Parcel-Based Estimates of Population for Nonstandard Geographies:
Combining Data from the ACS and the County Assessor's Office to Create Accurate Estimates of Children and Adults

Katherine Nesse, PhD
City of Bellevue
450 110th Ave NE
Bellevue WA, 98004
knesse@bellevuewa.gov | 425-452-2042

Abstract. Current estimates of population are needed for assessing the capacity of fire, utility and transportation services and facilities, among other things. However, in many cases, the area of a city or county relevant to the assessment for a policy or project is not one that conforms to the standard set of geographies used by the Census Bureau, such as census tracts and block groups. Although most local governments do not conduct regular household censuses or surveys, every county must maintain detailed property records to assess property values for taxation purposes. Since characteristics of residential properties have a predictable relationship with the population living in those properties, one can use property characteristics to estimate population at the smaller parcel level geography and then summarize these estimates to the larger area of interest. Using the ACS, I have established the relationship between the number of bedrooms in a housing unit and the number of adults and children living in those units. I then apply that relationship to the number of bedrooms for each housing unit recorded in the county property records to estimate the average number of adults and children in that housing unit. I use Bellevue, Washington (in King County) as a case study for this method. I find that the estimate for total population for Bellevue less than half a percent different from the official population estimate calculated by the state of Washington for 2020. This method could be applied to other counties that record the number of bedrooms in the course of assessing property value to create current population estimates.

Introduction.

It is common for city planners and other government officials to need estimates of population in geographic areas that are not covered by the Census or American Community Survey (ACS). Often officials will approximate the area with census tracts. However, it is possible to use information about the housing units recorded in the County Assessor files to estimate the number of people living in each unit. Based on this the population in the nonstandard geographic area can be estimated.

The Nesse Method builds on the Housing Unit Method but uses some of the ideas of Aerial Interpolation to refine the method. It is based on the number of bedrooms in a housing unit since that is closely related to the number of people. Beyond that, it is also closely related to the number of children in a household. This paper outlines the method and demonstrates it using Bellevue, Washington, a mid-size city adjacent to Seattle, as an example.

Literature Review.

There have been many methods of using data from one set of geographic boundaries to estimate population in another set of boundaries. These fall into two primary categories. The first is aerial interpolation methods. This method uses population in a set of geographies which is used to estimate population in another set of geographies based on the overlap of the spaces. The other category, housing unit methods, is based on administrative records such as water or other utility hook-ups. This method counts households and uses a known multiplier such as household size to estimate the population. The Nesse Method in this paper falls into Housing Unit Methods category. It refines the basic Housing Unit Method by using information from overlapping geographic areas contained in the ACS.

Aerial interpolation methods became widely applicable with the use of computational cartography. Goodchild, Anselin and Deichmann (1993) laid out a framework for the method and applied to areas in California in a computer-based geographic information system. More currently, it has been used in tandem with land cover data to estimate both population and population density (Reibel & Agrawal 2007; Wu & Murray 2005). These methods have the advantage of covering large areas, especially those where there is little additional information. However, they are comparatively crude when there is more detailed information about the housing units on the ground.

Housing unit methods are the most common way of estimating population in small, nonstandard geographic areas (Smith 2012). This method is based on a count of occupied housing units and the average household size for the area. Occupied housing units has been counted in a variety of ways including, electricity or other utility hook-ups, building permits, tax roles, in the past telephone service has been used as well. Average household size or people per household usually comes from a survey. The decennial census has long been the primary source for this information because of its comprehensiveness. However, for some areas, particularly small and rural areas, the information in the census doesn't fit the particular geographic area of interest. Roe, Carlson and Swenson (1992) developed a "local expert" method that combined local knowledge with random sampling to determine average household size. Occupied Housing units times the average household size gives the household population. For total population, the population in group quarters is added based on administrative records of the college, prison or nursing home that administers the group quarters. Smith and Cody (2013) evaluated the accuracy of the method in 2010 and found it to be reliable in a variety of geographic types.

The Nesse Method is not the first to improve the Housing Unit Method with information or techniques from spatial interpolation. Deng, Wu and Wang (2010) evaluated the possibility of estimating both the number of housing units and the average number of people per household using remote sensing techniques. This method could be helpful in places where administrative records or reliable surveys are not able to provide these variables with enough accuracy. Tanton, Vidyattama, Nepal and McNamara (2011) did not use the housing unit method but did use a method of reweighting characteristics of the population based on the distribution of the population. The Nesse Method is the Housing Unit Method with more detailed information on housing units and households and incorporating additional information from the ACS, weighted by the number of housing units on parcels, to create population estimates for any grouping of parcels.

Method.

The basic method is to multiply the average household size by each housing unit. If there is more detailed information about the average household size and about the housing units, it is possible to estimate the population more precisely. Equation 1) shows the general method used in the housing whatever method. In this chapter, I will show step-by-step how this basic equation is developed with additional information.

$$POP = \sum_{i \in PIN} HHS * HU_i \quad (1)$$

In equation (1) POP is the total population of the geography. HHS is the average household size in the geography and HU_i is the number of housing units associated with each Property Identification Number (PIN).

The Nesse Method uses additional information about the property to apply a more specific average household size. Specifically, household size varies based on whether the housing unit is in a single-family building or a multi-family building and based on the number of bedrooms.

$$POP = \sum_{i \in PIN} \overrightarrow{HHS} * \overrightarrow{HU_\iota} \quad (2)$$

In equation (2), HHS is no longer a static average household size but a vector of average household size for the number of bedrooms in single-family and multi-family units. HUi is not the number of units on that parcel but a vector of the number of single-family and multi-family units by the number of bedrooms. For a single-family parcel, the vector would contain all zeros except one element would have a 1. For a multi-family parcel such as an apartment building, the vector would contain all zeros for the single-family elements but perhaps 20 for the 1-bedroom multifamily element and 10 for the 2-bedroom multifamily element of the vector.

Household size can be further disaggregated into adults and children. The relationship between the number of bedrooms and whether children are present is particularly strong for housing units in the 0- to 3-bedroom range. Few children live in 0-bedroom housing units and the average number of children per housing unit increases as the number of bedrooms increase.

$$POP = \sum_{i \in PIN} \vec{A} * \overrightarrow{HU_\iota} + \sum_{i \in PIN} \vec{K} * \overrightarrow{HU_\iota} \quad (3)$$

In equation (3), A is a vector of the average number of adults by the number of bedrooms and K is the average number of children by the number of bedrooms.

It is possible that the average household size varies across a geography irrespective of the number of bedrooms. In other words, a two-bedroom home in one neighborhood may, on average, house more people that of a two-bedroom home in another neighborhood. If it is possible to identify average household size by the number of bedrooms at the neighborhood level, then no adjustment is needed. However, for many places, the detailed information of average household size by the number of bedrooms is only available at a broad level – perhaps for the whole city or for the county. In that case, it may be useful to apply an adjustment factor.

Average household size can be calculated from ACS tables at the census tract level for both single-family and multi-family buildings. The adjustment factor is created by relating the census tract average to the larger area (such as the city) average as in equation (4).

$$\alpha_{tb} = \frac{HHS_{tb}}{HHS_b} \tag{4}$$

In this equation, the average household size, α, for census tract *t* and building type, *b,* (multi-family or single-family) is the ratio of the average household size, *HHS*, in census tract *t* and building type b to the average household size, *HHS*, of building type *b* in the city as a whole. These adjustment factors can be related to PINs based on the census tract where the parcel is located and the type of building on that tract. This scaler, α_i, can then be multiplied by the vectors of both the average number of adults and the average number of children as seen in equation (5).

$$POP = \sum_{i \in PIN} \vec{A}\alpha_i * \overrightarrow{HU_\iota} + \sum_{i \in PIN} \vec{K}\alpha_i * \overrightarrow{HU_\iota} \tag{5}$$

The final adjustment that is useful is adjusting for the occupancy rate. This may be constant across an entire city but more likely it varies by neighborhood. The occupancy rate, much like the average household size can also be calculated at the census tract level for both multi-family and single-family buildings using equation (6).

$$\gamma_{tb} = \frac{OHU_{tb}}{HU_{tb}} \tag{6}$$

In equation (6), OHU is the number of occupied housing units of building type b (multi-family or single-family) in census tract t. HU is the number of housing units in total of building type b in census tract t. The occupancy rate, γ, is related to each PIN based on which census tract the parcel is in and what type of building is on the parcel, just as the household size adjustment factor was. The scaler, γ_i, is multiplied by the vectors of the average number of adults and children using equation (7).

$$POP = \sum_{i \in PIN} \vec{A}\alpha_i\gamma_i * \overrightarrow{HU_\iota} + \sum_{i \in PIN} \vec{K}\alpha_i\gamma_i * \overrightarrow{HU_\iota} \tag{7}$$

Data Sources.

This method is possible if the number of bedrooms for each unit on a parcel is available from the county assessor's office. The rest of the data comes from the ACS either from tables available on data.census.gov or through IPUMS. Below are the specific data sources for each variable in equation (7).

Housing Units ($\overrightarrow{HU_\iota}$)

In many counties, the number of bedrooms is a characteristic used to assess the improved value for property taxes and therefore available administratively. Even within a state, many counties organize their assessment records differently and have different levels of access. Some counties may make these files publicly available in a spreadsheet or database format on an open data website, while others may not even keep electronic records of property assessment details and bedroom information may only be available through a public records request. In addition, not all assessment offices use the number of bedrooms in assessing the value of the property some use only the square footage of the building. It is possible to make these estimates across county boundaries but one should make sure that the assessment is done using the same methods, at least as far as counting housing units and bedrooms is concerned.

Adults ($\vec{A}$) and Children ($\vec{K}$)

The average number of adults and children by the number of bedrooms is not in any standard ACS table but it can be constructed at the PUMA level using IPUMS. To do this, people must be categorized as either adults (18+) or children (0-17) and then summed to the household level. Each household has an integer number of bedrooms associated with it and an integer number of units in the structure. The records should be divided into groups based on the number of bedrooms and the structure type.

Housing types are regionally specific. Some places consider row-house style development where units are attached but internally autonomous to be single-family while other places consider this to be multi-family. Some places do not have multi-family units that have in excess of 3 bedrooms. It is helpful to look at the number of records in each of the groups to be sure that there are enough records that the potential error is small. If there are only 5 or 10 records in a group, it is a good idea to combine it with other groups in a way that makes sense for that particular region.

For each group, the numerator is the number of adults or children and the denominator is the number of households. Because this characteristic is a household-level characteristic, the household weight would be used for all of these numbers.

The PUMAs used to make this calculation should match as closely to the study area as possible but they will not likely match exactly unless the analysis is being done at the county level. In some large cities that contain multiple PUMAs, it may be possible to estimate the average number of adults and children at a sub-city level, in which case, the household size adjustment factor (below) may not be necessary.

The data should be from the same ACS sample as the Household Size Adjustment and the Occupancy Rate. There are a few reasons for this. The first is that it is a consistent reference point and so the adjustments are not adjusting for a different sample but only for the specific characteristic. The second is that the 5-year sample has more records and therefore will contain less sampling error than the 1-year ACS.

Household Size Adjustment (α_i)

The adjustment to household size is a ratio of the average household size in a census tract to the average household size at the city level (or whatever the larger area is that is being used). There is not an ACS table with this information but it can be calculated using tables B25032: Tenure by Units in Structure and B25033: Total Population in Occupied Housing Units by Tenure by Units in Structure. Since this calculation is at the census tract level, it must be done using 5-year estimates.

Occupancy Rate (γ_i)

The Occupancy Rate is a ratio of occupied housing units to all housing units. It is not available by the number of units in the structure in an existing table but it can be constructed by combining two tables: B25032: Tenure by Units in Structure which only includes occupied housing units and DP04:

Selected Housing Characteristics which has total housing units by the number of units in the structure. Since this calculation is at the census tract level, it must be done using 5-year estimates.

Example Application.

The example application is the City of Bellevue in Washington State. It is a mid-size city adjacent to Seattle with a strong downtown. Housing units are about half multi-family and half single-family ranging in size from small studio apartments to sprawling mansions. As a mid-size city, it has many of the data availability constraints that smaller jurisdictions face. At the same time, it is large enough to contain many distinct neighborhoods. One advantage that the city has is that it is situated in King County which maintains detailed, publicly available assessment records. Because the data is electronically available it is possible, with time and effort, to shape the data into a format that can be used in this application.

The assessor data used for this was downloaded on February 18, 2020. The data from the ACS came from the 2014-2018, 5-year sample. The estimate of total population for the City of Bellevue made using the Nesse Method for April 1, 2020 was 148,487.

To check the results, it is necessary to have an outside estimate of the population that is independent of the inputs into this estimate. In Washington State, the Population Division in the Office of Financial Management creates estimates of population for each jurisdiction in the state as of April 1 each year. The method for making these estimates is entirely different are relies primarily of administrative records of birth, death and migration. It does use construction permit data from each of the jurisdictions as an adjustment factor so there is some slight overlap in the methods as the permit data is related to the county's assessor records. The official OFM estimate of population for the City of Bellevue for the same date was 148,100. The Nesse Method yielded a population total with a difference of 387 or less than half of a percent.

Since the Nesse Method appears to estimate the population in Bellevue fairly accurately, it is possible to break down the estimate to the parcel level and then build them up to neighborhoods, which do not follow census tract boundaries. This was particularly helpful for places that are experiencing rapid growth in housing. Other applications include population in the 8-minute fire engine access area, the area within a quarter mile of the Frequent Transit Network, police precincts, and school attendance areas. To estimate the population in one of these areas using equation (7), you can simply sum across that area instead of across the entire city.

It is tempting to make estimates of any area using this method however there are some things to keep in mind. First, household sizes are based on averages from a sample survey. If there are too few observations, the results could be nonsensical (such as 2.34 adults in a household and 0.02 children) or just not based on enough parcels for the sample to be valid. For Bellevue, I do not make estimates for any area that has fewer than 1000 parcels. The second thing to keep in mind is that because they are averages, if you were to, say, estimate the number of people in blue houses and green houses, you will get about the same answer if there are a similar number of houses. That is because there is no difference between houses except for the number of bedrooms and the census tract they are in. It works best when the area is contiguous.

Conclusion.

Using the relationship between bedrooms and household size, it is possible to estimate the population at the parcel level and then build the parcels up to the relevant geographic boundary. The equation is based on the Something Method. It uses data from the county assessor on the number of housing units on each parcel and the number of bedrooms in each parcel. The relationship between bedrooms and the number of adults and children is drawn for the IPUM sample of the ACS. Adjustments can be made to the average household size and occupancy rate using tables from the ACS at the census tract level. The resulting parcel-level estimates can be used to estimate population in geographic areas that are not standard in the ACS such as school attendance area, police precincts or transit networks.

Works Cited

Deng, Chengbin, Changshan Wu and Le Wang. 2010. Improving the housing unit method for small area population estimation using remote-sensing and GIS information. International Journal of Remote Sensing 31 (21): 5673-5688.

Goodchild, Michael F., Luc Anselin and Uwe Deichmann. 1993. A Framework for the Areal interpolation of socioeconomic data. Environment and Planning A 25 (3), 383-397.

Reibel, Michael & Aditya Agrawal. 2007. Areal interpolation of population counts using pre-classified land cover data. Population Research and Policy Review 26 (5-6), 619-633.

Roe, Linda K., John F. Carlson and David A. Swenson. 1992. A Variation of the housing unit method for estimating the population of small, rural areas: A casestudy of the local expert procedure. Survey Methodology 18 (1): 155-163.

Smith, Stanley K. 1986. A review and evaluation of the housing unit method of population estimation. Journal of the American Statistical Association 81 (394): 287-296.

Smith, Stanley K. and Scott Cody. 2013. Making the housing unit method work: An evaluation of 2010 population estimates in Florida. Population Research and Policy Review 32: 221-242.

Wu, Changshan & Alan T. Murray. 2005. A Cokriging method for estimating population density in urban areas. Computers, Environment and Urban Systems 29, 558-579.

Tanton, Robert, Yogi Vidyattama, Binod Nepal and Justine McNamara. 2011. Small area estimation using a reweighting algorithm. Journal of the Royal Statistical Society 174 (4): 931-951.