only unsheltered PIT count conducted in 2022 was the RDS unsheltered PIT count discussed in the last section. To build a comparative case study, we constructed an autoregressive integrated moving average model of historical PIT county data (ARIMA[23];) to impute the 2021 and 2022 unsheltered PIT count. We used the R package "forecast"[23] to find the best fitting model over the 14-year period (2007-2020) of unsheltered PIT data in King County, WA.[24] We employed the Akaike information criterion (AIC) model fit criterion to find the best-fit model.[25] We found an ARIMA(0,1,0) with a covariate of the shelter count (available in all years) was the best-fitting model (model parameters are available in the online appendix, Section C). In Table 1, the RDS estimate of 7,685 unsheltered people is compared to the forecasted count of 6,819 unsheltered people. The confidence interval highly overlaps, and the 2 estimates would not be statistically distinguishable.

Comparison between 2024 RDS PIT and forecast of the 2024 PIT in King County

Similar to the 2022 case, we forecast the expected number of unsheltered people with an ARIMA model (see online appendix: Section C). Table 2 compares the RDS estimate of 9692 unsheltered people (over 18) with the forecasted amount of 8946 unsheltered people. The confidence interval highly overlaps, and the 2 estimates are not statistically distinguishable.

Table 1. Respondent-driven sampling estimate of the total unsheltered population in 2022 with statistical confidence interval (95%) computed using the δ method[21] compared to the best prediction of the visual PIT count forecasted by an ARIMA(0,1,0) with covariate of the shelter count (available in all years) and selected by AIC criterion.[25]

	RDS	ARIMA[a]
Point Estimate	7685	6819
Lower Bound (95% CI)	6816	5277
Upper Bound (95% CI)	8555	8360

[a]Forecasted visual unsheltered count vs RDS estimate for King County, WA, 2022.

Overall, this is strong evidence that the 2 methods should be highly correlated. Still, there are some obvious advantages of RDS over flashlights and clipboards: (1) the demographic survey is conducted at the same time and on the same population as the unsheltered count (the historic PIT count model requires decoupling these measures to avoid waking people in the night to interview them)[14]; (2) the RDS estimate has a confidence interval and statistical uncertainty in its formulation that the visual census does not (although this requires explanation); and (3) clear strategies are going forward to reduce estimate error. Further, this method allows people experiencing homelessness to voluntarily engage in the process and invite their contacts to do likewise. Both approaches (middle-of-the-night census and our proposed RDS) benefit from trained volunteers whose participation deepens community engagement with this social problem.

Discussion

HUD delegates to CoCs the responsibility for counting the number of people in outdoor and sheltered homeless situations. Although soundly critiqued, the "one night" crude census approach to estimating unhoused populations has become entrenched.[26] Enumeration strategies have been limited by a lack of innovation, driven by a desire to ensure compliance with CoC funding requirements (most CoCs are small, with minimal staff).

Policymakers and the general public will focus on the central tendency number, so low statistical bias is the most important feature of an estimator for policy uses. While we acknowledge the RDS estimator does provide more variance than we would like (this can be improved; see our proposals at the end of this section), the simulation and comparison studies show that we expect this single most used number, the mean estimate, to be of high-quality and also to provide some basic guidance (e.g., actual statistical bounds) to remind policymakers that it is a statistical estimate and that some caution around the interpretation of the number

Table 2. Respondent-driven sampling estimate of the total unsheltered population in 2024 with confidence interval (95%) computed using the δ method compared to the best prediction of the historical PIT count forecasted by an ARIMA(0,1,0) with drift selected by AIC.

	2024 RDS	ARIMA[a]
Point Estimate	9692	8946
Lower Bound (95% CI)	8458	7236
Upper Bound (95% CI)	11 107	100 657

[a]Forecasted unsheltered count vs RDS estimate for King County, WA, 2024.

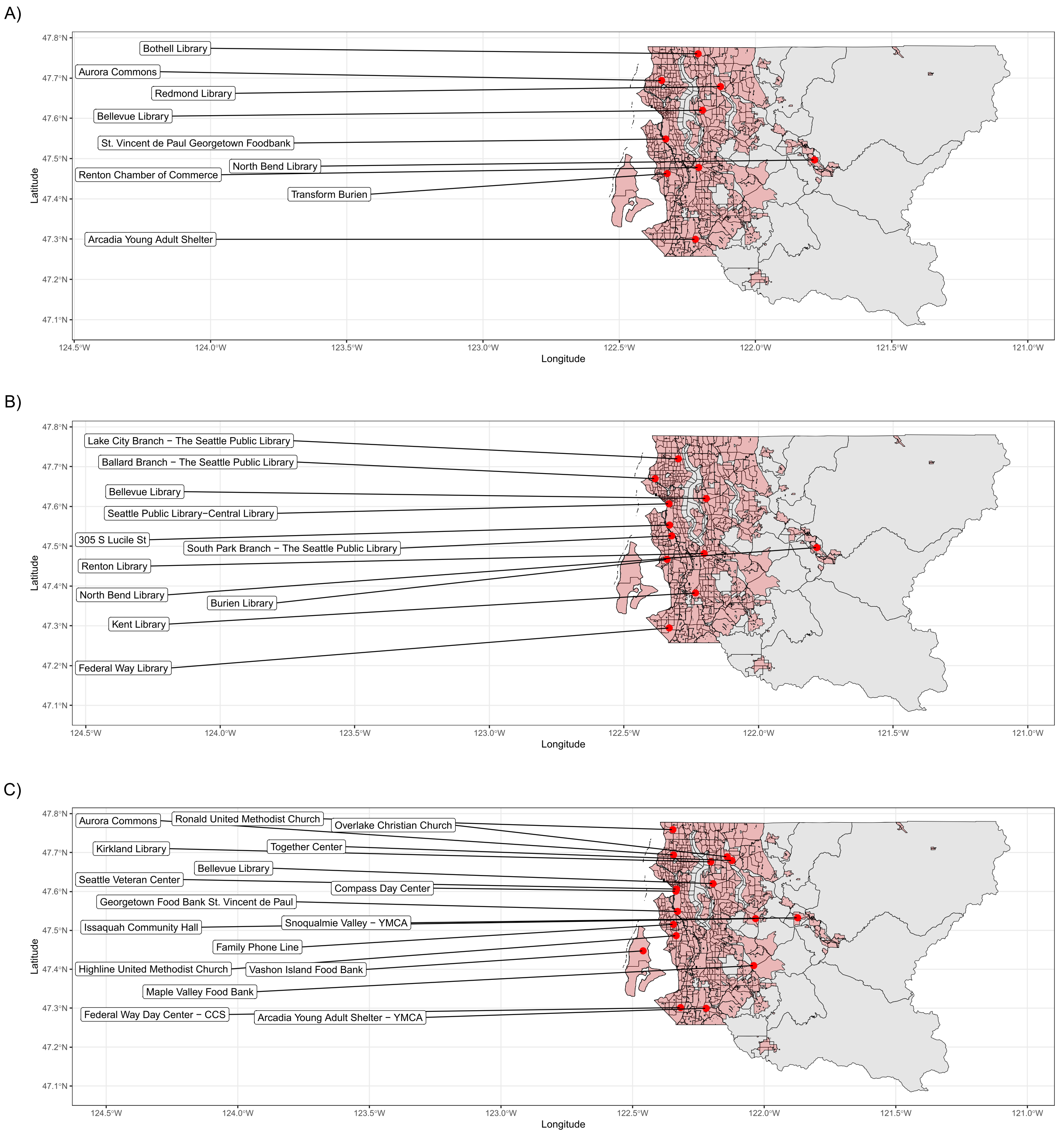


Figure 4. Continues

is warranted. This method also provides a lower and upper bound for logistical purposes in expectation of the support one will likely need for the unsheltered population.

There are several ways demographic information on unsheltered people experiencing homelessness could be acquired. Our solution provides a good framework for a sampling-based approach (with uncertainty bounds) and a straightforward way to offer community members a voluntary means to share their experiences. Here, we introduced a tested strategy built on the large body of work in public health for measuring hard-to-reach populations. We demonstrated in Davidson, TN, and King County, WA, that respondent-driven sampling can obtain an estimate of the percent of unsheltered people, which can be extrapolated to estimate the count of the unsheltered population of people experiencing homelessness by leveraging the known population of people using emergency shelters at the same time. Further, the RDS framework provides an ethical approach (ie, giving people the chance to volunteer to be in the count [or not] without being contacted by a researcher) and does not require enumerators prowling around with flashlights in the night hoping to catch sight of people "sleeping rough," as is done with traditional PIT counts.

Our sampling design and strategy have limitations. First, we know that even with all our efforts with community stakeholders and people with lived experience, we may not be able to access

D)

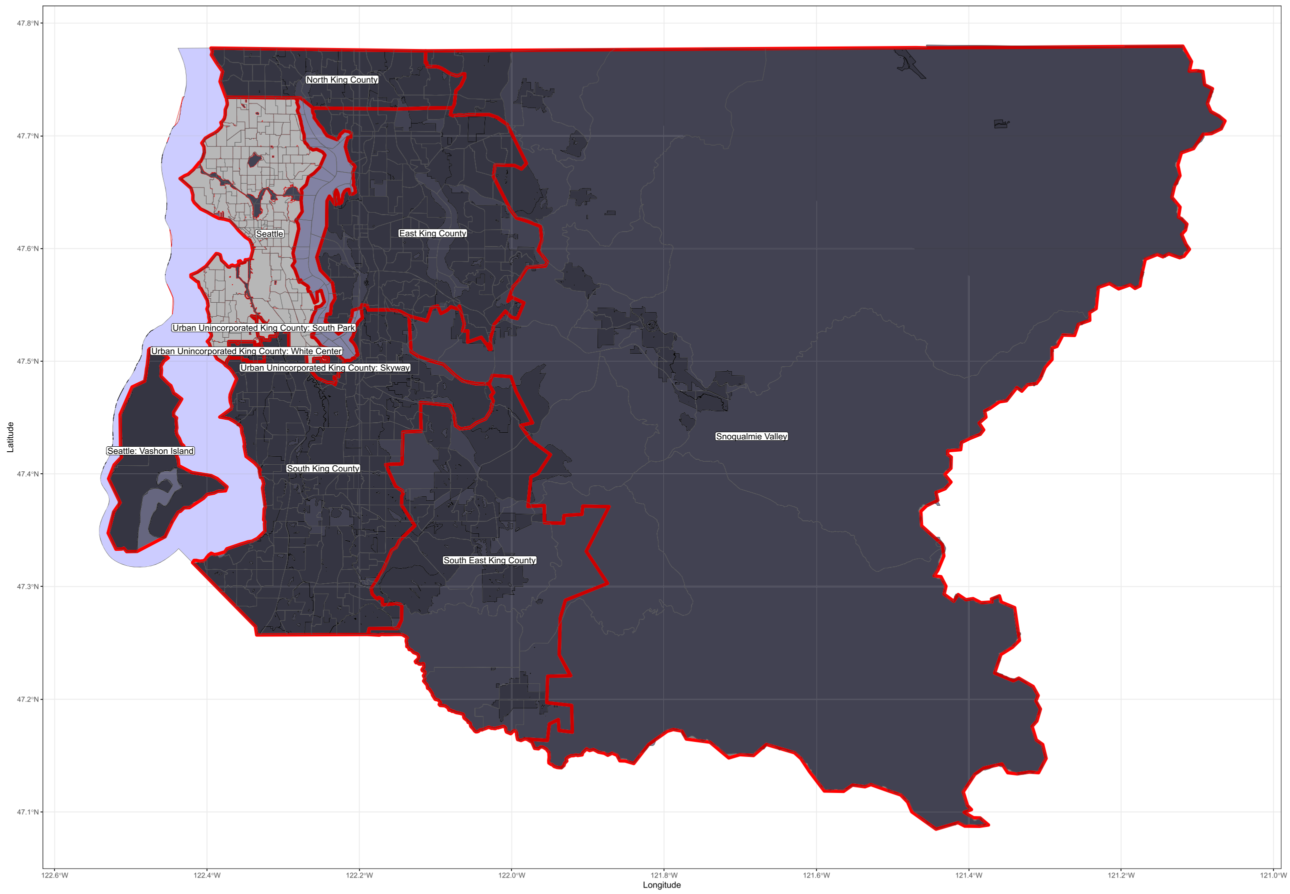


Figure 4. Nine hub locations used for the 2022 PIT count in King County. US census tracts are colored in gray, Seattle City US census tracts are shaded in blue, and incorporated and unincorporated urban areas are in red. The red dot is a hub location labeled with a named location. King County is the 12th most populous county in the United States, home to dense urban and remote rural communities.

some people who are sufficiently isolated; thus, as with all counts of people experiencing homelessness, we expect our result to be more of a lower bound than the upper bound of the problem. Hub locations must change over time because the population, organizations, or services have moved. We know some individuals took the survey multiple times (see online appendix: Section C), but there is room for applying modern machine learning (ML) techniques to refine what we did over the 3 surveys. Ideally, we would have had a simultaneous comparison between a visual unsheltered PIT count and the RDS unsheltered PIT count. Still, even without the direct comparison, our results show that the RDS method captures the same information within the statistical range while also providing people with compensation and the ability to choose how to engage with the count. Visual PIT count strategies are relatively straightforward. Teams of volunteers go out at night to count the number of people. After some deduplication efforts, which vary in rigor, the final unsheltered PIT count reflects the total counted by all teams. Our proposed method requires more technical skills than a typical CoC may have on staff. A CoC like KCRHA has all the logistical support necessary for implementing the RDS survey approach but has relied on statistician partners at the University of Washington for final count numbers. Communities of Care new to this process should, like KCRHA, look to team up with universities or outside statisticians; however, in the long run, we hope to develop an open-access app and dashboard, allowing most regional jurisdictions to run this process with in-house staff.

The pros to this method are (1) that it combines the demographic survey and counting exercise into a single process, so the demographics align with the enumeration; (2) it is a statistical procedure, with quality and statistical bounds on the results; and (3) it allows the people experiencing homelessness themselves to choose to participate in the count. The major con is that our proposed process is rather technical, so statistical support from an academic partner is required to ensure fidelity to the process; however, we propose open-access software and instructions that could reduce the barriers for most jurisdictions that now engage in a traditional PIT count. The new method also requires explaining the concept of statistical uncertainty to the public and policy members (who should be familiar with the basic idea from polling exercises).

Over the 3 years of surveying, we found that geographic coverage was important for obtaining a representative sample. Overall, the survey time could be reduced by recruiting sufficiently available volunteers to run all hubs simultaneously (see the 2024 results). Improved software and QR printing greatly improved network and survey data collection, with incentives for referrals being very important (see the length of chains in 2022 compared to 2023 and 2024). With sufficient resources, obtaining a high-quality, large sample in less than 2 weeks with sufficient preparation appears possible. Future work should look at methods for conducting RDS surveys close in time, say for a quarterly or monthly survey, and to consider using RDS as a method for generating county-wide representative samples for large-scale

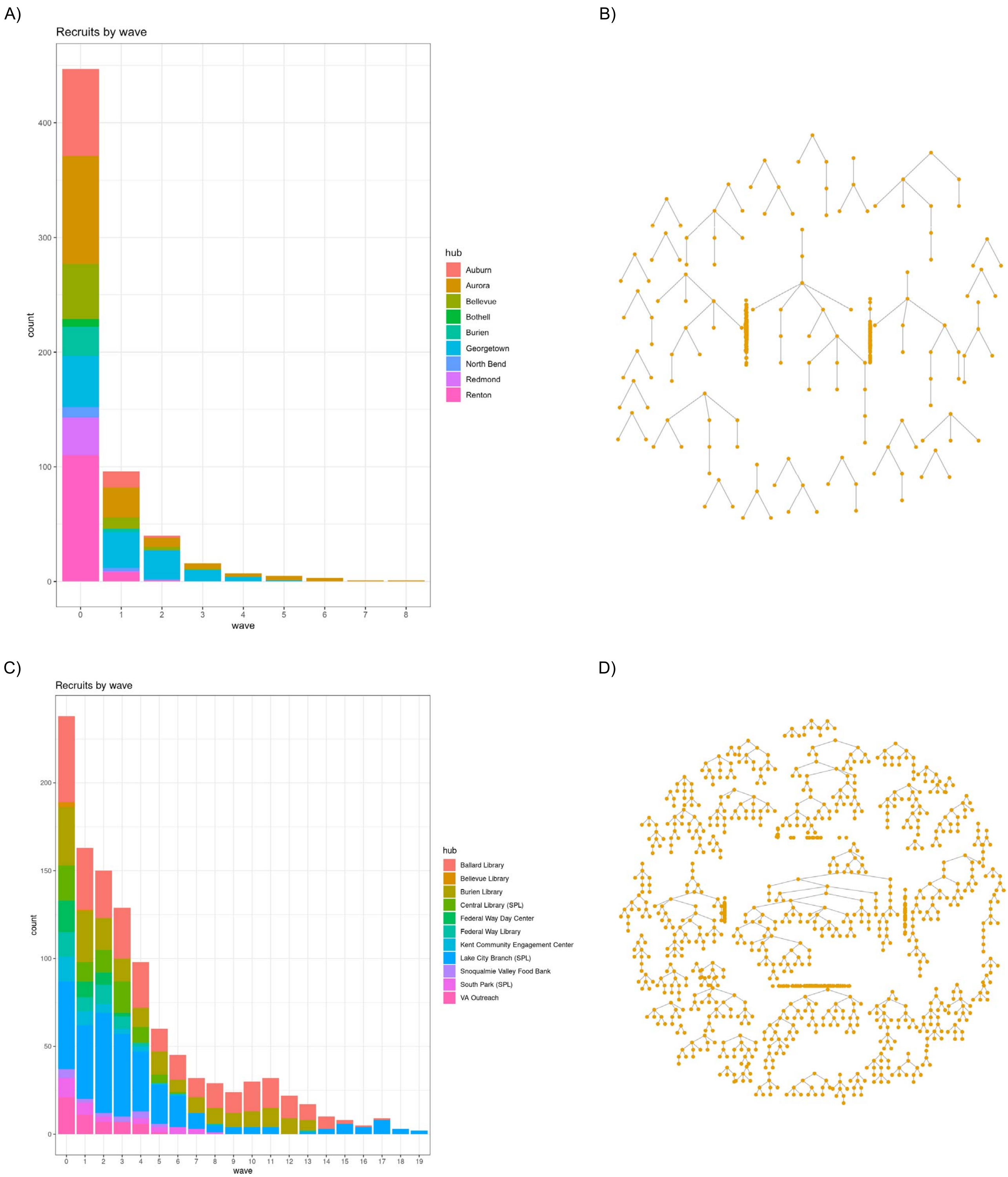


Figure 5. Continues

state-wide assessments like that done in California (see Margot et al for an example).

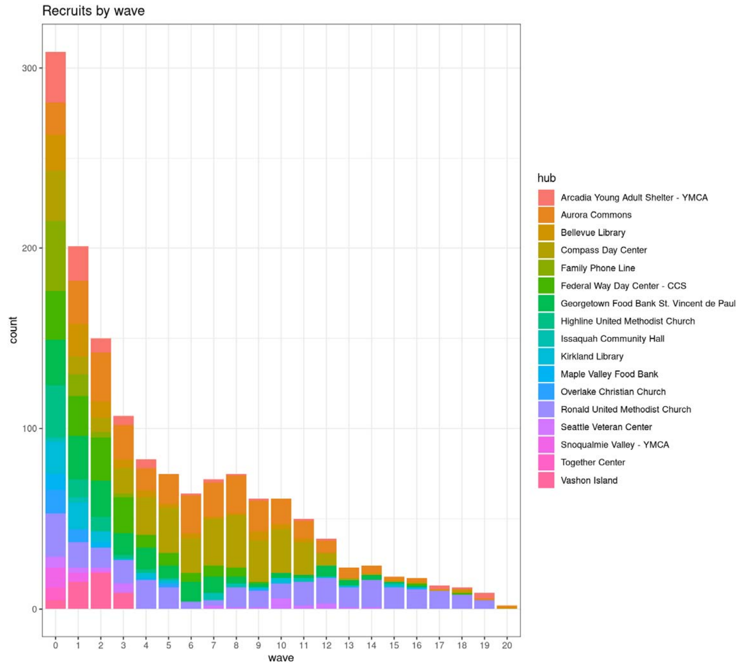
Organizational interests

The King County Regional Homelessness Authority has a stake in addressing homelessness in King County and may be impacted by the outcomes of this survey project. As the Community of Care (CoC) lead, the King County Regional Homelessness Authority must complete a biannual unsheltered Point-in-Time count by both Federal and State statutes as a requirement for continued funding from the Department of Housing and Urban Development. As such, they have a stake in identifying valid methods of enumeration of people experiencing unsheltered homelessness.

Acknowledgments

We wish to thank the King County, WA community and all the people who participated in this survey.

E)



F)

Figure 5. Plots of the core RDS statistics by year.

Supplementary material

Supplementary material is available at *American Journal of Epidemiology* online.

Funding

Partial support for this research came from a Eunice Kennedy Shriver National Institute of Child Health and Human Development research infrastructure grant, P2C HD042828, to the Center for Studies in Demography & Ecology at the University of Washington; National Science Foundation CAREER Grant #SES-2142964; University of Washington Population Health Initiative Tier 2 and 3 Grant. The content is solely the authors' responsibility and does not necessarily represent the official NIH or National NSF views.

Conflict of interest

Z.A., I.K., A.H., M-C.A., L.A., and A.H. have no conflict of interest to declare. O.K., J.R., and C.G., as King County Regional Homelessness Authority employees, acknowledge the importance of transparency and accountability in scientific research and peer review. The King County Regional Homelessness Authority employed O.K., J.R., and C.G. during the 2022 Point-in-Time count. The King County Regional Homelessness Authority, in its role as the Community of Care (CoC) lead, both conducted and provided funding for the King County 2022 RDS survey, which was carried out as a requirement of the HUD-mandated biannual 2022 Point-in-Time count. Financial Interests: O.K., J.R., and C.G. declare that they have no financial interests, such as stocks, patents, or research grants, that may be perceived as affecting their objectivity in the peer review process.

Data availability

HUD Data exchange provides all official PIT totals from KCRHA (see https://www.hudexchange.info/resource/3031/pit-and-hic-data-since-2007/). Individual data for 2022 and 2024 is available via request to KCRHA and to the University of Washington through Zack W. Almquist for the 2023 PIT data.

References

1. Roncarati JS, Byrne TH, McInnes DK. Invited commentary: data sources for estimating numbers of people experiencing homelessness in the United States–strengths and limitations. *Am J Epidemiol*. 2021;190(11):2437-2440. https://doi.org/10.1093/aje/kwab104
2. Snow DA, Leon A. Researching the homeless. In: *A Case for the Case Study*. Vol. 148. 1991:173.
3. David HJ, Scott L. The Human Right to Adequate Housing: Chronology of United Nations Activity; 1945 to 1999. Centre on Housing Rights and Evictions (COHRE); 2000. https://utoronto.scholaris.ca/items/3ce3a23c-80de-4122-a22c-85e82ea0dccf
4. Tsai J, Alarcón J. The annual homeless point-in-time count: limitations and 2 different solutions. *Am J Public Health*. 2022;112(4):633-637. https://doi.org/10.2105/AJPH.2021.306640
5. Johnston LG, Hakim AJ, Samantha D, et al. A systematic review of published respondent-driven sampling surveys collecting behavioral and biologic data. *AIDS Behav*. 2016;20(8):1754-1776. https://doi.org/10.1007/s10461-016-1346-5
6. Kushel M, Moore T. Towards a new understanding: the California statewide study of people experiencing homelessness, UCSF BHHI. United States of America; 2023. Accessed March 13, 2025. https://coilink.org/20.500.12592/nbcc61
7. Salganik MJ, Heckathorn DD. 5. Sampling and estimation in hidden populations using respondent-driven sampling. *Sociol Methodol*. 2004;34(1):193-240. https://doi.org/10.1111/j.0081-1750.2004.00152.x
8. Heckathorn DD. Respondent-driven sampling: a new approach to the study of hidden populations. *Soc Probl*. 1997;44(2):174-199. https://doi.org/10.1525/sp.1997.44.2.03x0221m
9. Sarah R, DeVost MA, Digitale JC, et al. Respondent-driven sampling: a sampling method for hard-to-reach populations and

beyond. *Curr Epidemiol Rep*. 2022;9(1):38-47. https://doi.org/10.1007/s40471-022-00287-8
10. Erik V, Heckathorn DD. Probability based estimation theory for respondent driven sampling. *J Off Stat*. 2008;24(1): 79-97.
11. Gile KJ, Johnston LG, Salganik MJ. Diagnostics for respondent-driven sampling. *J R Stat Soc Ser A Stat Soc*. 2015;178(1):241-269. https://doi.org/10.1111/rssa.12059
12. Fellows IE. On the robustness of respondent-driven sampling estimators to measurement error. *J Surv Stat Methodol*. 2022; 10(2):377-396. https://doi.org/10.1093/jssam/smab056
13. Anderson M-C, Hazel A, Perkins JM, et al. The ecology of unsheltered homelessness: environmental and social-network predictors of well-being among an unsheltered homeless population. *Int J Environ Res Public Health*. 2021;18(14):7328. https://doi.org/10.3390/ijerph18147328
14. All Home. *Seattle/King County Point-in-Time Count of Individuals Experiencing Homelessness*. King County Regional Homelessness Authority; 2020. https://kcrha.org/wp-content/uploads/2020/06/Count-Us-In-2020-Final.pdf
15. Toldeo L, Codeço CT, Bertoni N, et al. Putting respondent-driven sampling on the map: insights from Rio de Janeiro, Brazil. *J Acquir Immune Defic Syndr*. 2011;57(Supplement 3):S136-S143. https://doi.org/10.1097/QAI.0b013e31821e9981
16. Jenness SM, Neaigus A, Travis W, et al. Spatial recruitment bias in respondent-driven sampling: implications for HIV prevalence estimation in urban heterosexuals. *AIDS Behav*. 2014;18(12):2366-2373. https://doi.org/10.1007/s10461-013-0640-8
17. Gile KJ, Handcock MS. Respondent-driven sampling: an assessment of current methodology. *Sociol Methodol*. 2010;40(1):285-327. https://doi.org/10.1111/j.1467-9531.2010.01223.x
18. Sharad G, Salganik MJ. Assessing respondent-driven sampling. *Proc Natl Acad Sci*. 2010;107(15):6743-6747. https://doi.org/10.1073/pnas.1000261107
19. Gile KJ. Improved inference for respondent-driven sampling data with application to HIV prevalence estimation. *J Am Stat Assoc*. 2011;106(493):135-146. https://doi.org/10.1198/jasa.2011.ap09475
20. Handcock MS, Fellows IE, Gile KJ. *RDS: Respondent-Driven Sampling (Version 0.9-6) [Computer software]*. Los Angeles, CA: CRAN; 2004. https://CRAN.R-project.org/package=RDS
21. Oehlert GW. A note on the delta method. *Am Stat*. 1992;46(1): 27-29. https://doi.org/10.1080/00031305.1992.10475842
22. Baraff AJ, McCormick TH, Raftery AE. Estimating uncertainty in respondent-driven sampling using a tree bootstrap method. *Proc Natl Acad Sci*. 2016;113(51):14668-14673. https://doi.org/10.1073/pnas.1617258113
23. Hyndman RJ, Yeasmin K. Automatic time series forecasting: the forecast package for R. *J Stat Softw*. 2008;27(3):1-22. https://doi.org/10.18637/jss.v027.i03
24. U.S. Department of Housing and Urban Development. *Point-in-time count and housing inventory count*. HUD Exchange. Accessed March 13, 2025. https://www.hudexchange.info/programs/hdx/pit-hic/
25. Akaike H. Information theory and an extension of the maximum likelihood principle. In Petrov BN, Csaki F, eds. *2nd International Symposium on Information Theory*. Budapest: Akademiai Kiado; 1973. pp. 267-281. https://doi.org/10.1007/978-1-4612-1694-0_15
26. National Academies of Sciences, Engineering, and Medicine. *Permanent Supportive Housing: Evaluating the Evidence for Improving Health Outcomes Among People Experiencing Chronic Homelessness*. National Academies Press, 2018.

American Journal of Epidemiology, 2025, **00**, 1–10
https://doi.org/10.1093/aje/kwae342
Original Contribution

Online Appendix: Innovating a Community-driven Enumeration and Needs Assessment of People Experiencing Homelessness: A Network Sampling Approach for the HUD-Mandated Point-in-Time Count

Zack W. Almquist*[1], Ihsan Kahveci[1], Mary Ashley Hazel[2], Owen Kajfasz[3], Janelle Rothfolk[3], Claire Guilmette[3], MC Anderson[4], Larisa Ozeryansky[1], and Amy Hagopian[1]

[1]University of Washington
[2]University of California, San Francisco
[3]King County Regional Homelessness Authority
[4]Stanford University

Contents

Online supplement

A RDS power analysis

In this section, we will introduce a simulation-based method for obtaining a power analysis for an RDS sample to estimate the count of unsheltered people in a given community. To determine the sample size needed, we first need a way to simulate the entire process and compute the core statistic of interest (size of the unsheltered population). To do this, we fit a general statistical model for social networks on an egocentric (personal network) sample of people experiencing homelessness in Davidson County, TN (Nashville metro). Then, we use this model to simulate a complete network for the county. This is followed up with a repeated simulation of a peer-referral process to give us a general understanding of the bias and variance of our estimator. We can then map the results against different sample sizes to guide the recommended sample size for our RDS in another setting.

*Corresponding Author: zalmquist@uw.edu

ERGM of People Experiencing Homelessness in Nashville, TN, 2020

	Estimate	Std. error	MCMC %	z value	$Pr(> \|z\|)$
offset(N)	-6.342	0.000	0	-Inf	0.000
edges	2.037	1.071	0	1.901	0.057
degree2	0.338	0.491	0	0.689	0.491
degree3	0.117	0.903	0	0.129	0.897
degree4	0.854	1.161	0	0.735	0.462
degree5	1.300	1.144	0	1.136	0.256
degree6	1.239	1.304	0	0.950	0.342
I(Summer)	–	–	–	–	–
I(Fall)	-0.881	0.841	0	-1.048	0.295
I(Winter)	-0.867	0.865	0	-1.003	0.316
I(Spring)	-0.804	0.904	0	-0.889	0.374
I(White)	–	–	–	–	–
I(Black)	-0.016	0.433	0	-0.038	0.970
I(LatinX)	-0.731	0.878	0	-0.833	0.405
I(Asian)	0.308	1.482	0	0.208	0.835
I(AIAN)	-0.202	0.694	0	-0.291	0.771
I(NHPI)	-0.552	0.793	0	-0.696	0.487
I(Female)	–	–	–	–	–
I(Male)	0.277	0.221	0	1.251	0.211
Gender Node Match	-0.423	0.188	0	-2.246	0.025

Table 1: ERGM parameters fit the Nashville, TN egocentric data. – represents the reference group. I indicates a node factor (i.e., an indicator function), and the "node match" term provides a weight for any time the two individuals have the same gender (known as homophily).

A.1 A statistical model for the social network of people experiencing homelessness

An important method for the statistical modeling of social networks is the so-called Exponential Random Graph Model (ERGM), which allows for the inference and simulation of social networks. Formally, ERGM provides a framework for writing a generative probability model for social networks, see for example (1, 2, 3, 4, 5)). In the field of demography and public health, ERGM has been employed to understand racial mixing and other issues (6). Given a random graph G on support $\mathcal{G}$, we may write ERGM formally as follows:

$$\Pr(G = g|s, \theta) = \frac{\exp \theta^T s(g))}{\sum_{g' \in \mathcal{G}} \exp(\theta^T s(g'))} \mathcal{I}_{\mathcal{G}}(g)$$

where $\Pr(\cdot)$ is the probability mass of its argument, $\mathcal{G}$ is the support of G, g is the realized (observed) graph, and s is the function of sufficient statistics, θ is a vector of parameters (e.g., the degree distribution of the graph) and $\mathcal{I}$ is an indicator function. This model can be fit by MLE; see (7) for details. When fit, the results can be displayed in a regression table that can be interpreted in a conditional probability framework similar to logistic regression.

We have an egocentric sample (connection of focal person and their alters (e.g., other people experiencing homelessness) and their perception of their alters relationships) of Davidson County, TN (Nashville metro) from (8), representing about 50% of the total number of unshelterd population based on the 2020 PIT count (9). We can fit an ERGM to this sample to simulate the complete network of people experiencing homelessness. This process is straightforward for exponential family models (discussed above), where we build the mean statistics for the model out of sample data and fit with MCMC-MLE. Using egocentric data to fit ERGMS was developed by (10) and implemented in software by (11, 7). Here, we fit a model that takes into account the population size (this is handled with an offset to the density term), edges (analogous to density), degree up to six, and fixed effects (node factor) for the time period of data collection (fall, winter, spring, and summer), and race/ethnicity and gender. The results are presented in Table 1. We will use this model to develop our power analysis for our method.

A.2 Power Analysis Strategy for RDS Sampling based on simulated social network between people experiencing homelessness

We can use the estimated ERGM in the last section of people experiencing homelessness in Davidson County, TN, to simulate a realistic, complete network for people experiencing homelessness in Davidson County, TN (Nashville metro). This follows the same basic strategy as that employed by (7) or (12). To simulate an RDS process, we can then employ the (13) R package for generating a peer-referral process on a social network of people experiencing homelessness. The results of the simulated network can be visualized in Figure 1:C, and the RDS simulation can be visualized in Figure 1:D.

Simulated Unhoused Network for Nashville, TN, CoC with a Sequence of Simulated RDS Processes

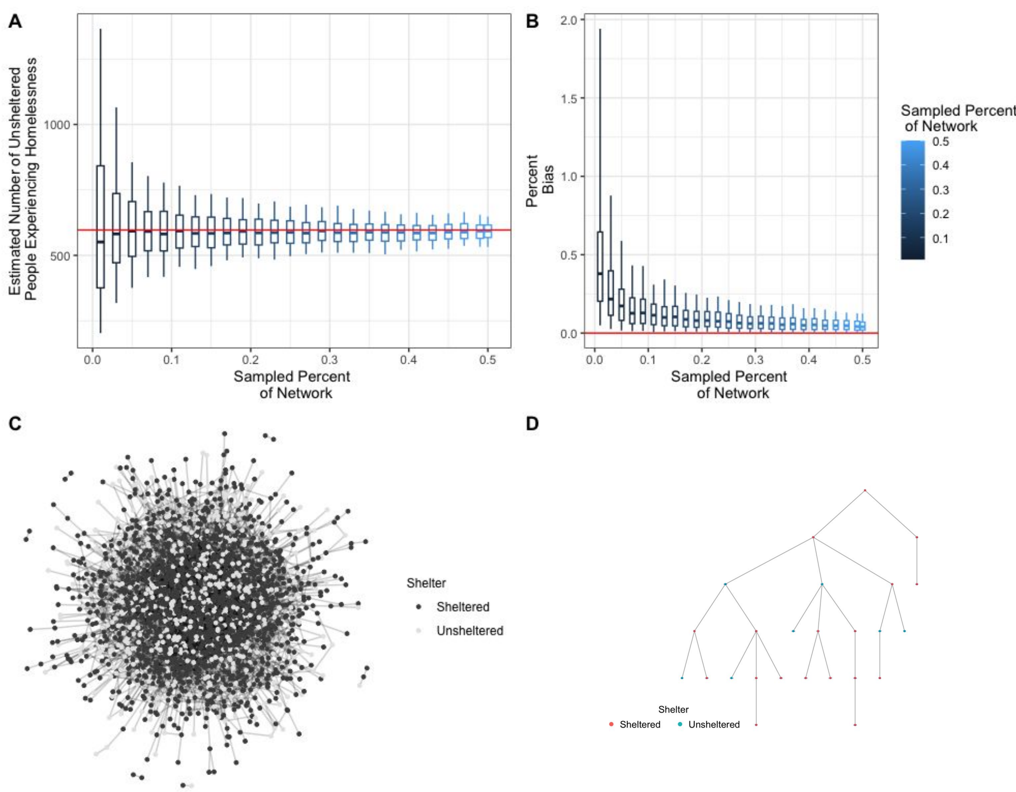


Figure 1: ERGM Simulation of a complete (shelter and non-shelter users) network of people experiencing homelessness in Nashville, TN (2,035 people) with 597 unsheltered people and 1,438 people using shelters in the network. **A** Represents the estimate of the total number of unsheltered people compared against the sample size in relative terms (i.e., as a percent of the total population). The red line represents the true value. **B** Plots the bootstrap estimate of bias against the sample size in relative terms (i.e., as a percent of the total population). **C** Represents the complete network of 2035 people. **D** Plot of an example RDS tree.

We can then employ this simulated network of the unhoused population to provide a general *power analysis* for an RDS sample for estimating the total number of unsheltered people. For example, let's take the case of Davidson County, TN (Greater Nashville, TN). Using our ERGM model fit from the egocentric data; we can extrapolate a complete network of 2,035 with 597 unsheltered people and 1,438 people using shelters in the network. In Figure 1: A, we see the estimated total unsheltered people and a bootstrap 95% confidence interval plotted against the percent of the population sampled. From this data, we see that our bias levels off, and our variance is close to minimum at around 5% of a sample and fully stabilizing at around 20% of the sample. Similarly, in Figure 1: B, the sample's statistical bias (sample estimate minus the true value) quickly shrinks with 0.05% of a sample, fully stabilizing at around 0.2%. Overall, the statistical bias is quite small in all cases, and the mean statistic is quite good. However, the variance, typical with straight RDS, is wider than standard survey methods with a sampling frame.

A.3 Summary of the power analysis

Following the results from our simulation analysis (Figure 1), we find that ideally, we would not have less than 5% and that, if possible, a 20% sample would be preferred.

B PIT count method

In this section, we derive a general estimator when a target population can be broken into two groups and one group has a known size. To do this, we Leverage the RDS estimator, built from the classic generalized Horvitz-Thompson estimator ((14); what (15) describes as the Psuedo Horvitz-Thompson estimator).

Let us first consider a population of N individuals with a known probability of being sampled, π_i of ith individual in the sample. We can estimate the population mean, μ, of any quantity, z_i, measured on the sampled individuals using a Horvitz-Thompson estimator (14):

$$\hat{\mu} = \frac{1}{N} \sum_{i=1}^{N} S_i \frac{z_i}{\pi_i}, \tag{1}$$

Where S is the random N-vector representing the sample, such that $S_i = 1$ if unit i is sampled and is otherwise 0. There are two major drawbacks to this estimator in the network sampling context: (1) the population size N is often unknown (in fact, the whole problem is that it is unknown in our case!), and (2) the inclusion probabilities π_i are also typically unknown. The first issue can be solved by plugging in the unbiased estimator of N, $\hat{N} = \sum_{i=1}^{N} S_i \frac{1}{\pi}$, to obtain

$$\hat{\mu} = \frac{\sum_{i=1}^{N} S_i \frac{z_i}{\pi_i}}{\sum_{i=1}^{N} S_i \frac{1}{\pi_i}}, \tag{2}$$

which is the ratio of two unbiased estimators and tends to estimate μ with small bias for large sample sizes. This variant of the Horvitz-Thompson estimator is known as the *generalized Horvitz-Thompson estimator* or the Hajek estimator (14).

Salganik-Heckathorn Estimator: (16) introduced an estimator that leverages the relationship between two groups, A and B, where $\bar{d}_A$ is the mean degree of a group A, $\bar{d}_B$ is the mean degree of a group B, and c_{AB} and c_{BA} are the cross-ties between the two groups. The total for the groups is thus N_A, N_B such that $N = N_A + N_B$. Given the above definitions, we can write the following:

$$\frac{N_A}{N} = \frac{\bar{d}_B \cdot c_{BA}}{\bar{d}_A \cdot c_{AB} + \bar{d}_B \cdot c_{BA}} = \mu_A \tag{3}$$

To employ this estimator, we need a way to estimate,

- The mean degree of $\bar{d}_A$ and $\bar{d}_B$ (for example, the people experiencing homelessness who are using an emergency shelter and those individuals who are not on a given night).

- The proportion of social relations of group A to group B, and visa versa (for example, the count of people in an emergency shelter community that are known to the unsheltered community and visa versa).

Let t_{AB} represent the total number of relations between groups A and B. Then, we can write an estimator for the cross ties:

$$c_{AB} = \frac{t_{AB}}{N_A \bar{d}_A}$$
$$c_{BA} = \frac{t_{AB}}{N_B \bar{d}_B}$$

In practice, (16) assumed edges were sampled randomly, which allows one to estimate c_{AB} out of the sample proportion of subjects in group A who recruit participants in group B. c_{BA} may be estimated in the same way as c_{AB}. The mean degrees can be estimated using the Hajek estimator introduced earlier.

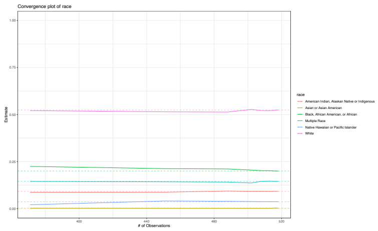
If we assume the size of group B, N_B is known with little or zero error, we write an estimator of the size of N_A:

$$\begin{aligned} N_A + N_B &= N \\ N_B &= \frac{N_A}{\mu_A} - N_A \\ N_B &= N_A \frac{1-\mu_A}{\mu_A} \\ N_A &= N_B \frac{\mu_A}{1-\mu_A} \end{aligned} \tag{4}$$

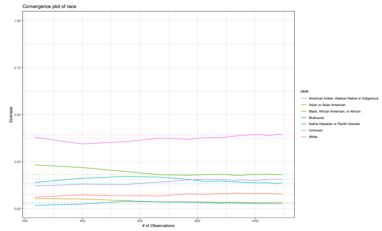
Notice that $N_B = N_A \frac{\mu_B}{1-\mu_B}$ analogously. We can further estimate standard errors (SE) and confidence intervals through bootstrap methods (17). In our case, we take A to be the population of people not using shelter (U) and B to be the population using emergency shelter (S) on a given night so that $N_U = N_S \frac{\mu_U}{1-\mu_U}$.

RDS Diagnostics

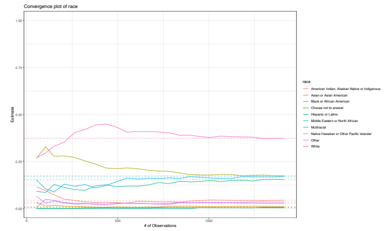
We follow the basic guidance of Gile et al. (18) proposal to examine the dynamics of the RDS estimate through convergence plots. Based on (18), we have made convergence plots for 2022, 2023, and 2024 RDS samples of unsheltered people experiencing homelessness. We can see the convergence plot for race/ethnicity in Figure 2 and gender in Figure 3. In all three cases, the final percentage settles down by the end of the sampling period.



(a) Convergence plot of race/ethnicity for 2022.

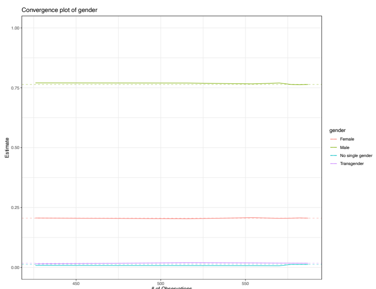


(b) Convergence plot of race/ethnicity for 2023.

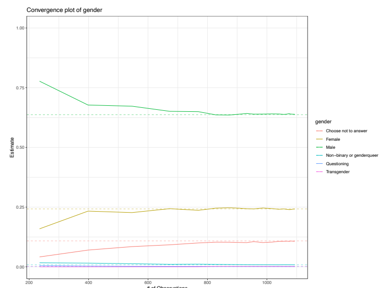


(c) Convergence plot of race/ethnicity for 2024.

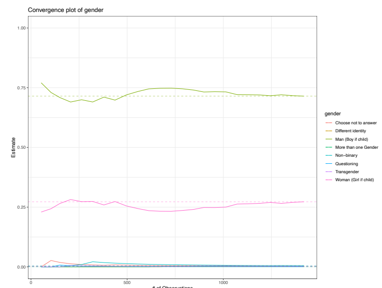
Figure 2: RDS convergence plots suggested by (18) as diagnostic tool.



(a) Convergence plot of gender for 2022.



(b) Convergence plot of gender for 2023.



(c) Convergence plot of gender for 2024.

Figure 3: RDS convergence plots suggested by (18) as diagnostic tool.

C ARIMA Model

RDS versus Forecast of Visual PIT Count

	King County Unsheltered Count
	Coefficient (Standard Error)
drift	306.69** (144.94)
log(Shelter Count)	−7,580.84* (4,573.39)
Observations	13
Log Likelihood	−99.52
σ^2	309,249.90
Akaike Inf. Crit.	205.05
Note:	*p<0.1; **p<0.05; ***p<0.01

Table 2: Fitting an ARIMA(0,1,0) regression to King County Unsheltered PIT data, 2007-2020, to forecast the 2022 visual PIT count. The best-fit model selected by AIC.

RDS versus Forecast of Historical PIT Count, 2024

	King County Unsheltered Count
	Coefficient (Standard Error)
drift	420.231**
	(205.358)
Observations	14
Log Likelihood	−104.340
σ^2	593,915.600
Akaike Inf. Crit.	212.680
Note:	*p<0.1; **p<0.05; ***p<0.01

Table 3: Fitting an ARIMA(0,1,0) with drift to regression to King County Unsheltered PIT data, 2007-2022, to forecast the 2024 PIT count. The best-fit model selected by AIC.

Methods used for discovering duplication of respondents

In 2022 and 2023, we relied on our survey taker notes to remove respondents who completed the survey more than once (less than 1%) of total surveys (and removed from the final total).

In 2024, to detect potential repeat takers, we relied again on survey taker notes but also used analytic and network methods to detect repeat respondents. We improved our record keeping by training our volunteers to leave a note in the app if they recognized a person before. While, we did not collect personally identifiable information to maintain anonymity and improve our relationship with this vulnerable population; however, we collected several weakly identifiable pieces of information (birth month and year, first two letters of first and last name) that we used to identify matching records when combined. For example, the same individual: "X" (based on the first two letters and last two letters, month and year of birth), has appeared in multiple coupons (we can observe this in Figure 4 – we believe "X" was using his own coupons). We pruned the tree by keeping the first one observed. In total, we have identified 11 entries collected from 4 respondents (again sub 1%).

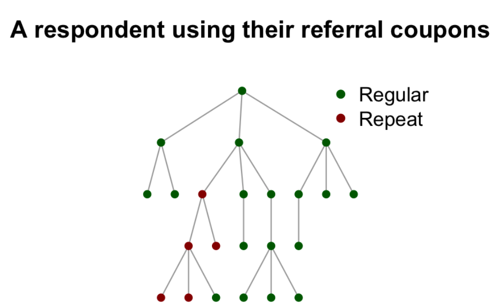


Figure 4: Tree where the respondent used their own coupons to take multiple surveys.

References

[1] Cranmer Skyler J, Desmarais Bruce A. Inferential network analysis with exponential random graph models *Political analysis.* 2011;19:66–86.

[2] Wasserman Stanley, Pattison Philippa. Logit models and logistic regressions for social networks: I. An introduction to Markov graphs and p *Psychometrika.* 1996;61:401–425.

[3] Snijders Tom AB, Pattison Philippa E, Robins Garry L, Handcock Mark S. New specifications for exponential random graph models *Sociological methodology.* 2006;36:99–153.

[4] Hummel Ruth M, Hunter David R, Handcock Mark S. Improving simulation-based algorithms for fitting ERGMs *Journal of Computational and Graphical Statistics.* 2012;21:920–939.

[5] Blackburn Bart, Handcock Mark S. Practical network modeling via tapered exponential-family random graph models *Journal of Computational and Graphical Statistics.* 2023;32:388–401.

[6] Goodreau Steven M, Kitts James A, Morris Martina. Birds of a feather, or friend of a friend? Using exponential random graph models to investigate adolescent social networks *Demography.* 2009;46:103–125.

[7] Hunter David R, Handcock Mark S, Butts Carter T, Goodreau Steven M, Morris Martina. ergm: A package to fit, simulate and diagnose exponential-family models for networks *Journal of statistical software.* 2008;24:nihpa54860.

[8] Anderson Mary-Catherine, Hazel Ashley, Perkins Jessica M., Almquist Zack W.. The Ecology of Unsheltered Homelessness: Environmental and Social-Network Predictors of Well-Being among an Unsheltered Homeless Population *International Journal of Environmental Research and Public Health.* 2021;18:7328. Number: 14 Publisher: Multidisciplinary Digital Publishing Institute.

[9] Nashville MDHA . Nashville and Davidson County Point-in-Time Count 2020 2020.

[10] Krivitsky Pavel N, Morris Martina. Inference for social network models from egocentrically sampled data, with application to understanding persistent racial disparities in HIV prevalence in the US *The annals of applied statistics.* 2017;11:427.

[11] Krivitsky Pavel N.. *ergm.ego: Fit, Simulate and Diagnose Exponential-Family Random Graph Models to Egocentrically Sampled Network Data.* The Statnet Project (https://statnet.org) 2022. R package version 1.0.1.

[12] Almquist Zack W. Large-scale spatial network models: An application to modeling information diffusion through the homeless population of San Francisco *Environment and Planning B: Urban Analytics and City Science.* 2020;47:523–540.

[13] Handcock Mark S., Fellows Ian E., Gile Krista J.. *RDS: Respondent-Driven Sampling.* Los Angeles, CA 2023. R package version 0.9-6.

[14] Gile Krista J, Handcock Mark S. Respondent-driven sampling: An assessment of current methodology *Sociological methodology.* 2010;40:285–327.

[15] Fellows Ian E. On The Robustness Of Respondent-Driven Sampling Estimators To Measurement Error *Journal of Survey Statistics and Methodology.* 2022;10:377–396.

[16] Salganik Matthew J, Heckathorn Douglas D. Sampling and estimation in hidden populations using respondent-driven sampling *Sociological methodology.* 2004;34:193–240.

[17] Baraff Aaron J, McCormick Tyler H, Raftery Adrian E. Estimating uncertainty in respondent-driven sampling using a tree bootstrap method *Proceedings of the National Academy of Sciences.* 2016;113:14668–14673.

[18] Gile Krista J, Johnston Lisa G, Salganik Matthew J. Diagnostics for respondent-driven sampling *Journal of the Royal Statistical Society. Series A,(Statistics in Society).* 2015;178:241.