

**Tracking Group Quarter Facility Data:
The Washington State Experience**

**Applied Demography Conference
San Antonio, Texas
January 2007**

Mike Mohrman
Geographer
Washington State Office of Financial Management
Forecasting Division
(360) 902-0602 - Phone
(360) 725-5174 - Fax
Mike.Mohrman@ofm.wa.gov

Abstract

The Washington State Office of Financial Management's (OFM) Small Area Estimate Program (SAEP) began publishing population and housing data for sub-county areas such as school districts, legislative districts, census tracts, and census block groups starting in December 2005. The SAEP draws data from OFM's official April 1 city/county population estimate program and distributes it to census blocks based on geocoded point data such as group facilities locations, housing starts, and polygonal data such as postal delivery statistics. Group quarters (GQ) are particularly important to both the city/county estimate program and the SAEP because changes in GQ populations can have a dramatic effect on local area estimates.

Group quarters data are inherently difficult to maintain. When facilities change ownership they are frequently renamed; they may merge with other facilities; and they can open or close just like any other business or institution. Facilities which serve multiple functions may mix GQ and non-GQ populations such as when nursing homes also serve as rehabilitation centers. The licensing of adult family homes frequently results in a change in use from "traditional" residential housing to small GQ facilities, muddling the accounting of both household and GQ populations. Group facilities are also surprisingly difficult to locate using a geographic information system. Administrative addresses are often reported for college dormitories, prisons, and other large institutions. The administrative offices are sometimes located far enough away from the physical location of GQs that the populations are incorrectly assigned to the wrong census tracts or blocks.

This paper focuses on how OFM collects, manages, and maintains GQ data for the city/county and SAEP programs. The primary topics addressed include data quality, data integrity, geocoding, the handling of multipart facilities, and various data elements that have proved useful in the estimation process.

Introduction

Group quarters (GQ) data are an integral part of many population estimate systems, including OFM's. The focus of this paper is to show how GQ data are used in OFM's Small Area Estimate Program (SAEP) and how the SAEP is intimately linked to OFM's existing city/county estimate program.

Washington State's city/county estimate program has been producing April 1 estimates of population and housing for cities and unincorporated parts of counties for state revenue distribution and planning purposes since 1942 — the first state program of this type in the nation. OFM's city/county estimates are developed using a combination of methods including the housing unit method, Component Method II, the ratio correlation method, and state certified local special censuses. Development on the SAEP began in early 1999. The SAEP was developed to support OFM's statutory estimates of population for special districts such as Nuclear Energy Benefit Areas, Justice Court Districts, and Federal Highway Urban Areas.¹ Special district estimates have been produced by OFM staff for a number of years using various areal interpolation methods (Packard 1967, 1968). With the availability of the 1990 census block data, the census block became OFM's preferred geographic unit to base the areal interpolation process upon. In the early 1990's areal interpolation was a labor intensive process in which OFM analysts would compare special district boundaries to census block maps and estimate the proportion of the population inside or outside of each special district by hand. The process could take weeks depending on the size and complexity of the area involved. One of the goals of the SAEP was to leverage the power of a geographic information system (GIS) to automate the areal interpolation process and reduce the amount of time required to produce the special district estimates.

Group Quarters Data Collection, Management, and Tracking

At the heart of both the city/county and SAEP estimate programs is an annual survey of population and housing. Early each spring OFM sends every city in Washington State a survey requesting the number of permits issued and completed for new frame housing by structure type, the number of mobile/manufactured housing units, counts of special populations (e.g., houseboats, travel trailers, and recreational vehicles), and counts of persons living in GQs by type of facility. Any changes in population and housing due to annexation or de-annexation are also accounted for in the estimates as well.

One of the goals of the SAEP was to build upon the success of the city/county estimates program. SAEP estimates are designed to match the official state and county estimates and mirror the official city estimates as much as possible. In this way, statistics based on SAEP estimates are comparable to statistics developed from official city/county estimates.

¹ See Appendix 1 to view the census area and special district estimates that are presently available for public download from OFM's website.

In order to integrate the two estimation systems, GQ data collected through the annual city/county survey had to be managed in an innovative way. Prior to 2000, the city/county survey questionnaire consisted of a blank worksheet (see Appendix 2). If the same staff person filled out the survey form each year, the data would often be consistent. Conscientious city staff would make note of any changes from the previous year such as a change in ownership or the renaming of a particular facility. In contrast, if a different staff member responded to the survey, OFM's personnel often found themselves looking at an entirely new or modified set of GQ facilities relative to the previous year. Any discrepancies in the GQ information had to be resolved by OFM analysts, usually by tracing the facilities back to the previous year's survey by address, phone number, facility type, or a combination of these attributes. In this way, the analysts could verify that no facilities were missing from the current year's survey and make sure that any additions to the city list were truly new, legitimate GQ facilities. For small cities the annual data cleansing process could be a simple task, but for larger cities with numerous GQ facilities, the process was usually quite time consuming.

In an attempt to improve the quality and consistency of the survey data, OFM developed a database capable of tracking individual GQ facilities over time. Under the new system, each facility is assigned a unique identification number. The identification number is tied to the building or institution, regardless of whether the name or ownership changes. For each unique GQ, the survey information is entered into the GQ database including facility type, street address, ZIP Code, contact phone number, and reported population. This information is remitted back to cities each year as part of their annual survey form (see Appendix 3). By returning this information to the cities, OFM helps the cities maintain a consistent list of facilities. This saves city staff members considerable time since they no longer have to refer to the form from the previous year and reenter the GQ information by hand. City staff can make updates or corrections by simply crossing off existing information or by entering information about new GQs since the time of the last survey. Under this new system, any work by OFM analysts related to address cleaning and standardization is not lost because the corrected address information is returned to the cities the following year. Furthermore, if a facility closes, the current population is reported as zero so the facilities do not completely disappear from the annual survey form without cause.

New GQ facilities have to be explicitly documented in the 'New Group Quarters Facilities' section of the survey form. The city is asked whether this is a completely new structure or a conversion of an existing building. If the facility is new (e.g., a new structure intended for use as a group home), the OFM analyst must confirm with the city that they did not also report the facility in the "New Frame Housing" section of the survey questionnaire. It is important to make sure that it is not counted both in the new housing stock and GQ categories. If the facility is not new, and it appears that it is the result of a conversion of an existing structure or the city believes the facility was inadvertently left off the prior year's list, then the city analyst has more investigative work to do.

The OFM analyst usually begins the data validation process by comparing current information with that from the previous year. Sometimes, a facility will reappear on a city's list every few years even though OFM and city staff agreed that the facility would be considered housing stock and not a GQ. If the facility has been investigated before, the analyst will call the facility to verify whether there has been a change in use that would warrant moving the facility out of the housing stock and into the GQ category. If the facility has not been investigated before, the analyst must call the facility and ask a series of questions which can vary depending on facility type. Determining whether or not the facility was counted in the 2000 census is a common starting point. If the facility was in operation at that time, the analyst will try to determine if the census counted the facility as part of the housing stock or as a GQ. The analyst must also make sure that the address reported by the city is the true site address and not the address of an administrative office. Once the site address is determined, the analyst will need to pinpoint the 2000 census block that the facility is located in. This is accomplished using a variety of sources including online map services, GIS, paper maps, or even driving directions from facility staff. The analyst will examine the 2000 block data and try to determine how the facility was counted in the census. The analyst will either accept the facility as a legitimate GQ or present the city with information that the building should be counted as part of the housing stock. If the facility began operating after the year 2000 and is not considered a new structure, the OFM analyst must ask what the building was used for before it became a GQ. Establishing whether a building changed ownership or operated under more than one name helps determine how to properly count the facility. If the facility is truly the result of a conversion from existing housing stock to a GQ, then the analyst must find out the number of units that were converted. The number of converted units will be subtracted from the city's housing unit count in order to prevent the double counting of people and housing units. Maintaining the integrity of the housing stock is vital to the city/county estimate process.

OFM staff attempt to derive long-term stay counts for facilities with highly transitory populations, such as jails, whenever reliable information is available. Most of the city and county detention centers in Washington are short-term stay facilities. Persons with longer sentences are typically housed in state or federal institutions. Detailed information about local jail populations is difficult to come by since jail data systems are not designed with the needs of demographers in mind. OFM staff recently received some data from Snohomish County for their primary detention facilities and work release programs in the City of Everett. Preliminary analysis of the data yields important information related to the counting of persons at local correctional facilities. Table 1 shows the number of persons released in the month of April by year and the number of releasees who were incarcerated less than 72 hours.

Table 1. Snohomish County Jail and Work Release Programs - April Releases

<i>Year</i>	<i>Total April Releases</i>	<i>Persons Incarcerated < 3 Days</i>	<i>Percent of Persons Incarcerated < 3 Days</i>
2000	1,531	916	59.8%
2001	1,654	963	58.2%
2002	1,772	924	52.1%
2003	1,780	981	55.1%
2004	1,849	1,126	60.9%
2005	1,644	968	58.9%
2006	2,009	939	46.7%
Total	12,239	6,817	55.7%

Source: Snohomish County Data Systems, SnoCo Jail Analyzer.

Snohomish County has also provided information about the long-term jail population in the City of Everett. The long-term population count is defined as the number of persons incarcerated for 182 continuous days or more on April 1st. The information contained in Table 2 indicates that nearly five percent of the total jail population can be considered long-term incarcerates. The county does not collect information on whether or not detainees have a permanent place of residence. As such, it is not possible to count persons without a usual place of residence as part of the legitimate jail population. Nonetheless, the information shown in Tables 1 and 2 make clear that these facilities primarily serve short-term stay populations.

Table 2. Snohomish County Jail and Work Release Programs - April 1st Population Counts

<i>Year</i>	<i>Total April 1 Population Count</i>	<i>Long-Term Population Count</i>	<i>Percent Long-Term Population</i>
2000	836	31	3.7%
2001	1,011	51	5.0%
2002	1,111	36	3.2%
2003	975	40	4.1%
2004	932	44	4.7%
2005	866	33	3.8%
2006	1,270	103	8.1%
Total	7,001	338	4.8%

Source: Snohomish County Data Systems, SnoCo Jail Analyzer.

Given the short-term nature of incarceration in local jail facilities, persons could be enumerated as part of the GQ jail count and at their usual place of residence. The jail and work release facilities in the City of Everett are located in the same census block. The Census 2000 GQ count for this particular block is 516 persons—a difference of 485 persons compared to the Snohomish County’s year 2000 long-term stay count. Part of this difference may be explained by the timing of the census count versus the timing of the jail count, but as this example demonstrates, the potential for double counting is quite serious.

Accurately counting the GQ population associated with health care facilities also proves cumbersome. The process is complicated by issues surrounding the classification of service types. Many health care facilities offer a variety of services. They may be licensed to provide both short-term beds for patient rehabilitation and long-term beds for patients who require constant care. A nursing home may offer full-time care services, independent living apartments, or a variety of services in-between. The census taker, city official, and facility staff member are likely to be confused about the specific populations to include in a GQ count. OFM analysts provide guidance to cities to help with hard to classify institutions such as these. The analyst will often call facilities to determine which beds should be included in the facility count and relay that information back to city staff for future reference.

Accurate and consistent data are crucial for maintaining the integrity of housing and GQ counts. Any issues that arise are best addressed early on in the process since incorrect counts can lead to errors that propagate over time. Collecting and cleaning GQ data is time consuming work. There are few shortcuts other than leadership on the part of OFM and better communication between all of the parties involved.

Geocoding Group Quarters

The geocoding of GQ addresses is a time-consuming, labor intensive process. Addresses can be quite complex and typically include some combination of house number, unit number, street pre directional, street name, a street type, street post directional, rural route number, post office name, and ZIP Code. Reported GQ addresses may consist of post office box numbers, references to building names, institution names, or even street intersections. Before accurate address matches can be made, all of these various elements must be consistent with regards to spelling and abbreviations; structured in the proper order; and stripped of non-standard punctuation and extraneous information.

Modern GIS and geocoding products make the task of address matching quite easy. Thousands of addresses can be geocoded with just a few mouse clicks. Some software vendors try to improve match rates by standardizing and correcting address elements on-the-fly. They might allow commonly misspelled street names to be matched (e.g., 'Mane St' matches to 'Main St'). They may change street directionals; 'NW Main St' to 'NE Main St', or even change street types (e.g., 'Road' to 'Street' or to 'Avenue', etc.). Most address matching software will then score the reliability of the resulting match based on changes made to the input address and to the assumptions made by the software. The end user has to determine, based on the match score, whether the quality of the match suits his/her needs.

Address quality is very important to the success of the address matching process. Most organizations that geocode at the statewide or national levels focus on input address quality because they have very little control over the quality and accuracy of the geospatial files they are being matched to. In practice it can take years for new streets to be included in national level files. This is especially true for addresses in rural areas.

Accurate estimates of GQ populations at the block level require accurate geocoding. The well known adage of ‘garbage in equals garbage out’ applies here but with geocoding it is not simply a matter of misspellings and transposed numbers although these errors often result in bad geocodes. There are many non-standard, out of sequence addresses that cannot be accurately geocoded through batch processing.

Larger institutions — especially colleges/universities, state run hospitals, and correctional facilities — will often have their own campus mail system. This means that the U.S. Postal Service (USPS) will deliver all mail addressed to the institution to a single location, and the campus mail system takes over from there. This simplifies mail delivery for the USPS and may improve delivery from the perspective of the institution but it confounds most geocoding software. The problem centers on the fact that as soon as the mail leaves the USPS system, the address is no longer required to conform to USPS standards. Sometimes a large institution may have a “street style” address, which is the intake location for the internal mail system. The address is essentially an administrative address rather than an address which represents the actual location of the facility. An administrative address that looks like a street address will be treated as such by the geocoding software, often resulting in a geocoded location for the GQ facility that does not accurately represent the true spatial location of the facility. If the central delivery address is not a standard street style address, then the geocoding software will likely produce an error code or result in an un-matched record. In both cases, the results are unsatisfactory. The McNeil Island Corrections Center in Pierce County, Washington is provided as an example. As the name suggests, the correctional center is located on an island. Other than the state prison, McNeil Island has little developed land and no industry to speak of. The correctional center has both a standard mailing address (1403 Commercial Street) and a P.O. Box address (Box 88900) with the same city, state, and ZIP Code (Steilacoom, WA 98388). The facility does not have standard street delivery and although the mailing address reads like a standard street address it is really the intake point for the prison mail system. If the analyst is not aware of these special circumstances, they might improperly geocode the facility based on the mailing address. The McNeil Island Corrections Center address can be geocoded (i.e., without errors or warning messages from the geocoding software), but the resulting location is approximately 3.6 miles from the actual location of the facility (see Figure 1).

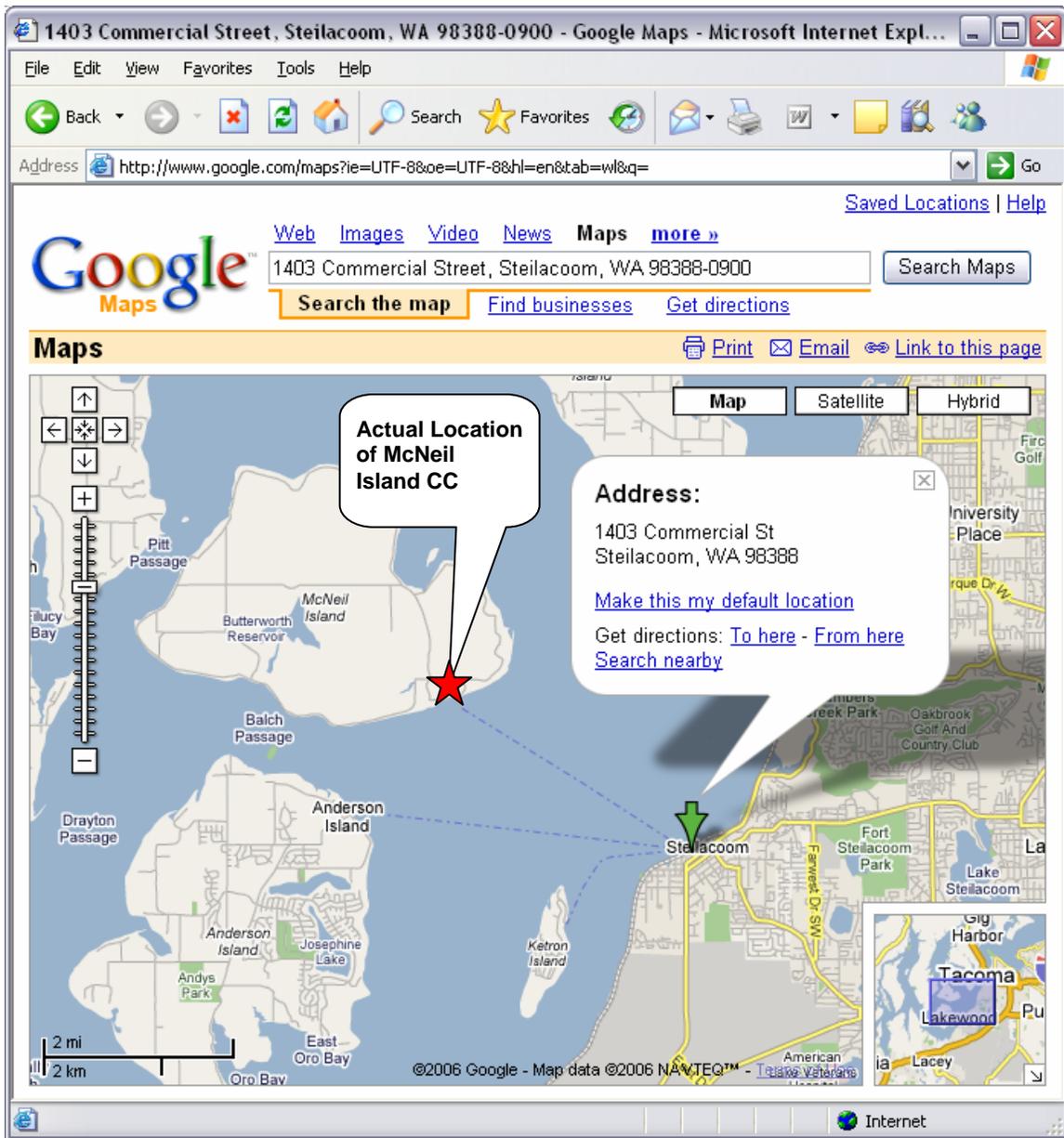


Figure 1. McNeil Island Corrections Center Geocodes

The geocoded address is located in the small town of Steilacoom, the source of the facility’s mail delivery system and the only source of ferry service to the island.

Over reliance on geocoding software products which attempt to correct misspelled or non-standard address elements can result in bad geocodes. Further complicating matters, the updating of roads and address attributes is neither random nor comprehensive. There is usually a pattern to the update frequency based on counties or cities with unusual amounts of building activity such as the Census CAUS program (U.S. Census Bureau 2006). This sometimes means that areas with relatively small amounts of new

construction may be passed over until there has been enough change to merit a region-wide update.

In preparation for the 2000 federal census, OFM staff started to geocode GQs after the 1999 city/county estimate cycle was complete. Facilities that were poorly geocoded or could not be successfully matched were contacted by OFM staff in order to collect additional address information that would improve geocoding accuracy. Many facilities were missing key address elements with ZIP Codes being the most common omission. Internet-based address lookup systems were utilized whenever possible to populate the database. After the final round of geocoding, a comprehensive review of the geocoded facility locations began. The first stage of the review process focused on large state institutions such as colleges, universities, prisons, and state run hospitals. These institutions feature many of the geocoding stumbling blocks described previously. Facilities were then checked to make sure that the geocoded locations were inside of the reporting city/county boundaries as well.

With the release of Census 2000 data, an additional review of the geocoded GQ facilities began in earnest. All facilities with a reported population over 25 persons were reviewed over a three year period. OFM's overarching rule was that GQ facility locations needed to be consistent with the federal census GQ blocks to the largest extent practicable. This constraint was necessary because, according to our methodology, any GQ populations not in census blocks in the year 2000 would appear in the SAEP estimates as new GQ populations. Consistency with the federal counts was considered so important that facilities with seemingly good geocodes were sometimes moved across a street or block face so that they were positioned in a census block with a similar type and size GQ. Throughout the review process, OFM staff made use of tax parcel and building permit information provided by local agencies as well as aerial imagery to reposition facilities. Local data has become more accessible over time, helping to improve the positional accuracy of the geocoded facilities over the course of the decade.

The use of locally derived data produces another set of issues however. Partway through the review process it became apparent that the positional accuracy differences between local data sources and the TIGER/Line data was often so great that georeferencing facilities using locally produced (and more accurate) data such as county assessor's records or GIS tax parcel layers would sometimes result in the placement of a facility outside of the proper 2000 census block. At first these discrepancies were noted and staff simply moved the facility to a representative location inside the appropriate block. Eventually it became apparent that this situation was recurring frequently enough that it needed to be tracked in a consistent manner. To help track the most accurate facility location, OFM added several new database fields to hold the parcel ID number and the positionally accurate coordinates. The two coordinate fields are populated only when the difference in alignment between the local data and the TIGER data is such that using the local data would place the facility outside of its proper census block (see Figure 2).

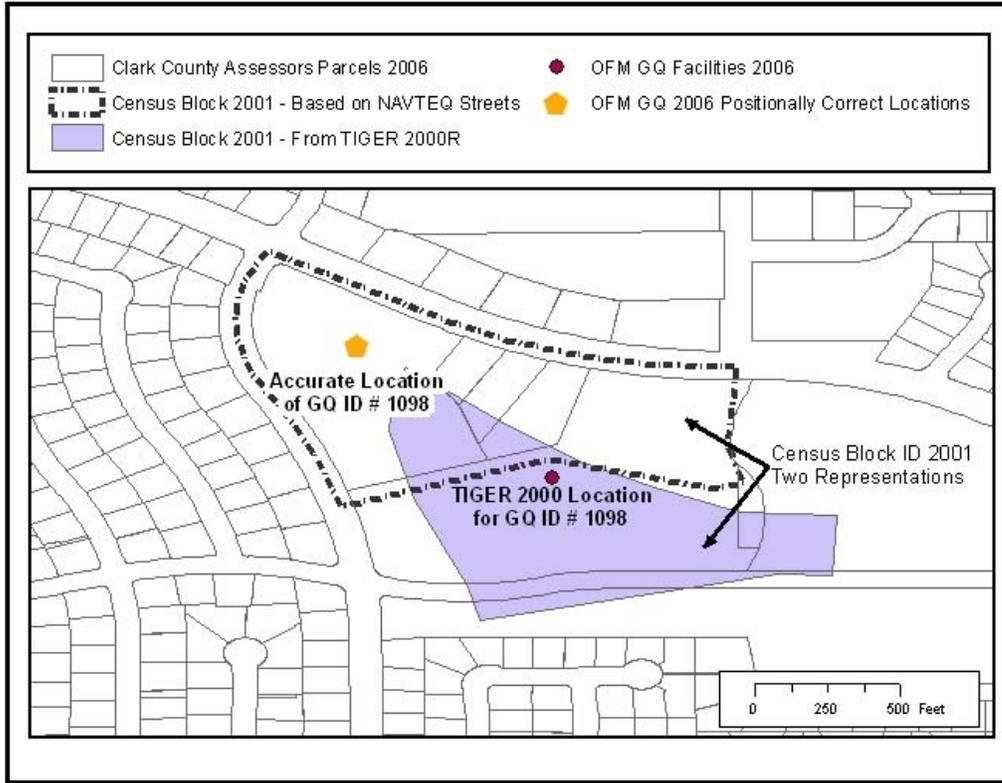


Figure 2. Positional Variation Between Local Data Sources and 2000 Census Boundaries

As stated earlier, the SAEP methodology requires that geocoded GQ facilities be consistent with 2000 census blocks. This means that SAEP uses the facilities geocoded to the TIGER line files rather than the more geographically accurate local data. The incidence of this is relatively low, representing only 1.4 percent of the total number of records in the GQ database. More telling of the overall geocoding accuracy is the 32 percent of facilities that have been repositioned by OFM (see Table 3). For GQ facilities with 50 or more people, the percent of repositioned GQ facilities is 46 percent, indicating that larger facilities may be more problematic.

Table 3. OFM Group Quarters Geocoding Information, 2006

<i>Geocode Type</i>	<i>Number of Records</i>	<i>Percent of Total</i>
Successfully Matched Geocodes	714	66.9%
Re-Positioned Geocodes	346	32.4%
Not Geocoded	7	0.7%
Total Records	1,067	100.0%

Estimate Methodology

As stated earlier, precisely geocoded facilities are necessary for the SAEP estimate process. OFM's city-level estimates only require GQ counts by city. The basic process for the city estimates is to calculate the change in city reported GQ counts from the base census year and add it to the Census 2000 GQ count as shown in the following equation:

$$\text{Current Year GQ Estimate} = \text{Federal Census 2000 GQ Count} + (\text{City Reported Count Current Year} - \text{City Reported Count 2000})$$

An example using data for the City of Battle Ground, Washington is provided in Table 4. If a city conducts a State Certified Special Census then the change would be calculated from the special census year rather than the federal census year.

Table 4. Group Quarter City Estimate Example, City of Battle Ground, WA

<i>Year</i>	<i>2000 Census Count</i>	<i>City Reported Total</i>	<i>City Reported Change</i>	<i>OFM GQ Estimate</i>
2000	116	108	--	--
2001	116	101	-7	109
2002	116	99	-9	107
2003	116	94	-14	102
2004	116	100	-8	108
2005	116	111	3	119
2006	116	112	4	120

The above formula produces reliable estimates on the city level as long as the city is consistent in its reporting and that the city and the Census Bureau are classifying essentially the same buildings as GQ. Using this methodology, if a city is not reporting any changes to a particular facility, the last known good count, the census count, is used for that facility. Legitimate new facilities would have their populations added to the city total.

When applying this methodology it is possible for a city's GQ population estimate to sum to a negative number. This happens when the city's reported base number (usually for the year 2000) is higher than the federal 2000 census count and one or more facilities close. When the GQ population drops below zero, OFM assumes the difference in population is actually being counted in the household population and therefore takes the difference from that population. In practice, this situation typically only happens in small towns with few GQ facilities, affecting only four of Washington's 281 cities over the 2001 through 2006 estimate cycles.

The SAEP uses the same basic methodology as the city/county estimates program. To begin the SAEP GQ estimate process, population change from the base census year is calculated at the facility level. As with the city/county process, this change gets tracked according to the location of the GQ's, but for SAEP the change gets recorded at the level of the individual census block. The 2006 SAEP GQ estimate for the City of Battle Ground is provided as an example in Table 5. All blocks with a GQ population are

included, irrespective of whether it was the Census Bureau, the City, or both that provided the information.

Table 5. Group Quarter Block Estimate Example, City of Battle Ground, WA

<i>County FIPS Code</i>	<i>Census Tract Code</i>	<i>Census Block Code</i>	<i>2000 Census GQ Count</i>	<i>OFM GQ ID#</i>	<i>City Reported Count 2000</i>	<i>City Reported Count 2006</i>	<i>City Reported Change 2006</i>	<i>Initial SAEP Block Estimate</i>
011	040405	2007	40	31	40	40	0	40
011	040406	2021	65	30	66	71	5	70
011	040407	1022	11	--	--	--	--	11
011	040407	1004	0	32	2	1	-1	-1
City Total			116		108	112	4	120

Similar to the city/county estimate process, city reported population loss for some institutions may be greater than the total 2000 census block GQ population count. Because of this difference it is mathematically possible for a block's GQ population to be less than zero using this estimate method. An example of this situation (Census Block 1004) is shown in Table 5. When negative block population estimates are generated, the estimate program overwrites any negative values with zeros and places the negative populations into a pool of temporarily unallocated GQ population by jurisdiction. This pool includes records that could not be geocoded as well as records where no detailed geographic information is available due to facility confidentiality. After the initial block level population adjustments are made, any population change that cannot be allocated to the appropriately identified census blocks are allocated on a proportional basis to all of the census blocks that contributed GQ population to a jurisdiction's total GQ population (see Table 6). 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 estimate when this occurs.

Table 6. Group Quarter Block Estimates Raking Example, City of Battle Ground, WA

<i>Census Block Code</i>	<i>Initial SAEP Block Estimate</i>	<i>Non-Negative SAEP Estimate</i>	<i>Unallocated Population Pool</i>	<i>Block Share of Total City GQ Pop</i>	<i>Block Share of Unallocated Pop</i>	<i>Final SAEP Estimate</i>
2007	40	40	0	0.331	-0.331	39.669
2021	70	70	0	0.579	-0.579	69.422*
1022	11	11	0	0.091	-0.091	10.909
1004	-1	0	-1	--	--	0.000
City Total	120	121	-1	1.001	-1.001	120.000

* In the SAEP any remainders of +/- 0.001 are added to the block with the largest share of population in the jurisdiction.

Unfortunately, it is not always possible to obtain precise geographic information for all institutions. Some institutions, typically colleges and universities, cover multiple census

blocks. OFM collects institution-level data for colleges and universities by facility type. The agency requests information on the number of students living in dorms, student apartments, social houses, and other university owned housing but not by individual building. Population changes for these types of institutions are distributed to census blocks based on the proportion of the GQ population for the institution as counted in the 2000 census.

Group Quarter Database Elements

While a considerable amount of information is needed to estimate GQ populations at the city level, even more information is required to produce block level estimates for the SAEP. A complete list of the individual data elements OFM currently stores in its GQ database for use in both the city/county and SAEP estimate programs is included in Appendix 4. Although, many of the individual data elements have been discussed previously, the purpose of this section is to discuss in greater detail some of the functions those variables serve.

The facility population is probably the most important piece of information in the database. Without it, most of the other data elements are useless for estimation purposes. OFM actually stores two annual population counts for each facility. The database contains a 'reported population' field and a 'accepted population' field. The reported population field, as the name implies, holds the city or institution reported population figure. The accepted population field holds the population which is actually used in the city/county or SAEP estimates. As discussed earlier, OFM analysts will adjust the population of a facility if there is sufficient reason for doing so. For example, if a city reports the total bed count instead of head count, the analyst will adjust the population accordingly. Additionally, adjustments will be made if a facility is better represented in the housing population rather than the GQ population. In these cases, the reported population is left as is but the accepted population value is changed. By storing the data in two fields, the analyst can query the database and identify facilities requiring further scrutiny. The two GQ population fields, coupled with a comment field, serve as an administrative record of any changes. Any changes made to the GQ population are also noted on the survey forms. In 2006, there were 64 cases (six percent of the total) where OFM analysts adjusted a GQ facility population.

The facility ID number, the facility name, street address, and contact information are essential for identifying and tracking GQs as well as locating them in geographic space. The GIS database retains all of the information generated during the geocoding process including latitude and longitude, the standardized street address used by the geocoder, post office name, ZIP Code, match scores, and the geocode date. Analyst comments about a particular geocode are stored in an additional field. All of the various data elements are included in the GIS metadata. When updating the GIS data each year, the analyst only needs to be concerned with the geocoding of new facilities. As shown in Table 3, approximately 32 percent of OFM's geocoded facilities have been repositioned

for one reason or another. By correcting issues as they become apparent and then carrying the changes forward, the analyst has more time to focus on the new facilities and other issues that may arise.

The census MAF/TIGER Accuracy Improvement Project is in the process of positionally realigning census geographies, making TIGER boundaries coincident with local data files where available. As this process continues, the GQs which were painstakingly repositioned to display correctly with TIGER 2000, may thus be out of alignment with the improved census boundaries. OFM's geocoded facilities will need to be adjusted again to match the updated boundaries. Beginning in 2007, OFM plans to store additional latitude and longitude fields for use with realigned census boundaries. The existing coordinates will be maintained because the GQ data will have to work with 2000 TIGER data as well as the realigned features.

Final Remarks

Collecting, maintaining, and using GQ facility data will continue to be a challenge. Because GQ populations can be large in size and typically do not share the same demographic characteristics as the household population, they will remain a necessary component of local estimate systems.

Although OFM has done its best to maintain a consistent set of GQ information, the nature of services has changed over time and the lines between which populations belong in households and which belong in GQs has been blurred. As GQ definitions change, the ability to compare data over time becomes more difficult. For example, changes to GQ definitions used in the American Community Survey (ACS) may make comparisons of decennial census GQ counts to ACS GQ estimates virtually meaningless (Lowe 2007).

The accurate geocoding of GQs is vital to the success of the SAEP. The quality control measures incorporated into the data collection process are important but are often not enough to ensure accurate geocodes. Many facilities use an administrative address rather than the actual site address, which can be far from the actual location of the facility. Reliance on geocoding programs to batch process addresses is frequently inadequate. Each geocoded GQ facility must still be manually checked for accuracy.

Changes to census boundaries as a result of the Census MAF/TIGER Accuracy Improvement Project will force OFM to once again review each GQ location and reposition geocoded facilities. As this process unfolds over the next several years, the database will need to be structured to work with both sets of boundaries. Continual improvements to OFM's GQ data collection, management, and tracking procedures will lead to more reliable estimates and thus benefit data users in the long run.

References

- Flowerdew, R. & Green, M. (1994). Areal interpolation and types of data. (In: S. Fotheringham & P. Rogerson (Eds.), *Spatial Analysis and GIS* (pp. 121-146). London: Taylor & Francis.)
- Jones, R. (2002). *Nonsampling Issues Resulting from the Use of Census 2000 and 2001 Supplementary Surveys as a Sampling Frame for National Epidemiological Survey on Alcohol and Related Conditions (NESARC)*. (Paper presented at the 2002 Annual Meeting of the International Conference on Improving Surveys, Denmark)
- Lowe, T. J. (2007). *The American Community Survey: Clothes without an Emperor*. (Paper to be presented at the 2007 Applied Demography Conference, San Antonio)
- Mohrman, M., Reese-Cassal, K. & Lowe, T. (2006) *Census Block Population Estimates: Purpose and Practice*. (Presented at the 2006 Population Association of America Conference, Los Angeles)
- Packard, W. S. (1967). *Staff Document No. 1: Determination of Population County Justice Courts*. Olympia, WA: State Planning and Community Affairs Agency, Population and Research Division
- Packard, W. S. (1968). *Staff Document No. 4: Determination of April 1, Population Highway Department Urban Areas*. Olympia, WA: State Planning and Community Affairs Agency, Population and Research Division
- Starsinic, D. E. & Zitter, M. (1968). Accuracy of the housing unit method in preparing population estimates for cities. *Demography*, 5(1), 475-484
- U.S. Bureau of the Census (2003). Notes and Errata, 2000 Census of Population and Housing/Washington State. Prepared by the U.S. Census Bureau. Washington, DC: Government Printing Office
- U.S. Bureau of the Census (2001). Census 2000 Summary File 1 Technical Documentation. Prepared by the U.S. Census Bureau. Washington, DC: Government Printing Office
- U.S. Bureau of the Census (2001). Census 2000 Summary File 1/Washington State. Prepared by the U.S. Census Bureau. Washington, DC: Government Printing Office
- U.S. Census Bureau (2006). *Design and Methodology, American Community Survey, Technical Paper 67*. Retrieved 11-1-2006, from U.S. Census Bureau's Web site: <http://www.census.gov/acs/www/Downloads/tp67.pdf>
- Washington State Office of Financial Management (2006). *2006 Population Trends*. Olympia, WA: Washington State Printing Office

Appendix 1. Publicly Available SAEP Estimates

Small Area Estimate Program (SAEP) | OFM - Mozilla Firefox

File Edit View History Bookmarks Tools Help

http://www.ofm.wa.gov/pop/smallarea/def Google

ESTIMATES FOR CENSUS AREAS, 2000-2006 BY SINGLE YEAR

Metadata Link	Estimate	GIS - Shapefile	File Size
American Indian Reservation (includes Tribal Designated Statistical Areas)	Excel	Download	281 KB
Block Groups and Maps	Excel	Download	595 MB
Census-Designated Places (CDP)	Excel	Download	643 KB
Legislative Districts (2002) and Maps	Excel	Download	917 KB
Legislative Districts (2002) and County Parts	Excel		
Public Use Microdata Area (PUMA) - 1 percent	Excel	Download	522 KB
Public Use Microdata Area (PUMA) - 5 percent	Excel	Download	812 KB
School Districts	Excel	Download	1,658 KB
Census Tracts and Maps	Excel	Download	3,476 KB
Urban Areas	Excel	Download	832 KB
Urban Areas and County Parts	Excel		
ZIP Code Tabulation Areas (ZCTA)	Excel		
ZIP Code Tabulation Areas (ZCTA) and County Parts	Excel		

ESTIMATES FOR "OTHER" AREAS, 2000-2006 BY SINGLE YEAR

Metadata Link	Estimate
Selected Islands	Excel
Water Resource Inventory Areas (WRIA) (Washington State Department of Ecology)	Excel Revised 10/11/2006

Last modified: December 05, 2006
E-mail: OFM.Forecasting@ofm.wa.gov

Done 0.141s

Appendix 2. 1999 City/County Annual Survey Form: GQ Population

Office of Financial Management

1999 FORM A

D. Group quarters population...persons living in places that reflect other than ordinary household life. Generally places where unrelated people eat and/or sleep together in a "group" living situation.

Note: Do not include any group quarters population in areas annexed from April 2, 1998 through April 1, 1999

Type of group quarters facility*	Population April 1, 1999
Nursing/convalescent home	546
College dormitory, fraternity, sorority, etc.	
Mental/correctional facilities	
Jail (persons with stays of <u>six</u> months or more or who have no usual residence)	
Military barracks, bachelor officers' quarters	
Other, specify Adult Foster Homes	48
Total	594

*The following are NOT group quarters: apartments for the elderly, student apartments, and short-term care facilities such as detention centers or drug/alcohol rehabilitation centers. If you have questions, call and ask for assistance.

Detailed information:

Name and type of facility	Address	Contact/phone number	Population
Delaware Plaza	926 Delaware	Admin. 423-3333	92
Cottonwood Lodge	1524 - 3rd Avenue	Admin. 636-5090	16
Americana Nursing	917 - 7th Avenue	Admin. 425-5910	63
Frontier Nursing	1500 - 3rd Avenue	Admin. 423-8800	118
Park Royal Conv.	910 - 16th Avenue	Admin. 423-2890	38
Manor Nursing Home	1330 - 11th Avenue	Admin. 425-6706	36
Total:			

Use additional sheets if necessary.

Canterbury Gardens / 1457 - 3rd Avenue / Admin. 423-2200 ----- 51
Please describe the procedure used to get the information in the above tables.

Campus Towers / 1767 20th Avenue / Admin. 423-6200----- 102

Fremont Village / 1416 - 3rd Avenue / Admin. 577-5913----- 30

West Castleman Estates / 2304 W. Castleman / Admin. 636-3886----- 6

Birchwood #1 / 2437 - 50th Avenue / 425-5919----- 12

Birchwood #2 / 4326 Pine Street / 425-6436----- 12

Heron Crest #1 / 2446 - 50th Avenue / 575-8741-----p- 6

Heron Crest #2 / 2441 - Hickory Street / 575-9259----- 6

Appendix 3. 2006 City/County Annual Survey Form: GQ Population

D. Group quarters population...persons living in places that reflect other than ordinary household life. Generally places where unrelated people eat and/or sleep together in a “group” living situation.

- Do *not* include any group quarters population in areas annexed from *April 2, 2005 through April 1, 2006*. These annexed populations are handled separately.
- Group quarters counts from Census 2000 (or special census) are shown and used for those cities not tracking annual changes in specific group quarters facilities.
- Report current counts for the facilities your city/town has been reporting. *These facilities are already listed in section D.3 beginning on the next page.*
- Add new group quarters facilities (if any) below in D.2.

D.1. Total Group Quarters Population by Type

Type of group quarters facility	Population April 1, 2006	FOR REFERENCE	FOR REFERENCE
		Population Reported by City April 1, 2005	Population Counted in April 1, 2000 Census
D.1. = D.2. + D.3.			
Nursing/convalescent home	596	609	401
College dormitory, fraternity, sorority, etc.		0	0
Mental/correctional/jail (<i>persons with stays of six months or more or who have no usual residence</i>)/ other institutions		0	148
Military barracks, bachelor officers' quarters		0	0
Other non-institutional: (<i>please specify</i>)		0	312
Total Group Quarters Population: Insert D + Newly Added Facilities	596	609	861

Note: The following are NOT group quarters: apartments for the elderly, university controlled family housing (including apartments), and short-term care facilities such as detention centers or drug/alcohol rehabilitation centers. If you have questions, call and ask for assistance.

D.2. List New Group Quarters Facilities

Name of new facility	Type of Facility	Address incl. zip code	Source of new facility*	Resident Population Count
Use additional sheets if necessary		Total Population in New Group Quarters Facilities		

*Conversion of residential/commercial structure? Annexation? New construction?

D.3. Group Quarters Detail Worksheet for Previously Reported Facilities 2006.**Municipality of: Longview in Cowlitz County**

Please make changes/corrections to facility information if needed. Post new counts.

Group Quarters Facility Type: Nursing/Convalescent Home						
OFM ID	Facility Name	Address	Zip Code	Phone Number	2005 Population	2006 Population
530	Another Option	2274 34TH AVE LONGVIEW	98632	360-575-8196 1796	6	6
1212	Another Option #3	2827 OCEAN BEACH HWY LONGVIEW	98632	360-578-1209	4	4
514	Canterbury Gardens	1457 3RD AVE LONGVIEW	98632	360-423-2200	60	62
519	Canterbury Inn	1324 3RD AVE LONGVIEW	98632	360-425-7947	155	152
506	Cedar Gardens Adult Family Home	28 W PINE LN LONGVIEW	98632	360-578-0279	5	6
1259	Colorado House	368 COLORADO ST LONGVIEW	98632		2	2
508	Delaware Plaza	926 DELAWARE ST LONGVIEW	98632	360-423-3333	88	100
522	Elder Care Connections	2437 50TH AVE LONGVIEW	98632	360-575-1684	6	6
511	Evergreen Americana Health & Rehab.	917 7TH AVE LONGVIEW	98632	360-425-5910	60	50
509	Evergreen at Park Royal II	910 16TH AVE LONGVIEW	98632	360-423-2890	41	34
507	Evergreen Frontier Rehabilitation and Extended Care	1500 3RD AVE LONGVIEW	98632	360-423-8800	104	94
518	Evergreen Manor Nursing Home	1330 11TH AVE LONGVIEW	98632	360-425-6706	30	29
517	Heroncrest 1	2446 50TH AVE LONGVIEW	98632	360-575-9412 8741	6	6

Appendix 4. Variables in OFM's GQ Database

<i>Variable Name</i>	<i>Variable Description</i>
Facility_ID	Unique numbers that are automatically generated to identify each facility in the database.
Facility_Name	Name of the group home, facility or institution.
Confidential_Flag	Yes/No field identifying the record as being confidential.
Group_Type	Group quarter facility type code.
Source_Code	Data source type code
Source_Type	Data source type description
County_Name	Name of county where the facility is located.
Jurisdiction	Current name of the municipality where the facility is located. Note: due to city annexations and incorporations some facilities change jurisdiction over time.
Address	Best available facility address.
Post_Office_Name	Post Office name associated with facility address.
State	State name associated with facility address.
Zip_Code	Zip Code associated with facility address.
Phone_Number	Facility phone number.
Year_Facility_Opened	The year OFM began tracking the group quarter facility as an individual record in this database. Note: database created in 1999.
Year_Facility_Closed	The year the facility closed, if known. Used as an indicator for the the facility closing as opposed to simply not being reported.
R2000	Source reported population 4/1/2000.
A2000	Accepted population used in OFM's official estimate 4/1/2000.
...	Annual repeats of reported and accepted populations.
R2006	Source reported population 4/1/2006.
A2006	Accepted population used in OFM's official estimate 4/1/2006.
FIPS	Three digit county FIPS Code.
Place00	2000 Census Place Codes for incorporated cities and unincorporated counties.
Block00	2000 Census Block ID Number.
Latitude	Latitude that best represents the facility location in NAD83 decimal degrees.
Longitude	Longitude that best represents the facility location in NAD83 decimal degrees.
Address2	Geocoder standardized address.
City2	Geocoder standardized Post Office Name.
State2	Geocoder standardized State.
ZipCode2	Geocoder standardized Zip Code.
Zip_4	Geocoder assigned four digit Zip Code add on number.
GeoResult	The geocoder assigned accuracy code, which describes the quality of the address match. Any code equal to 'MANUAL' means OFM staff has manually placed the facility at the current location. Only street level geocodes are accepted from geocoding software packages.
DtGeoUpdate	Date facility was geocoded or date when the geocode was updated.

TxtGeoComment	Comments about the placement or accuracy of the records which have been manually geocoded.
PID	County Parcel ID number.
Alt_Latitude	Alternate Latitude. This code is assigned when the OFM geocoded census block is different from the block assigned by the 2000 census. This field is only filled when we are very sure there is a problem with the latitude/longitude and or the census 2000 block assignment. This difference is often apparent when we use local data for geocoding. Sometimes is the difference caused by the difference in geo-referencing or accuracy of the locally received data vs. the TIGER/Line files, other times it appears that the census geocode is incorrect. We have chosen to consistent with the census bureau so the official locations and codes for the facility are stored in the 'LATITUDE', 'LONGITUDE', and 'BLOCK00' codes. These codes are maintained to help resolve local data discrepancies and hopefully issues in the 2010 census. These changes do not appear in the Census Count Revisions assigned by the census CQR program.
Alt_Longitude	Alternate Longitude. This code is assigned when the OFM geocoded census block is different from the block assigned by the 2000 census. This field is only filled when certain there is a problem with the latitude/longitude and or the census 2000 block assignment. This difference is often apparent when using local data for geocoding. Sometimes the difference is caused by differences in the positional accuracy of the locally derived data vs. TIGER/Line files. Other times it appears that the census geocode is incorrect. OFM has chosen to be consistent with the census bureau so the official locations and codes for the facility are stored in the 'LATITUDE', 'LONGITUDE', and 'BLOCK00' codes. These codes are maintained to help resolve local data discrepancies and potential issues in the 2010 census. These changes do not appear in the Census Count Revisions assigned by the Census CQR program.
Alt_Block00	Alternate 2000 Census block code associated with Alt_Latitude and Alt_Longitude coordinates.

Acknowledgements

I would like to thank my colleagues in the OFM Forecasting Division for their help and support in developing the SAEP database and for their assistance with this paper. I would also like to extend my appreciation to Washington's 281 cities and 39 counties for providing OFM with the data necessary to support our estimate programs.