Prevalence of Unsheltered Homelessness in King County, WA: Spatial Analysis of KCRHA Service Regions

Alejandro D. Hernandez, M.S.[a] and Zack W. Almquist, Ph.D., M.A., M.S.[b]

[a.] *Department of Biostatistics, University of Washington, Seattle, Washington*
[b.] *Department of Sociology, University of Washington, Seattle, Washington*

Report history: Last modified March 21, 2025
Keywords: Unsheltered homelessness; Respondent-driven sampling; Small-area estimation

A B S T R A C T

Purpose: Accurately estimating unsheltered homelessness is crucial for policy and resource allocation. Traditional Point-in-Time (PIT) counts have limitations, prompting alternative approaches like peer-referral respondent-driven sampling (RDS) to obtain a representative sample of unsheltered people. This study aims to leverage the strengths of the novel RDS method and classic spatial statistics to estimate the number of unsheltered people in jurisdictions within King County, Washington.
Methods: Using data from a multisite survey of 1,466 individuals, we applied spatial analysis techniques, including small-area estimation with Fay-Herriot models, to estimate the number of people experiencing unsheltered homelessness. To obtain a fair representation of our target population, our estimates rely on weights derived from the RDS method.
Results: Direct (RDS-II weighted) estimates produced similar results as the spatial and non-spatial Fay Herriot models.
Discussion: Traditional PIT counts often underestimate unsheltered populations due to visibility bias and logistical constraints. By incorporating RDS and small-area estimation, we improve accuracy and precision, yielding more reliable data for decision-makers.

Accurately enumerating the number of people experiencing homelessness is essential for policymaking, resource allocation, and assessing the effectiveness of intervention programs. In Washington state's King County, enumeration is federally mandated by the U.S. Department of Housing and Urban Development (HUD) to several "Continuum of Care" (CoC) jurisdictions. The tally is divided between individuals living sheltered and unsheltered. The sheltered population—those staying in emergency shelters, transitional housing, or similar facilities—are counted from administrative records. The unsheltered population, which includes individuals sleeping in public spaces, vehicles, or encampments, is considerably harder to measure due to their mobility, geographic dispersion, and lack of formal institutional ties. Historically, unsheltered homelessness has been estimated through point-in-time (PIT) counts, conducted by volunteers with flashlights and clipboards at night, sometimes known as a "street count" (HUD, 2014). This approach, while widely used, suffers from several methodological limitations (e.g., volunteers may disproportionately report individuals residing in visible and accessible locations). Constrained budgets and logistical challenges have restricted the development of more sophisticated methodologies, leaving many jurisdictions reliant on a crude and outdated approach.

Conflicts of interest: The authors have no conflict of interest to declare.
Address correspondence to: Dr. Zack W. Almquist, Department of Sociology, University of Washington, 211 Savery Hall Mail Code 353340, Seattle, WA 98195. *E-mail address*: zalmquist@uw.edu.

To improve estimation accuracy, researchers at the University of Washington developed an alternative respondent-driven sampling (RDS) method to enumerate the unsheltered homeless population, which relies on peer-referral to access hard-to-reach populations (). The method relies on a group of seed participants, who receive a limited number of referral coupons to recruit others from their social networks. This process continues iteratively, forming a chain-referral structure that expands the sample in a way that more effectively reaches socially connected yet geographically dispersed populations. One of the key advantages of RDS is its ability to account for the structure of social networks. Each respondent reports the size of their personal network—that is, how many other unsheltered individuals they know—which allows researchers to assign statistical weights. These weights follow the basic logic of Horwitz-Thompson estimators, adjusting for the likelihood that individuals with larger social networks are more likely to be recruited into the study and thereby reducing selection bias. Weighted estimates generated through RDS aim to approximate the true population distribution, providing a more representative picture of unsheltered homelessness than traditional PIT counts. Despite the advantages of RDS-based estimates, they, like other weighted estimates, are prone to high variance, particularly when sample sizes are small. Because statistical weights are derived from self-reported network sizes, errors in reporting can introduce additional variability. Furthermore, RDS relies on pseudo-random recruitment, meaning that sampling biases can persist if recruitment chains fail to reach certain subpopulations. These factors contribute to unstable population estimates, particularly when applied to small geographic areas where spatial heterogeneity and clustering effects are more pronounced.

To address these challenges, spatial statistical techniques may be incorporated to borrow strength from neighboring regions and improve estimate stability. By leveraging spatial dependence, which assumes that geographically proximate areas exhibit similar characteristics, estimates can be smoothed to reduce fluctuations while preserving meaningful regional variations. One approach to improving precision is to apply global and local smoothing techniques, which use data from adjacent areas to refine estimates. These techniques align with small-area estimation (SAE) methods, commonly used when survey data are sparse at the subregional level. SAE improves statistical power by pooling information across space, allowing for more stable and reliable estimates while maintaining local specificity.

However, spatial consideration introduces additional challenges. One potential source of error arises when respondents travel significant distances before being surveyed. If the site of individual's survey in a located far from where they primarily reside, their data may be misclassified to the wrong geographic region, distorting area-level estimates. To mitigate this issue, we may analyze self-reported travel patterns, including travel distance, time in transit, and modes of transportation, to assess potentially invalid assumptions.

This study aims to integrate RDS-based sampling with small-area statistical techniques to improve the estimation of unsheltered homelessness in jurisdictions within King County, Washington. Specifically, we leverage strengths of respondent-driven sampling to estimate the size and characteristics of the unsheltered population and incorporate spatial statistical techniques, such as global and local smoothing, to improve the stability of small-area estimates. Secondly, we analyze travel of respondents to their survey site, particularly the distance traveled to survey sites, time in transit, and modes of transportation, to assess potential biases in regional estimates. By combining innovative sampling methods with spatial modeling, this study seeks to produce more reliable and geographically precise estimates of unsheltered homelessness. These findings will support better-informed policy decisions and resource distribution for addressing homelessness at the local level.

Methods

Study design and data source

A multisite study was conducted to assess the experiences and needs of people experiencing unsheltered homelessness. Enrolled participants were provided coupons to share with peers as referral to the survey. Surveys took place at seventeen sites across King County, Washington. Individuals were surveyed between January 22, 2024, and February 2, 2024. CoC areas were defined by subregion, with two subregions being split into locales: The Seattle Metro subregion divided into two locales: Seattle and Vashon. The Urban Unincorporated King County subregion divided into three locales: White Center, Skyway, and South Park. Except for the Skyway and South Park subregions, each subregion contained at least one survey site.

Classification of unsheltered homelessness. Conditions of living were reported in the survey and described as housed, sheltered, or unsheltered.

Respondent demographics and self-reported travel to survey site. Demographics were self-reported by respondents. Self-reported travel to a survey site was measured by distance (miles), time (minutes and hours), and mode of transit (walking, bicycle, etc.). Travel distances under half a mile and under a mile were recorded as indicators to protect anonymity. These missing distances were imputed as half a mile and one mile, respectively, to assess a possible extreme situation. Travel speed was calculated to identify outliers and inconsistencies. Mode of transportation was reported as Bicycle/Bike, Bus, Car, Ferry, Link light rail (a train line), Walking, or Other, which allowed self-entry. Mode was further characterized as involving or not involving motorized vehicles (i.e., Bus, Car, Ferry, or Link).

Statistical analysis

Descriptive statistics were used to characterize the sample by demographic and housed or sheltered versus unsheltered. They were also used to summarize self-reported travel distance, time, and mode of transportation for the overall sample, then by survey site. Maps displaying summary statistics of travel distance for each site were created to visualize typical and maximum travel distance.

We desire small area estimates for each CoC subregion, based on respondents from survey sites located within subregions. Two Urban Unincorporated King County subregions, Skyway and South Park, had no survey sites/respondents. However, the Urban Unincorporated King County subregion White Center did, and because the three regions share the same neighbors and are relatively close in proximity, we elected to produce estimates from the White Center region and duplicate them to Skyway and South Park. In modeling, these regions are collectively referred to as Urban Unincorporated King County.

Area-level estimates were generated using the RDS-II estimator, which utilizes the RDS weights to directly estimate prevalence. Area estimates were also produced from two variations of the Fay-Herriot model, a classic model for SAE, which are described below. The ordinary model pools information across the entire space to smooth estimates toward a global value and the spatial model pools information locally to smooth estimates toward neighboring values.

Fay-Herriot model. The Fay-Herriot model is a classic area-level model used to obtain small area estimators (Fay & Herriot, 1979). It is typically formulated as a two-stage model: the first stage defines the sampling model and the second stage specifies the linking model. Fay and Herriot proposed modeling a transformation of the weighted estimator using a linear mixed model. Define θ_i to be the log odds (logit) of the weighted prevalence estimate of a binary outcome (e.g., unsheltered homelessness) for some area i, p_i^w, and V_i to be the estimated design-based variance of θ_i. Then, the model is defined as:

$$\theta_i = \text{logit}(p_i^w) = \log\left(\frac{p_i^w}{1 - p_i^w}\right) \qquad V_i = \text{var}(\theta_i)$$

$$\hat{\theta}_{\text{i}} | \theta_{\text{i}} \sim N(\theta_{\text{i}}, \hat{V}_i)$$

$$\theta_i = \alpha + \delta_i$$

$$\delta_i | \sigma_\delta^2 \sim_{iid} \text{N}(0, \sigma_\delta^2)$$

where σ_δ^2, the between-area residual variance, is estimated. In this basic Fay-Herriot model, the area-specific random effects δ_i are assumed to be independent and identically distributed (iid) between areas. In practice, the delta method is used to obtain appropriate variance estimates.

Spatial Fay-Herriot model. The Besag-York-Mollié (BYM) spatial model accounts for spatial correlation by assuming that observations in neighboring areas are more similar than those in distant areas (Besag, York, & Mollié, 1991). This model includes a spatial random effect that smooths estimates based on neighboring values and an unstructured exchangeable component that captures uncorrelated noise. A spatially linked model extends this approach by introducing spatially correlated area effects,. This model includes a spatial random effect that smooths estimates based on neighboring values and an unstructured exchangeable component that captures uncorrelated noise. A spatial linked model introduces spatially correlated area effects, b_i, using BYM2 specification of random effects, whose hyperparameters estimate the marginal variance across areas, σ_b^2, and the proportion of variance assigned to the spatial term, ϕ.

$$\hat{\theta}_{\text{i}} | \theta_{\text{i}} \sim N(\theta_{\text{i}}, \hat{V}_i)$$

$$\theta_i = \alpha + b_i$$

$$b_i | \sigma_b^2, \phi \sim \text{BYM2}(\sigma_b^2, \phi)$$

Covariates may be included in linear combination with coefficient in the θ_i term, for any Fay-Herriot model. Frequentist inference proceeds by integrating out the random effects to give

$$L(\alpha,\theta) = \prod_{i=1}^{m} \int_{\delta_i} p(y_i|\delta_i,\alpha,\theta)p(\delta_i|\theta_i)\, \mathrm{d}\delta_i$$

where θ represent variance parameters. This likelihood may be maximized using maximum likelihood or restricted maximum likelihood, with random effects estimates obtained via empirical Bayes.

Area-level estimates were generated using the RDS-II estimator and compared to IID and BYM2 Fay-Herriot models. Point estimates of prevalence and estimates of uncertainty were mapped and plotted in comparison. Between-area residual variances were interpreted for both Fay-Herriot models, and the proportion of spatial variance was interpreted for the spatial model.

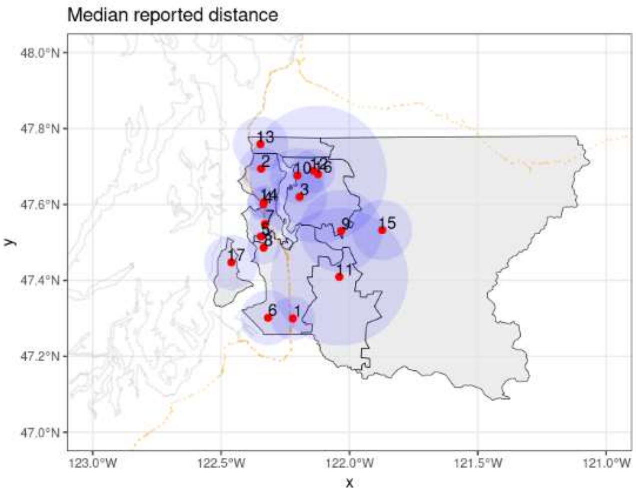
Results

Overall characteristics

A total of 1,466 individuals were surveyed between January 22, 2024, and February 2, 2024. 864 were unsheltered, 299 were sheltered, 191 were housed. Our analysis excluded 112 respondents with missing data regarding their living conditions. The sample was majority Non-Hispanic White, men, aged 45-54 (Table 1).

Table 1. Demographics of Survey Respondents.

	Overall (N = 1,354)	**Unsheltered (N = 864)**	**Housed or Sheltered (N = 490)**
Age			
18-24	57 (4.2%)	32 (3.7%)	25 (5.1%)
25-34	279 (21%)	196 (23%)	83 (17%)
35-44	394 (29%)	260 (30%)	134 (27%)
45-54	338 (25%)	202 (23%)	136 (28%)
55-64	235 (17%)	149 (17%)	86 (18%)
65 or older	47 (3.5%)	24 (2.8%)	23 (4.7%)
Did not respond or unknown	4 (<0.1%)	1 (<0.1%)	3 (<0.1%)
Race			
American Indian, Alaskan Native or Indigenous	56 (4.1%)	37 (4.3%)	19 (3.9%)
Asian or Asian American	22 (1.6%)	13 (1.5%)	9 (1.8%)
Black or African American	220 (16%)	117 (14%)	103 (21%)
Hispanic or Latino	206 (15%)	126 (15%)	80 (16%)
Multiracial	202 (15%)	125 (14%)	77 (16%)
White	553 (41%)	389 (45%)	164 (33%)
Other	75 (5.5%)	47 (5.4%)	28 (5.7%)
Did not respond or unknown	20 (1.5%)	10 (1.2%)	10 (2.0%)
Ethnicity			
Hispanic/Latinoax	298 (22%)	171 (20%)	127 (26%)
Non-Hispanic/Latinoax	1,046 (77%)	683 (79%)	363 (74%)
Did not respond or unknown	10 (0.7%)	10 (1.2%)	0 (0%)
Gender			
Man	966 (71%)	629 (73%)	337 (69%)
Woman	371 (27%)	226 (26%)	145 (30%)
Different identity	12 (0.9%)	6 (0.7%)	6 (1.2%)
Did not respond or unknown	5 (0.4%)	3 (0.3%)	2 (0.4%)



	Hub	Min	Median	Max
1	Arcadia Young Adult Shelter - YMCA	0.0	4.0	25
2	Aurora Commons	1.0	4.0	35
3	Bellevue Library	1.0	5.0	20
4	Compass Day Center	0.0	3.0	32
5	Family Phone Line	0.0	0.0	0
6	Federal Way Day Center - CCS	1.0	5.0	40
7	Georgetown Food Bank St. Vincent de Paul	1.0	3.0	30
8	Highline United Methodist Church	1.0	3.0	30
9	Issaquah Community Hall	3.0	7.5	20
10	Kirkland Library	0.4	5.0	25
11	Maple Valley Food Bank	3.0	12.5	30
12	Overlake Christian Church	1.0	4.0	15
13	Ronald United Methodist Church	0.5	5.0	20
14	Seattle Veteran Center	1.0	3.0	20
15	Snoqualmie Valley - YMCA	1.0	5.5	20
16	Together Center	3.0	12.5	20
17	Vashon Island Food Bank	1.0	5.0	15

Figure 1. Median reported distance illustrated as buffers (blue) across 17 survey sites (red). Primary boundaries illustrate 10 CoC subregions (grey) in King County, Washington. Illustrated is shoreline of the Puget Sound (light grey) and a major train line (yellow).

Self-reported travel distance, time, and mode of transportation

Travel distance was categorized as “Less than a mile,” “Less than half a mile,” or recorded in miles. Currently, summary statistics exclude 414 individuals encoded as “Less than a mile” or “Less than half a mile,” leading to missing distance data. Consequently, these statistics likely overestimate average travel distance, as individuals with shorter travel distances are omitted. Among the respondents, total reported travel distance to a survey site was approximately 6,400 miles. Measures of central tendency suggest a typical travel distance of around 4 miles (±2 miles). Approximately 45% of respondents traveled 3 miles or less, 67% traveled 6 miles or less, and 90% traveled 12 miles or less. Nearly all travel times were under 30 miles, with 95% of respondents traveling fewer than 20 miles. For reference, an average healthy adult walks approximately 3 miles per hour. Across the 17 survey sites, the median travel distance was around 5 miles, with a maximum of 12.5 miles at Together Center and Maple Valley Food Bank (Figure 1). The highest cumulative travel distances were reported from Aurora Commons and Ronald United Methodist Church (1,000 miles, equivalent to the driving distance from Seattle to Los Angeles), followed by Compass Day Center (800 miles, comparable to the distance from northern to southern California).

Travel time was reported in hours and minutes. Three extreme values (9, 10, and 30 hours) were excluded from visualizations for clarity but were included in summary tables. Collectively, respondents accumulated 27 days of travel time. Measures of central tendency indicate a typical travel time of approximately 20 minutes (±10 minutes). Around 50% of respondents traveled 15 minutes or less, 30% traveled between 15 and 30 minutes, and 16% traveled between 30 and 60 minutes. Nearly all travel times were under 2 hours, with 95% of respondents traveling less than 1 hour. A few respondents reported distances and times suggesting speeds exceeding 150 mph, likely due to data entry errors (e.g., reporting "2 minutes" instead of "2 hours"). Across the

17 survey sites, median travel time was usually around 20 minutes, with the greatest median of 45 minutes at Maple Valley Food Bank. The highest cumulative travel times were reported at Aurora Commons, Ronald United Methodist Church, Federal Way Day Center, and Compass Day Center, each accumulating 4 to 4.5 days of travel.

Modes of transportation included walking, bicycle, bus, car, Link light rail (a train line), ferry, and an "Other" category allowing self-entry. A total of 974 respondents (66%) used a motorized vehicle at some point, while 449 respondents (31%) traveled exclusively by foot or bicycle. The remaining 49 individuals were categorized as "Other," though their specific modes have not yet been examined. The most common transportation methods were bus (50%, 729 respondents), walking (30%, 446 respondents), car (16%, 241 respondents), and bicycle (2%, 29 respondents). When stratified by mode of transport, total reported travel distance was 5,681 miles for motorized travel and 554 miles for walking or biking. The median distance for motorized travel was 5 miles (IQR: 3–10 miles), compared to 2 miles (IQR: 1–3 miles) for walking/biking. Median distance for those that reported walking was 2 miles (IQR: 1-5 miles), as compared to those who rode a bus or took a car (both with a median distance of 5 miles and IQR of approximately 3-10 miles). Travel involving motorized vehicles accumulated 20 days of traveling, compared to 6 days by walking or riding a bike. Median time for motorized travel was 20 minutes (IQR: 14-40 minutes), compared to 10 minutes (IQR: 3-20 minutes) for walking/biking. Median travel time for those that reported walking was 15 minutes (IQR: 10-30 miles), as compared to those who rode a bus (30 minutes, IQR: 15-45 minutes) or took a car (15 minutes, IQR: 9-20 minutes). Overall, travel distance and time was deemed to not dramatically violate the reliability of our data. We continue to conduct spatial analysis.

Small-area estimation

Direct estimates of prevalence were calculated from the RDS-II estimator (Volz and Heckathorn, 2008). This method approximates the population proportion by weighting it based on a repeated-sampling model for RDS, assuming that the inclusion probability is proportional to the degree of each respondent. RDS-II weights were on average 6.6 and ranged from 1 to 20.

Table 2. Prevalence estimates and standard errors of unsheltered homelessness in CoC subregions of King County, Washington.

	RDS-II		**Fay-Herriot**		**Spatial Fay-Herriot**	
Area	*Estimate (%)*	*Std. Error*	*Estimate (%)*	*Std. Error*	*Estimate (%)*	*Std. Error*
East King County	11.0	0.018	10.8	0.016	10.9	0.017
North King County	17.1	0.013	16.9	0.013	17.0	0.014
Seattle Metro, Seattle	42.9	0.034	42.4	0.031	42.7	0.035
Seattle Metro, Vashon Island	2.5	0.008	2.7	0.009	2.7	0.008
Snoqualmie Valley	0.5	0.002	0.6	0.003	0.6	0.003
Southeast King County	1.5	0.010	2.1	0.014	2.2	0.014
South King County	20.4	0.034	20.0	0.031	19.8	0.034
Urban Unincorporated King County (White Center, Skyway, South Park)	40.0	0.021	4.4	0.023	4.3	0.025

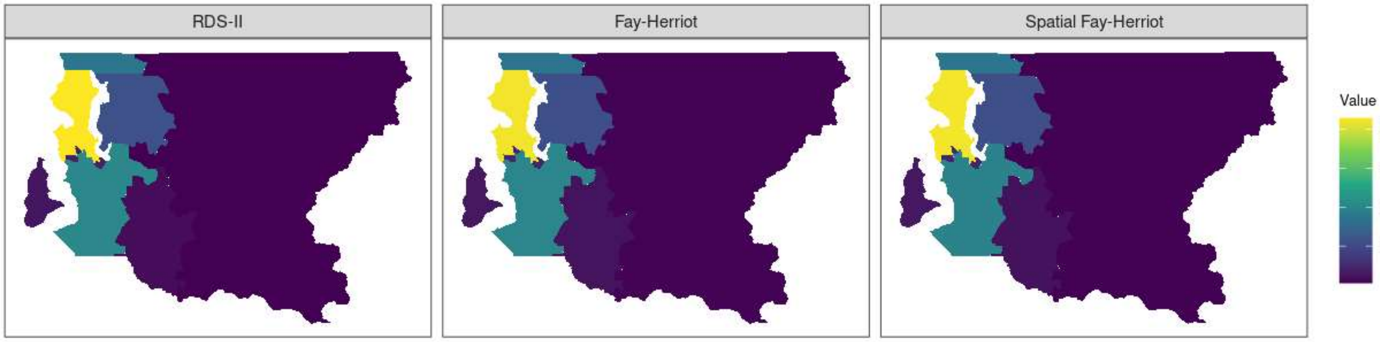


Figure 3. Prevalence of unsheltered homelessness in 10 CoC subregions in King County, Washington. Small area estimates were generated from a direct (RDS-weighted) estimator, a smoothing Fay-Herriot model, and a spatial smoothing Fay-Herriot model.

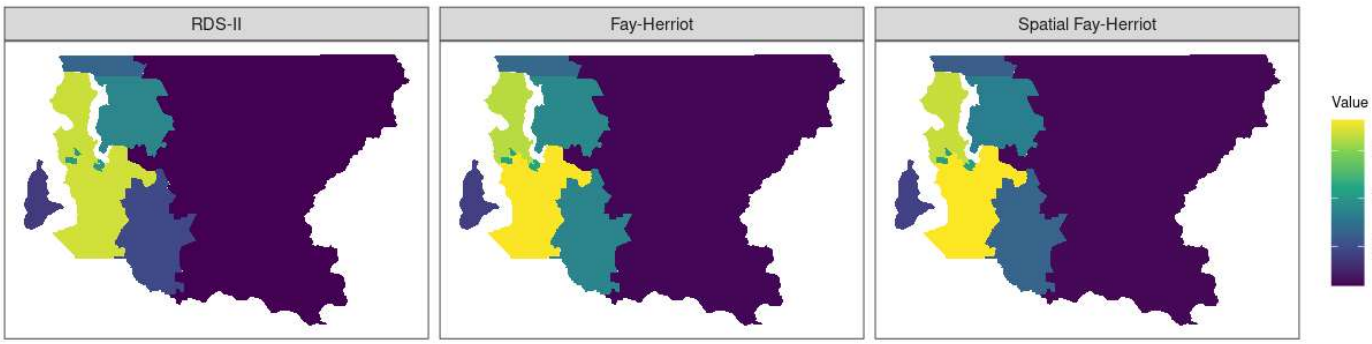


Figure 2. Standard error of prevalence estimates of unsheltered homelessness in 10 CoC subregions in King County, Washington. Small area estimates were generated from a direct (RDS-weighted) estimator, a smoothing Fay-Herriot model, and a spatial smoothing Fay-Herriot model.

Overall, the three estimators produced similar estimates (Table 2, Figures 2-4). Across all models, prevalence was lowest in Snoqualmie Valley and highest in Seattle Metro, Seattle. RDS-II prevalence for the 10 King County CoC subregions ranged from 0.48% to 42.9%, while Fay-Herriot estimates ranged from 0.63% to 42.2%, and spatial Fay-Herriot estimates range from 0.61% to 42.3%. The Fay-Herriot smoothed area prevalence estimates toward a global value of 6.8%, whereas the spatial Fay-Herriot smoothed estimates toward a global value of 6.2%. The spatial model estimated a mean variance of 1.49 across areas, with 25.9% of variance accounted for spatially. Estimates of uncertainty were generally similar across models, though both Fay-Herriot models produced slightly higher standard errors for Southeast King County. Notably, Fay-Herriot estimates did not significantly shrink direct estimates and, in fact, exhibited greater variance (Figure 4), contrary to expectations. This outcome may be attributable to the limited number of areas, which could prevent the typical shrinking effect on extreme estimates.

Discussion

PIT counts suffer from visibility bias, inconsistent volunteer coverage, and weather-related fluctuations. These factors result in underestimation of unsheltered homelessness. RDS improves representation by leveraging social networks, while small-area estimation stabilizes variance and accounts for geographic disparities. The spatial Fay-Herriot model further refines estimates by borrowing strength from neighboring regions, reducing

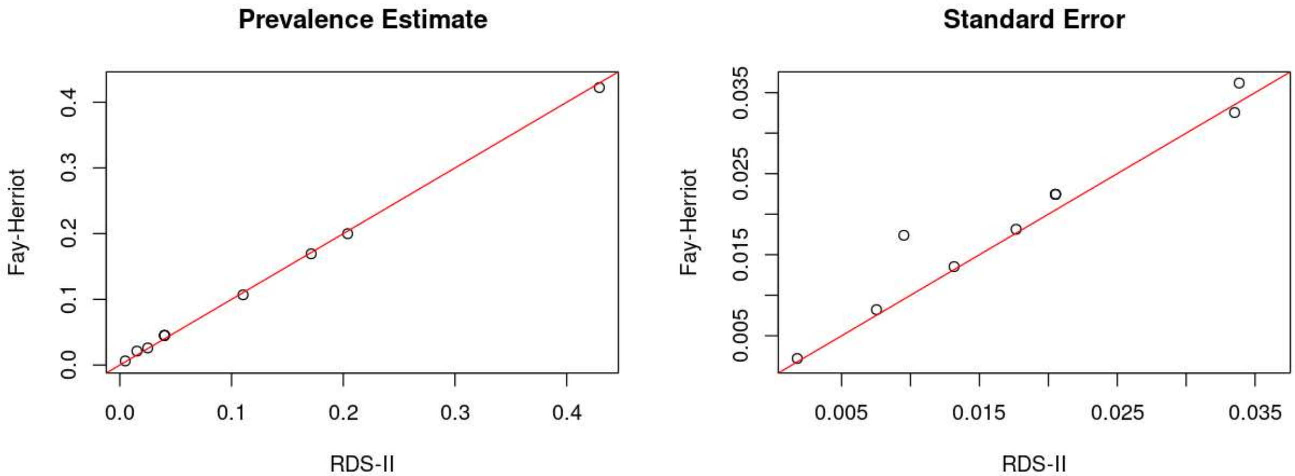


Figure 4. Prevalence estimates and standard errors of unsheltered homelessness in 10 CoC subregions in King County, Washington. Small area estimates were generated from a direct (RDS-weighted) estimator and a smoothing Fay-Herriot model.

local volatility. Although most respondents traveled reasonable distances, some outliers suggest potential misclassification. Incorporating travel data into small-area modeling may further enhance accuracy. Precise estimates offer better distribution of aid and services that more closely match regional need. Combining RDS with spatial models offers a scalable framework for estimating homelessness in other regions. Considering future research, longitudinal data collection may refine estimates and assess seasonal trends.

Conclusion. This study integrates RDS and small-area estimation to enhance the accuracy of unsheltered homelessness prevalence estimates. The results illustrate smoothing estimation and highlight limitations of spatial smoothing when there are few areas. These refined estimates may inform more equitable policy decisions and resource distribution strategies.

References

1. U.S. Department of Housing and Urban Development (2014). Point-in-Time Count Methodology Guide.
2. Almquist, Z. W., Kahveci, I., Hazel, M. A., Kajfasz, O., Rothfolk, J., Guilmette, C., Anderson, M.C., Ozeryansky, L & Hagopian, A. (2024). Innovating a Community-driven Enumeration and Needs Assessment of People Experiencing Homelessness: A Network Sampling Approach for the HUD-Mandated Point-in-Time Count. *American Journal of Epidemiology*.
3. Fay, R.E. and Herriot, R.A. (1979) Estimates of Income for Small Places: An Application of James-Stein Procedures to Census Data. Journal of the American Statistical Association, 85, 398-409. https://doi.org/10.1080/01621459.1979.10482505.

4. Besag, J., York, J. and Mollie, A. (1991) Bayesian Image Restoration with Two Applications in Spatial Statistics (with Discussion). Annals of the Institute of Statistical Mathematics, 43, 1-59. https://doi.org/10.1007/BF00116466.
5. Volz, E., & Heckathorn, D. D. (2008). Probability based estimation theory for respondent driven sampling. Journal of Official Statistics, 24(1), 79-97.