SIMTECH SOLUTIONS' PROPOSED METHODOLOGY FOR GEOGRAPHIC SAMPLING AND ENUMERATION FOR THE HUD POINT IN TIME COUNT

Larger Continuums of Care participating in the HUD Point in Time homeless census require a significant number of volunteers to effectively canvas the entire geographic area. Fortunately, there is HUD-supported alternative for regions that are unable to recruit sufficient volunteers: geographic sampling and enumeration. The functionality required to support this approach was developed with the guidance of statistician and University of Pennsylvania professor, Dan Treglia, PhD while working with the *Connecticut Coalition to End Homelessness (CCEH)* to conduct a statewide homeless count.

This document outlines the process implemented in Connecticut to derive a statistically reliable count estimates while only needing to canvas 40% of the census tracts within the state. This work is being shared as an overview of the methodology and to serve as a guide for other large CoCs interested in utilizing a similar approach to derive accurate count estimates.

While this paper is being shared as a guide, we encourage Count Administrators to also review the HUD guidance on *How to Use Sampling within a CoC to Conduct an Accurate Unsheltered Count* and to use the PIT help desk at PIT@SimtechSolutions.com if you need assistance.

PREPARING FOR THE COUNT

Sampling Strategy

The state of Connecticut is comprised of two Continua of Care (CoC): The Balance of State CoC (CT-505) and Fairfield County (CT-503).

A stratified random sample was employed to estimate the number of unsheltered homeless individuals in each CoC. For each CoC, CCEH and PIT Regional Coordinators designated each of Connecticut's 833 census tracts either "high" or "low" based on the probability of finding a person experiencing homelessness in that census tract. All high probability areas, and a random sample of low probability areas, were canvassed during the PIT count.

PIT Region	People Found in 2019	No People Found in 2019	Total
Bridgeport	4	62	66
Bristol	4	26	30
Danbury	1	49	50
Hartford	10	30	40
Hartford Suburbs	1	150	151
Litchfield	4	50	54
Meriden	2	25	27
Middlesex	5	31	36
New Britain	1	20	21
New Haven	10	21	31
New Haven East	1	23	24
New Haven North South	4	41	45
New Haven West	0	12	12
Norwalk	2	52	54
Norwich-New London	5	60	65
Stamford-Greenwich	5	35	40
Waterbury	12	38	50
Windham	4	29	33
Grand Total	**75**	**754**	**829**

Figure 1 – Census Tracts where homeless were found in 2019

Designation of the High Probability Census Tracts

For Connecticut, the high probability designations were based on results from the previous year's count and institutional knowledge from CCEH and PIT Regional Coordinators. A community may also choose to use locations gathered throughout the year by street outreach staff and/or information from the US Census to identify the census tracts with a high housing rent burden. Regions may also opt to use the "Known Locations Survey" that is an optional add-on to be used by community members to help identify areas known to contain people experiencing homelessness (PEH).

For Connecticut, Simtech used the results from the previous year's count to derive a total of 75 census tracts in which at least one person was counted – all of which were marked as high probability census tracts. This designation of the census tract as high probability is made regardless of whether that census tract was designated or sampled in the prior year, or if that census tract was not intended to be canvassed at all.

Review of the High Probability Census Tract Designations

Each PIT Region was configured by census tract coverage within the *Point in Time Regional Command Center*. PIT Regional Coordinators worked with CCEH in the Command Center to verify their region's coverage and use the functionality within the Command Center to overlay other geospatial data to automatically designate tracts as ones that are of a high probability to contain people experiencing homelessness. Geospatial data available to Regional Administrators to inform the designations include:

- PIT surveys from prior years,
- Known Location Surveys; (see this video);
- Interactions from Street Outreach teams using the Show The Way app;

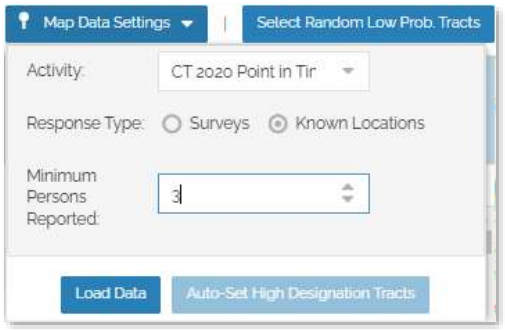


Figure 2 – Ability to Overlay Other GIS Data

The Count Administrator can designate High Probability tracts simply by clicking on "Map Data Settings" from the Sampling Setup tab. For Known Location Surveys, the Administrator can select a minimum threshold to be met for a tract is to be designated as High Probability. After clicking "Load Data" the pinpoints that correspond to the data are shown on the map. Once done, the administrator simply clicks "Auto-Set High Designation Tracts" to have the tracts flagged as High Probability which automatically sets them to be sampled.

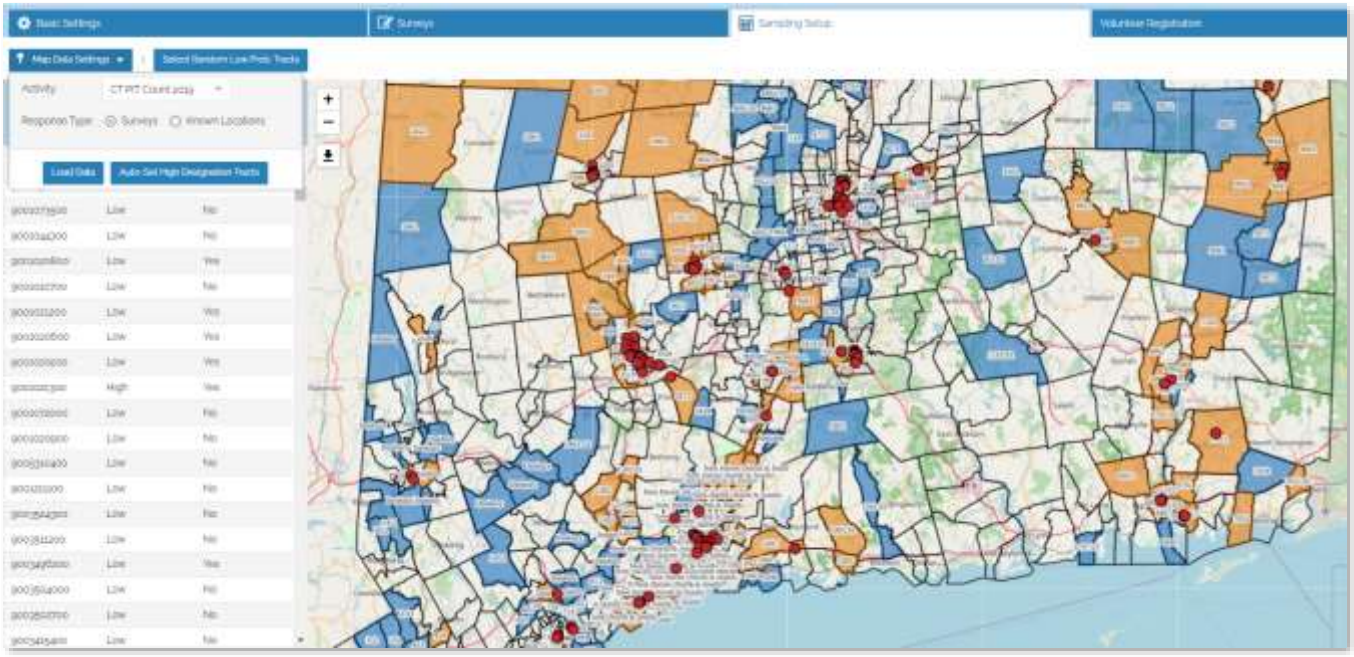


Figure 3 – Census tracts were designated as "high probability" by referring to locations from prior year's PIT count data and current outreach data. These are the areas in orange on the map.

PIT Regional Coordinators are subsequently able to add or remove high probability areas based on their institutional knowledge and information available to them through the usage of the region view in the Command Center (Region Management is an optional additional feature). If a Coordinator wanted to add census tracts, they provided justification to CCEH to ensure that the sample sizes did not exceed the capacity to recruit, train, and deploy enough canvassers. One hundred and eight (108) census tracts were added during this review process, which resulted in a total of 183 high probability areas to be surveyed. The HUD guidance on Sampling and Enumeration strategies refers to these tracts as "Certainty Areas."

Sampling Low Probability Census Tracts

The low probability census tracts to be canvassed are chosen through a random, computer-generated sample of all census tracts that were not previously designated as High Probability areas.

The Continuum of Care, rather than the PIT Region, was used as the basis for choosing the low probability sample to ensure that samples were large enough to be statistically valid. PIT Regions such as Meriden and New Haven East, for example (having twenty-seven (27) and twenty-four (24) census tracts, respectively), are insufficient in size for a statistically reliable sample. Small sample sizes can result in high variance and large confidence intervals, which would bring the reliability of the PIT count estimates into question. This is especially true for subpopulations (such as youth or veterans) where the small numbers expected to be counted could lead to artificially high or low estimates. Simtech Solutions' staff are available to ensure the sample size is sufficient to produce a statistically reliable estimate. For the State of Connecticut, there were 829 total census tracts, 183 of which were designated as High Probability. This left a total of 646 low probability areas to be randomly sampled, with 191 in CT-503 Fairfield County and the remaining 455 in the CT-505 Balance of State CoC.

Determining the Number of Low Probability Census Tracts to be Sampled

The sample size for each Continuum of Care was determined by Simtech and CCEH in consultation with PIT Regional Coordinators, with the goal of improving the accuracy of the PIT count by increasing the sample size, while understanding that the logistical constraints of finding, training, staging, and deploying canvassers across the state.

$$n_{LP} = \frac{1}{\frac{d^2}{N_{LP}^2 z^2 \sigma^2} + \frac{1}{N_{LP}}}$$

Simtech used the formula to the right to estimate the predicted precision of the count estimate, within each CoC, for any given sample size. Fortunately, there are free tools such as https://www.calculator.net/sample-size-calculator and https://goodcalculators.com/sample-size-calculator to help Count Admins determine the size of the sample that is needed. For Connecticut, the Confidence Level of 95% and a Confidence Interval of 10% were used. For CT-503 Fairfield County, this resulted in 64 of the 191 census tracts needing to be randomly sampled. This means 64 or more measurements/surveys are needed to have a confidence level of 95% that the real value is within ±10% of the measured/surveyed value.

CoC Details	
Total Census Tracts:	210
High Probability Census Tracts:	19
Low Probability Census Tracts:	191
Low Probability Sampled:	64
Low Probability Weighting Factor:	2.984

Figure 4- Sampling Set Up for CT-503 Fairfield County

CoC Details	
Total Census Tracts:	619
High Probability Census Tracts:	164
Low Probability Census Tracts:	455
Low Probability Sampled:	79
Low Probability Weighting Factor:	5.759

Figure 5 - Sampling Set Up for CT-505 Balance of State

For CT-505 Balance of State, this resulted in 79 of the 455 census tracts needing to be randomly sampled. This means 79 or more measurements/surveys are needed to have a confidence level of 95% that the real value is within ±10% of the measured/surveyed value.

The Role of the Weighting Factor in Deriving Estimated Count Figures

The *Low Probability Weighting Factor* is a simple calculation derived by dividing the total number of *low probability census tracts* by the number of *low probability tracts to be sampled*. For example, for CT-505 the weighting factor is 455/79 or 5.759. After all low probability areas that have been randomly selected have been canvassed, the count results from these areas are multiplied by the weighting factor to derive estimated count figures for all low probability tracts in the region.

Randomly Select Tracts in the Command Center

Based on the determination of the number of low probability census tracts to be sampled, the "Randomly Select Tracts" function within the Command Center was then used to determine which of the tracts are to be canvassed and which are to be ignored.

Figure 6- The Randomly Select Tracts Function in the Command Center

Define the Sub-Regions

From within the "Regions" tab within the Command Center, the Count Administrator can define sub-regions to break up the count activity into even more manageable components. Regional Count Administrators can be assigned to these sub-regions and have a restricted view into the command center that limits the volume of survey information to only include the surveys gathered by volunteers assigned to their designated geographic region within the CoC. All regions will be found under the "Regions" tab of the selected count activity. Additional guidance on how to use these features can be found within these instructions.

Recruit the Volunteers

For regions that opt for the Volunteer Registration Portal feature, the list of volunteers who have signed up will be displayed on the Users tab along with details of their preferred Region to volunteer in and key details such as their experience with the PIT count, ability to drive, and other information that is helpful when creating count teams.

Assign Volunteers to Count Teams

Within the sub-regions, Count Team areas were defined by the Count Administrator simply by clicking "Add Team," providing a team name, choosing the team, and then clicking on census tracts on the map to be covered by the team. The Administrator should consider the size of the area, the number of high probability tracts to cover, the number of low probability tracts to be sampled, the estimated number of people experiencing homelessness, and the estimated number of volunteers when choosing how many tracts to assign to a team. The image below is from the Bridgeport sub-region and shows how Count Team areas were established within Bridgeport. Volunteers who signed up through the Volunteer

Registration Portal were subsequently assigned to these Count Teams.

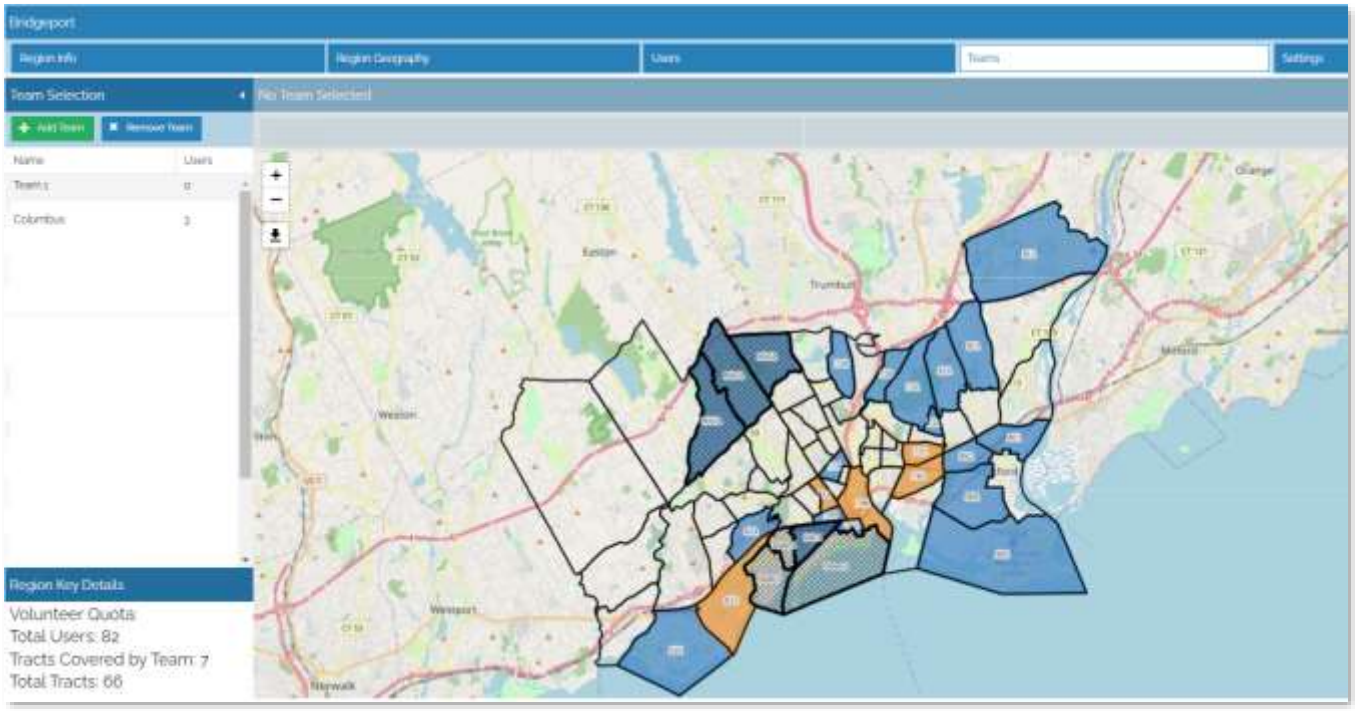


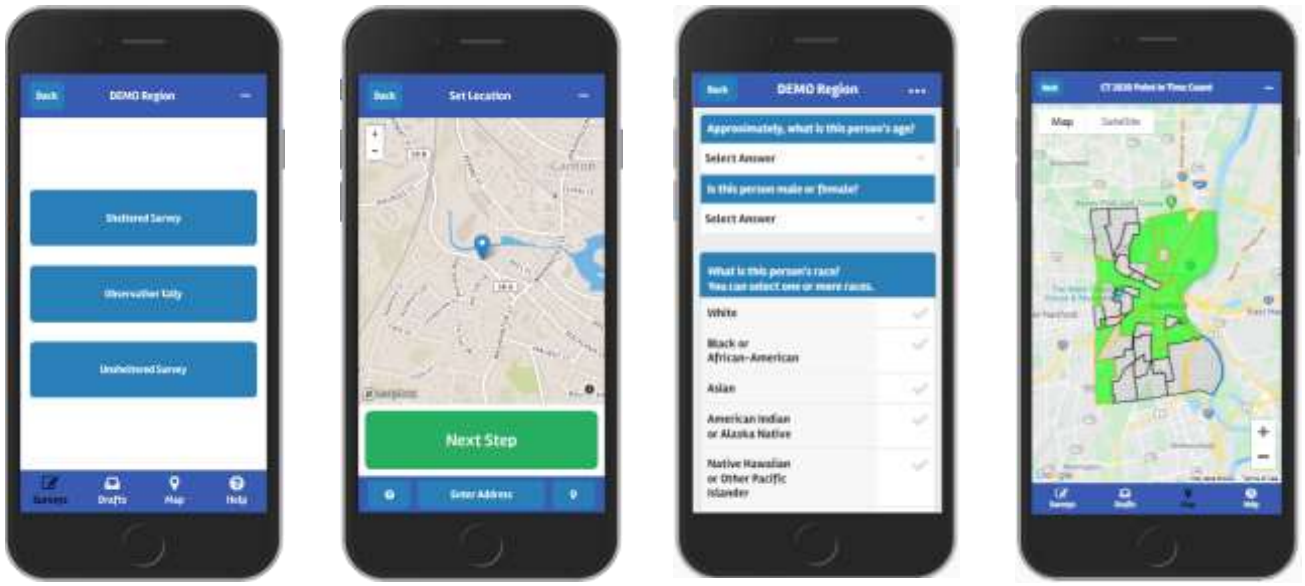
Figure 7- The Map-Based Interface Used to Create Regions and Count Teams

CONDUCTING THE COUNT

Utilization of Mobile Technology with Built-In GPS Capabilities

To conduct the surveys, volunteers download the *Counting Us* app from either Google Play or the Apple App Store, register an account, and join the appropriate count activity by entering the Setup Key attributed to the count project for which they are volunteering.

The survey questions found within Counting Us include such demographic information as age, race, and gender, as well as information on veteran status, disabling conditions, length of homelessness, and other questions that are included in the final PIT report submitted to HUD. A key feature of the Counting Us application is the built-in GPS functionality that pinpoints the exact physical location where each survey is conducted. If the CoC opts for the Region and Count Team Management feature, then the area the volunteer has been assigned to will be displayed on the map.



Monitor Count Activities in Real-Time

The *Point in Time Regional Command Center* receives survey data submitted by volunteers using the Counting Us mobile app, in real time. The map view from within the Command Center shows the location of each completed survey. Count Administrators can contact volunteers on their smart phones if surveys are conducted outside of the volunteer's designated count area.

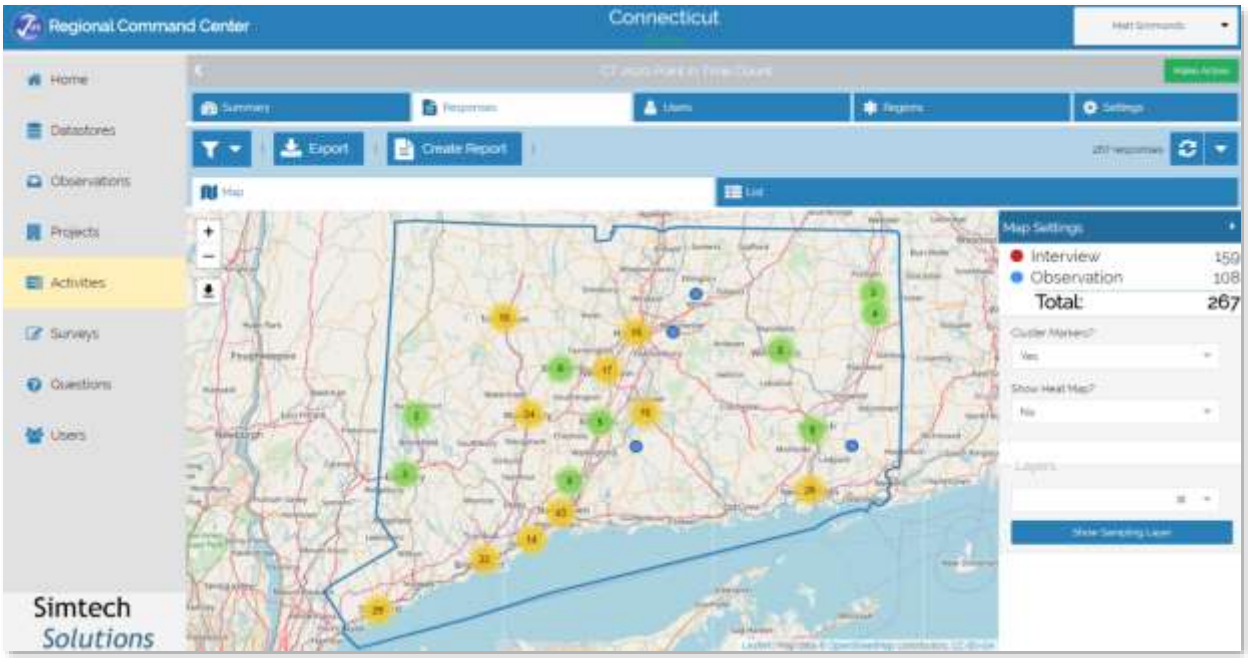


Figure 8 - Count Administrators watch the activities of Count Volunteers in real-time

Results are updated in real-time throughout the count and displayed on a dashboard found within the Command Center.

Figure 9 - Dashboard view of key demographic information collected by count volunteers

DERIVING THE FINAL COUNT ESTIMATES

Total Estimates for the CoCs and the State

Sampling weights, estimates, and confidence intervals are applied differently to high and low probability census tracts to produce final estimates. Since all high probability census tracts are to be fully canvassed, there are no sample weighting factors or confidence intervals applied. The weighting factor is applied to survey figures from the low probability census tracts to be sampled to derive estimated figures for all low probability census tracts.

To ensure the methodological rigor of the PIT estimate, only surveys from areas designated as high probability, or are within low probability census tracts to be sampled, are included in the results calculation. Surveys from outside of the boundaries of these designated tracts are not included in the results, as doing so would remove the randomness of the random sample.

As shown in Figure 10, the HUD Point in Time Report can be "Run with Sampling" simply by checking off the box shown before clicking "Run Report." An example of the results for the CT-505 CoC is shown in Figure 11. As demonstrated in Figure 5, the CoC has 619 total census tracts

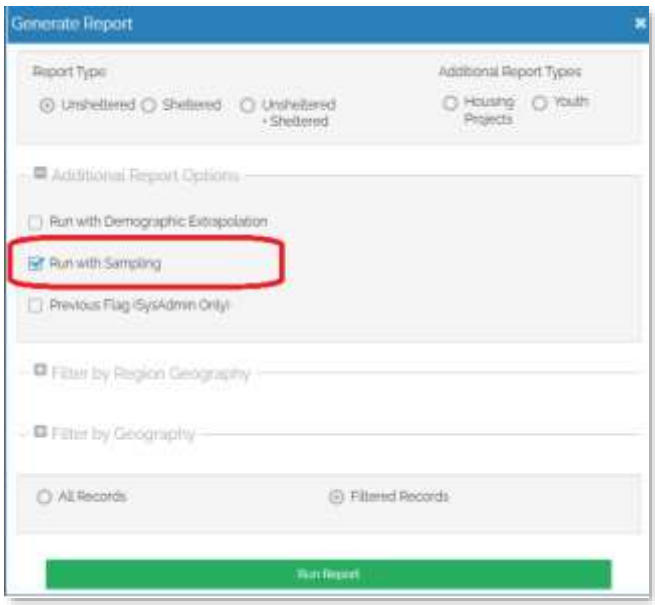


Figure 10 – Running the PIT Report with Sampling

with 164 designated as high probability and 455 as low probability. All 164 of the high probability census tracts were canvassed, along with a sample of 79 of the 455 low probability census tracts.

Household and Person Breakdown

	Region High	Region Low	Region Low Extrap.	Region Total
Total Number of Households	202	8	46.07	248.07
Total Number of Persons	205	8	46.07	251.07
Number of Young Adults (18-24)	5	0	0	5
Number of Adults (25+)	200	8	46.07	246.07

Figure 11 - Estimated Number of Homeless in CT-505 for the night of the 2020 Point in Time Count

After the count was completed, 205 people were counted in the high probability areas, and eight (8) were counted in the 79 sampled low probability areas. The number of people surveyed (eight (8)) was multiplied by the Weighting Factor of 5.759 (see page 4) to derive an estimated count of 46 people for all low probability census tracts. The Total Number of Persons estimated for the region was derived by adding the total counted in the high probability census tracts (205) to the estimate from the low probability census tracts (46), for a total of 251 persons.

Deriving Estimates in Sub-Regions (ONLY IF NECESSARY)

The State of Connecticut also requested estimates be provided for sub-regions within the state including PIT Region, Coordinated Access Network (CAN), and Sub-CoCs. The estimates produced were based on the number of individuals counted in high probability census tracts in that region as well the *weighted average number of individuals* estimated in the low probability sample. Complete HUD Point in Time Reports were produced within the Command Center using built-in reporting logic and geographic shape selections to define sub-region coverage.

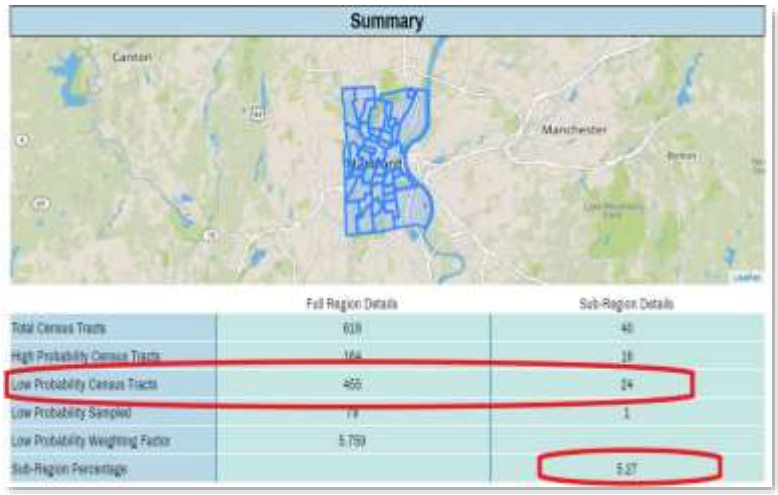


Figure 12 – Determining the Percentage of Low Probability Census Tracts within the Sub-Region

Since the number of census tracts within the sub-regions are an insufficient sample size for the purposes of estimations, the estimates for these sub-regions are based on the estimates derived at the CoC level. To accomplish this, the overall percentage of low probability census tracts within the sub-region must first be calculated. As shown in Figure 12, the Hartford PIT Region contains 24 low probability census tracts whereas the Balance of State CoC contains 455. Therefore, the percentage of low probability census tracts for the CoC that exist within the Hartford PIT Region is 24/455, or 5.27%.

Deriving the overall estimates for the sub-region is a two-step process. First, all count figures from the high probability census tracts are counted without any adjustments made. For the Hartford PIT Region, there were 13 people counted in the "high probability" census tracts. The second step is to derive the estimated counts for the low probability census tracts. This is done by multiplying the estimated figures

from the low probability census tracts for the entire CoC, by the percentage of all low probability census tracts that exist within the CoC that are from within the sub-region. The 8 individuals counted within the low probability areas were multiplied by the weighting factor of 5.759 to produce an estimated count of 46 people experiencing homelesssness within the low probability tracts.

The percentage of the low probability census tracts within the subregion (24) of all low probability census tracts within the CoC is 24 out of 455, or 5.27%. Multiplying the estimated count for low probability census tracts for the entire region by this percentage produces the total estimated count for the sub-region of 2.43. The count from the high probability census tracts (13) is then added to the estimated count from the low probability census tracts (2.43) to derive a final estimated count of 15.43.

Household and Person Breakdown

	Region High	Region Low	Region Low Extrap.	Region Total	Sub-Region High	Sub-Region Low Extrap.	Sub-Region Total
Total Number of Households	202	8	46.07	248.07	13	2.43	15.43
Total Number of Persons	205	8	46.07	251.07	13	2.43	15.43
Number of Young Adults (18-24)	5	0	0	5	0	0	0
Number of Adults (25+)	200	8	46.07	246.07	13	2.43	15.43

Figure 13- Deriving Estimated Counts for Sub-Regions

IN CONCLUSION

Geographic sampling and enumeration is an alternative strategy to conducting a full census and should be considered as an option for communities with insufficient volunteer resources. For this approach to be a viable one, the Continuum of Care needs to have:

- Enough census tracts to support a randomized sampling approach.
- The ability to determine areas within the CoC with a high probability of containing people experiencing homelessness.
- The willingness to ignore surveys conducted in areas that are not designated to have counts conducted within them.
- The ability to explain the approach and the rationale for the decision to use an alternative method.
- The willingness to expect a fluctuation in count figures when compared to any prior years when areas that were not canvassed were not accounted for with a process to derive estimates.

While this approach requires some upfront work, it has the ability to provide more reliable count estimates than a canvas-based approach in regions that lack sufficient volunteer resources. We would encourage reaching out to HUD Ask A Question (AAQ), and to the Simtech Solutions help desk, for any additional support or guidance that might be needed.