Census Block Population Estimates: Practice and Purpose

Washington State Office of Financial Management
Forecasting Division

Mike Mohrman
Geographer
Mike.Mohrman@ofm.wa.gov

Kyle Reese-Cassal
GIS – Demographer
Kyle.Reese-Cassal@ofm.wa.gov

Theresa Lowe
Chief Demographer
Theresa.Lowe@ofm.wa.gov

Manuscript prepared for presentation at the annual meeting of the
Population Association of America,
Los Angeles, March 30 – April 1, 2006.

Background

In 1942, Washington was the first state to develop city and town population estimates for program administration and revenue allocations. County projections were also done at an early date, with an established annual county population estimates program beginning in the late 1960s. Thus, Washington's demographic center, now seated within the Office of Financial Management (OFM), has been developing independent local area population estimates for over six decades.

OFM's population unit handles a tremendous amount and variety of requests for population data. The population unit provides comprehensive demographic information for government program administration and responds to requests from the legislature and executive branch for information needed for policy development, funding, and planning. The unit also serves as the key liaison to the Federal Census Bureau through the State Data Center (SDC) program and Federal State Cooperative for Population Estimates (FSCPE).

Population requests are often for "special" areas. These areas, such as health service districts, do not correspond with census geographies. Historically, special area estimates were done by hand on paper maps. Analysts reviewed the special district boundaries and compared them to census areas. The advent of Geographic Information Systems (GIS) has since replaced this cumbersome process. Using a GIS, OFM routinely merges special district boundaries with federal census blocks and estimates population using an areal interpolation scheme (Flowerdew and Green, 1994).

When working from the decennial census as a base, estimates from areal interpolation are likely to worsen as the time after the base census lengthens. For example, if Special District X accounted for 39 percent of the county's total population in 1990, then the 1998 estimate would be made by applying that same proportion, 39 percent, to the 1998 county estimate. The methodology, termed areal interpolation, is reasonable except for the underlying assumption that the population distribution was the same in 1998 as it was in 1990. Assumptions of this type are vulnerable to the "ecological fallacy," i.e., problems arising from assumptions about individual level decisions based on aggregate data (Gehlke and Biehl, 1934).

Experience shows that population change is sporadic and does not occur uniformly across space. The distribution of population and housing can change rapidly as new subdivisions, group quarters, or commercial structures are added to an area. For this reason OFM aims to capture population change at the most detailed level possible. Without some real indicator of the change in distribution, planning decisions are based on inappropriate and out of date information, thus succumbing to the ecological fallacy. The importance of OFM's population estimates further accentuates these dangers.

Because of the need for quality estimates, internal and external, OFM began developing a 'small area' estimates database in May of 1999. As intended, the database evolved into

the Small Area Estimate Program (SAEP). In short, this system creates block-level population estimates based on changes in the housing stock and group quarters since the last decennial census. Embarking on this tremendous and seemingly impossible six-year effort typifies the true spirit of applied demography.

Development of the Small Area Estimate Program

Data and Method

First, it should be acknowledged that block-level data are a challenge to work with due to: inaccurate census counts (Census 2000 Notes and Errata), the small size and often awkward shape of blocks, and the positional inaccuracy of the boundaries when compared to local geographic data sources. Despite the difficulties working with census blocks, block-level estimates are the best way to capture local change. Tracking growth on the smallest available geography allows disaggregate patterns to be analyzed. By increasing the spatial scale of analysis to census blocks, spurious zoning and aggregate effects are minimized. It is important to keep in mind that block estimates are not intended for use in and of themselves, but when combined they have been shown to produce reliable and accurate population estimates for small or special areas (Mohrman, Reese-Cassal, & Lowe, forthcoming).

OFM's long-standing city and county population estimate program is largely directed at the development of estimates for smaller areas. When first developing the database and estimate procedures OFM looked to regional planning agencies as a valuable source of data—or even as a source of small level estimates that could be combined with OFM's for statewide coverage. In Washington, the Puget Sound Regional Council (PSRC) serves the large Puget Sound counties containing 55 percent of the state's population. PSRC has been collecting building permit data and producing annual tract level population estimates with the Housing Unit method since 1971. Other planning agencies in the remaining metropolitan counties collect permit data and some produce estimates for their local areas as well.

Nearly all of the agencies use a variation of the housing unit method. New housing starts are added to the existing housing stock based on either geocoded building permits or new units from assessors' records. Data and estimates vary by the level of detail and geographic scale, but clearly other planning agencies would be an excellent source of housing data.

The importance of having good pre-existing working relationships and data sharing practices with many of the large regional planning councils and with county planning agencies is an important part in the overall success of this program. Tax revolt initiatives severely cut the local government resources over the past several years in Washington. Without these well-established relationships data collection efforts would be considerably more difficult.

OFM's established city/county estimate system collects the number of issued and completed building permits by structure type, the number of mobile homes, counts of group quarters by facility, and counts of special populations, from Washington's 281 cities. The information is provided for the city as a whole without spatial attributes. Similar data are collected for the unincorporated area of the state's 39 counties.

As part of OFM's role in determining the official state populations, the agency also collects, processes, and approves all incorporations, annexations, and de-annexations in the state. By statue, cities must conduct a door-to-door census of population and housing in all newly annexed or de-annexed areas. OFM approves and records all annexations; the population and housing counts they provide are an integral part of the annual city and county population estimate system.

If the SAEP is to utilize the official April 1 estimates as control figures, SAEP must compensate for boundary adjustments due to annexations and incorporations. This is an important issue because as cities change their boundaries, they do not follow census geographies. The SAEP compensates for these adjustments by combining census blocks into SAEP tabulation areas. SAEP tabulation areas are approximations of current city limits and the unincorporated balance of county. Like vintage 2000 census city limits, tabulation areas are made up of complete census blocks; each block is entirely inside or outside a jurisdiction. Each year, each census block is evaluated using a GIS and the available statewide city limit files. Each block is assigned a SAEP tabulation area based on the jurisdiction that comprises the highest percentage of the block's total area. These SAEP tabulation areas are used throughout the population estimation process as control areas.

Development of Group Quarters Estimates

Like other estimate systems utilizing the housing unit method, SAEP estimates household and group quarter populations separately. Group quarters are often a large source of population change for specific areas. Development and maintenance of a Group Quarters (GQ) database is a very important aspect of the SAEP.

SAEP's GQ database draws its data from the same database used in the agency's annual city/county population estimates program. There are approximately 1,000 facilities in this statewide database. Local government staff update a large part of the information each year in a standard reporting process. Forecasting staff update the information for specific types of GQ such as resident military personnel, college dormitory populations, state operated institutions and long-term care facilities.

A large advantage to having the SAEP and local government estimate systems integrated is that all GQ facilities receive a comprehensive evaluation each year. City and county staff provide annual GQ population and facility changes and then OFM analysts provide additional scrutiny on a regional assignment basis.

GQ facilities must be accurately geocoded to the smallest possible geography for a block level estimation system to accurately show population change. Great care must be taken to achieve correct placement of GQ facilities. Geocoding can be surprisingly inaccurate; the misallocation of large GQ facilities can lead to large, conspicuous errors. Even after address standardization and cleansing, OFM staff review and compare the locations of geocoded GQ facilities to GQ population counts reported by the census at the block level. In many cases the geocoded facilities were 'moved' or 'manually geocoded' to an adjacent block to be consistent with the Federal Census counts. Consistency with the Federal Census is important because of the methodology used to estimate a block's GQ population.

Group Quarter facility population is estimated by first subtracting the administrative count of the GQ population on the estimate date from the administrative count at the time of the last census. Then the difference, or change, is added to the GQ count as enumerated by the federal census in the appropriate block. In some cases, GQ administrative reports show a total population for the institution but the actual populations are distributed to multiple blocks. These types of institutions are usually colleges or military bases. Population changes for these facilities are distributed to census blocks based on the proportions of population counted in these facilities in the 2000 census.

Because of the difference between facilities reported counts and the 2000 census counts, it is mathematically possible for a block's population to fall below zero using this method. The program overrides any negative populations in SAEP with zeros, and places the unallocated change into a pool of temporally unallocated GQ population by tabulation area. This pool includes records that cannot be geocoded as well as records where no detailed geographic information is available due to confidentiality. After the initial block-level population adjustments are made, any population change that cannot be allocated to the appropriately identified census blocks is allocated on a proportional basis to all of the census blocks that contributed any GQ population to the SAEP tabulation area total GQ population. In some cases the jurisdiction as a whole will show a loss in GQ population; this loss is taken out of that jurisdiction's household population when it cannot be taken out of the GQ total population.

Development of Housing Unit and Household Population Estimates

While OFM is uniquely situated with a wide variety of population and housing data for cities and the unincorporated county, SAEP requires more geographically detailed data for accurate estimates of small areas. Like the regional planning agencies that were surveyed, OFM aims to collect new housing unit information at the parcel or address level of detail. Knowing precisely where new units are located is the best way to pinpoint population growth.

Address-level building permit data are available from regional planning offices and from some counties. In other counties, assessors can provide data on new housing units. In some areas no detailed data are available at all. Adding another layer of complexity,

different data sources operate on different timetables. While some jurisdictions report monthly, others may take up to two years to get detailed data to OFM. Issues with the timeliness of data were anticipated, and accommodations for gaps in data availability were devised.

To compensate for a lack of detailed data, postal delivery statistics from the United States Post Office are used as indicators of housing change. Although postal delivery stops do not equate one-to-one to housing units, (Lowe and Mohrman, 2003) marketing companies frequently use these data to help mold their population estimates. Postal data also have the advantage of being available on a timely basis for every state in the nation. As with building permits and assessor data, there are gaps in the area of coverage in postal delivery areas. In some of the most rural parts of the state mail delivery is limited to 'General or PO Box delivery'; the box holders come to the post office and pick up their mail themselves. In these cases knowledge is limited for a given Post Office service area.

Given the aforementioned limitations, the estimation of new housing units is based on a data dependent approach. The basic philosophy being; use the best data available for a given jurisdiction at a given time.

The total number and type of new and demolished housing units come from the existing city/county estimate system. This integration makes SAEP estimates consistent with the official April 1 estimates at the county level.

The distribution of city/county reported housing units to census blocks within the SAEP tabulation areas is based on a number of possible indicators of change. Geocoded building permits are preferred when available. In the absence of building permit data, assessor data are used. When assessor data are unavailable, postal delivery statistics are used to distribute the city or county level housing information to census blocks. If postal coverage is limited in the given jurisdiction, then housing counts from the federal census bureau are used as a fall back. *The SAEP is designed with the flexibility to take advantage of different data at different times.*

Very often geocoded building permit or assessor's data can be used as a reasonable means to distribute new housing, but there is usually a lack of geocoded demolitions. Subsequently, in many jurisdictions new units are allocated based on geocoded data, while demolitions are allocated based on the postal change or the federal census data.

After the housing units are allocated to blocks, a master occupancy and average Population Per Household (PPH) table is extracted from the city/county system. For each year, the change in occupancy and PPH is calculated on a jurisdiction-wide basis. Occupancy rates and PPH ratios are also calculated at the census block level based on the 2000 census data. The jurisdiction-wide change in occupancy rates and PPH ratios is applied to all census blocks within their appropriate SAEP tabulation area. If the mathematical result of adding the change in occupancy rate to the base census rate results in an occupancy rate above one, the occupancy rate is set to one. If the application of

change will result in an occupancy rate of less than zero, then the block rate is left at the base census rate.

For existing housing (i.e. units which are not new for the current estimate year) the number of occupied housing units is estimated by applying the new adjusted block occupancy rate to the existing block housing count. For new housing units, occupied units are estimated by applying the jurisdiction average occupancy rate for the appropriate housing type. This is done separately for existing housing and for newly added units because new units sometimes fall in blocks with little or no existing housing. The adjusted block rates do not apply in these cases.

Output from the housing unit allocation process is used to estimate household population. This is done separately for existing housing and for newly added units similar to the way the occupancy rates are applied to the estimate of housing stock. A population base estimate is made for existing housing by applying estimated block PPH ratios to estimated occupied units.

Population from new housing units is estimated using the average jurisdiction PPH ratio by structure type. In OFM's city estimates, adjustments are made for the occupancy rates in jurisdictions with large numbers of new units to reflect the lower occupancy rates associated with large new developments. A similar adjustment is made in the SAEP as well. High growth census blocks, those with more than 25 new units per year, have their occupancy rates decreased for new units by ten percent the first year to help compensate for the lower occupancy rates associated with high growth developments.

The total population for the jurisdiction is then calculated using the year's base population, the population estimated from new housing units and the group quarter population estimate. The difference from the control population is calculated and is then raked over the jurisdiction based on the occupied housing stock (base and new). When the raking process is complete, occupancy rates and PPH ratios are re-calculated at the block level based on the raked data. These values become the base data that feed the next year's population allocation process. The population allocation process starts over for the next year.

Summary

Census blocks are key to the SAEP program. By estimating at the block level very detailed patterns of population change can be analyzed. Without geographically precise estimates of the change in population distribution, planning efforts suffer because they assume a static pattern of growth.

Grouping blocks into SAEP tabulation areas allows for the full integration of the official city/county estimates as controls. The estimate program is not tied to the county as a control area, but breaks the county down into more meaningful parts.

Operating on county parts makes data collection and processing run more smoothly. Data collection efforts can be targeted at high growth or high interest areas. This also allows for limited requests for address level data from resource-poor jurisdictions, which already have difficulty providing summary level data for the official city/county system.

Boundary adjustments to SAEP tabulation areas on the block level enable us to make compensations for annexations and new incorporations. Maintaining current boundaries is an important part of the existing city/county estimates system. Between April 2, 2000 and April 1, 2005, 50,853 housing units, and 118,676 people have changed jurisdiction due to annexation/incorporation in Washington State (Population Trends, 2005). Data users appreciate OFM's ability to track boundary changes and maintain current population estimates that reflect those changes. This process preserves the integrity of the official April 1 estimates, therefore increasing the usability of the data across governmental agencies.

As the data are put to use they will be scrutinized by regional planners, applied demographers and others. This scrutiny will result in modifications to the program design and improvements in precision and accuracy.

It is hoped that these detailed estimates of change will allow for more realistic estimates of population for special districts and planning areas. This detail should prove to be extremely valuable for state and local government planning purposes and program management.

References

Flowerdew, R. & Green, M. (1994). Areal interpolation and types of data, pp. 121-145, in: S. Fotheringham & P. Rogerson (eds.), Spatial Analysis and GIS. London: Taylor & Francis Ltd.

Gehlke, C.E. & Biehl, K. (1934). Certain Effects of Grouping Upon the Size of the Correlation Coefficient in Census Tract Material, *Journal of the American Statistical Association* 29(185):169-170.

Lowe, T. & Mohrman, M. (2003). Developing Postal Delivery Data for Use in Population Estimates, Population Estimates & Projections 17:1-11. Olympia, WA: Washington State Office of Financial Management.

Mohrman, M. Reese-Cassal, K. & Lowe, T. (Forthcoming ~ 2007). Washington State's Small Area Estimate Project: Developing annual population estimates for ad hoc areas, *Population Research and Policy Review*.

2005 Population Trends (2005). Olympia, WA: Washington State Office of Financial Management.

U.S. Bureau of the Census (2000). Notes and Errata, 2000 Census of Population and Housing. Washington, D.C.: U.S. Government Printing Office.