

POPULATION ESTIMATES & PROJECTIONS

Research Brief No. 21
December 2003

Using a Geographic Information System in a Population Estimate Program: The Pasco Case

Contributors: Theresa Lowe and, Mike Mohrman

The Office of Financial Management's (OFM) April 1 population estimates program develops estimates for local jurisdictions that are used for revenue allocations and program administration (RCW 43.62.020). State collected locally shared revenues distributed to city and towns amount to about \$31 per person per year. This report discusses how a geographic information system (GIS) was used to determine if a low-cost sample survey was a financially feasible option to improve the 2003 population estimate for the City of Pasco.

OFM uses the Housing Unit Method to estimate city populations. A simplified version is shown below. Annual population estimates are benchmarked to the most recent federal census. Locally taken censuses are also used when based on federal census definitions and approved by OFM. Administrative data or special survey data may be used to *adjust* the base census occupancy rates and household size in years following a census.

$$\begin{array}{rcl} \text{Current City Housing X Occupancy Rate X Avg. Persons Per Occupied Unit} = & \text{Persons in Housing} & \\ & + & \\ \text{Current count of persons in nursing homes, correctional, other facilities} = & \text{Persons in Facilities} & \\ & \hline & \text{Total City Population} & \end{array}$$

Background

In the spring of 2002 the Tri-Cities area was a bright spot of economic expansion due to federal nuclear waste containment projects. Unemployment was low and residential construction was booming. Because of the strong housing demand, city officials felt the occupancy rates used by OFM when developing their population estimate were too low—particularly for recently built housing.

When cities officials feel that OFM's population estimates are not accurate they have two possible actions for the following year's population determination. One is to conduct an actual population count of the entire city. The other is to conduct a sample survey to update the occupancy rate and/or household size used in the Housing Unit formula. Censuses and surveys could cost \$15 or more per house depending on management, staffing, or other special considerations. The cost for a city the size of Pasco, with about 35,000 people, would be higher than for most cities of a similar size. Pasco is in a rural-agricultural area and is 56 percent Hispanic. Pasco's Hispanic population is difficult to enumerate because of transient agricultural workers, undocumented persons, and language barriers. A citywide census would involve enumerating the population, by name, in nearly 12,300 housing units and might cost more than \$200,000. The sample survey option to determine the current occupancy rates would involve contacting only 2,500 of the city's housing units. If sample units could be selected from an address listing, the survey cost would be minimal.

Developing a Master Address List for Sample Selection

OFM and the City of Pasco worked together to evaluate the lower cost sampling option. *This option required developing complete and accurate address lists for drawing independent samples for single-family units, multi-family units, and mobile homes/trailers. Most address lists, however, are administrative records that have been developed for specific purposes that are not related to survey taking. Thus, when these address lists are used for sample selection, the coverage of housing may be incomplete and therefore create a bias in the sample.*

Vacant housing tends to be excluded from address lists. For example, utility records may exclude vacant new constructions. Many vacation homes or other seasonal housing units are in rural area without street addressing. Their owners may use post office boxes or general delivery to receive mail. These seasonal homes would be missed in any sample selected from street address listings. Rooms for rent above commercial structures—many

which cater to somewhat transient populations, can have many vacant units. These too are typically missed because the rental units may not be individually addressed.

Housing records from six different sources were combined to develop the Pasco address list: city utility records, the city's multi-family housing list, city building permit records, the address list obtained from the 100 percent count of all housing annexed since 2000, assessor records, and housing authority records. Development began by standardizing the address in the input data sets. This was necessary to remove duplicate records and to prepare the data for geocoding and mapping. Duplicate addresses were expected because the different data sources overlap in coverage. For example the assessor's records included many units annexed by the city, but not units from the most recent annexations. Some datasets did not show much overlap. For example, the assessors classify 5+ unit structures as commercial so most of these units did not appear in the residential extract from the assessor or in the cities utility listing. In all, the six data sets yielded a total of 9,632 unduplicated addresses. This was 54 percent of the original 17,957 addresses submitted. The other 8,325 were duplicates.

Geocoding and Mapping the Master List for Fieldwork

The list of unduplicated addresses was geocoded to the county GIS parcel file. Records with parcel ID numbers were matched geocoded to their corresponding parcels. Unmatched records were then matched to the parcel file based on address. This was necessary because the parcel database had records which were more current than the local street centerline file, which is typically used. Finally, records, which were not matched by other methods, were matched to the available street centerline files. In all 9,596 or 99.6 percent of the addresses were matched.

Table 1. City of Pasco Address List Geocoding Information

Match Type	Number of Addresses	Percent Matched
Parcel ID	7,179	74.53
Assessor Address	1,784	18.52
Geocoder Address	633	6.57
No Match	36	0.37
Total Matched	9,596	99.63
Total Addresses	9,632	

The geocoded addresses were then used to make an atlas for staff to use during the field check of the list's coverage. A sample of census blocks was drawn and displayed on the map with the county parcel boundaries and the geocoded units. To aid the field crew, each housing unit was labeled with the house number from the address database.

Field Evaluation and Findings

The field check to evaluate the housing coverage of the prepared address list was conducted in September 2002. The blocks selected for examination contained a good representation of single-family housing, apartments, and manufactured homes. New residential construction areas were also identified for canvassing. The evaluation survey covered one-tenth of the blocks in Pasco and these blocks contained about 17 percent of the city housing. The actual housing on each block was compared to the housing expected on the basis of the prepared address list. The 2000 census counts of housing by block were also used to help resolve discrepancies identified in the field. The findings are summarized in Table 2.

City Survey Areas by Expected Unit Type and House Number

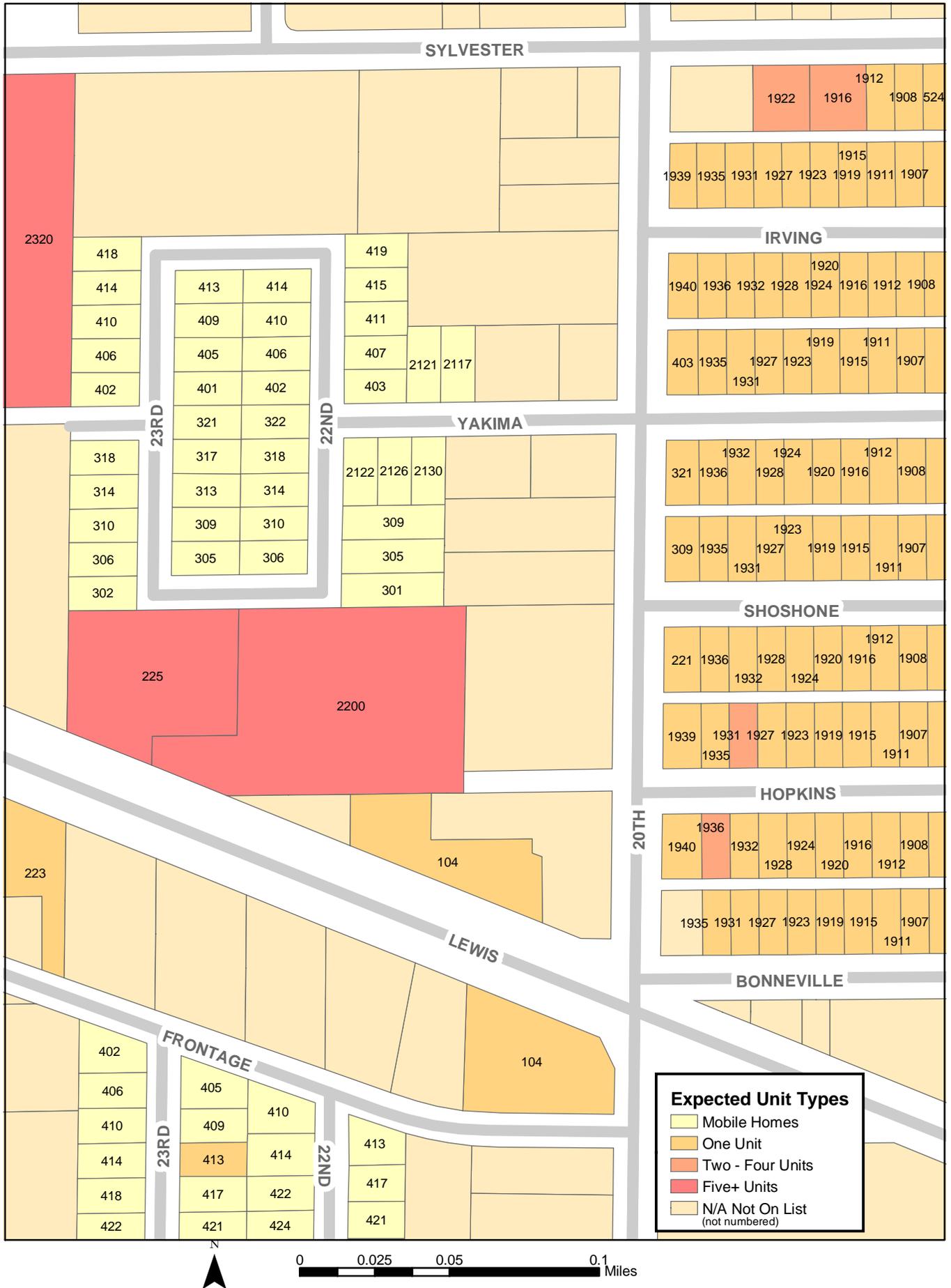


Table 2. Results of the September 2002 Field Survey in the City of Pasco

Goals	Comments	Findings
Evaluate the completeness and accuracy of the housing address list developed for the City of Pasco to draw housing samples for <ol style="list-style-type: none"> 1. Single-family, 2. Multi-unit, and 3. Manufactured housing. 	An address list was developed by The Office of Financial Management (OFM) from the following files provided by Pasco: <ol style="list-style-type: none"> 1. Assessor records 2. City utility records 3. City's multi-family housing list 4. Building permits since January 1, 2000 5. Housing authority records 6. Annexations since January 1, 2000 	<ol style="list-style-type: none"> 1. Overall, the housing address list appears to be good for single-family units. Uncertainties were limited to about 1 percent of the single-family units. 2. Manufactured/mobile homes and apartments are problematic. <ul style="list-style-type: none"> • A single address may represent many mobile homes or apartments. • Some entire blocks with mobile homes and multi-units counted in the census are excluded from the survey address list and would be missed in the sample. * This appears to be true for inconspicuous low quality units, which may have more vacant units. Lack of inclusion would create a bias.
Notes: * Some mobile/manufactured housing and multi-unit homes are missing from the survey address list developed for Pasco because they are in areas designated as "commercial" in the assessor's records where residential units are not counted. The supplemental list of multi-units provided by Pasco was incomplete for units in these "commercial" areas.		

Based on the field check, the primary address list would need correction before representative samples could be drawn for apartment unit and manufactured housing parts of the housing stock. Correcting the address list would require a systematic canvass of the entire city. Thus, the best option would be to systematically canvass the entire city—drawing the sample and conducting the occupancy survey at the same time. While this would still be considerably less costly than a full-scale census, the fieldwork requirements might raise the cost of the sample survey above what might be gained in revenue allocations. At this point, other alternatives for developing the 2003 population determination were discussed with Pasco. These included doing a limited survey of the vacant units in new¹ residential housing constructed over the last year.

Conclusions

Housing units or households are the typical sampling unit used in most surveys to obtain population and housing characteristics—or opinions, purchasing behavior, and other information of interest. GIS technology assisted in the development of a master housing address list for the City of Pasco by structure type, and the mapping of these units by address to facilitate field checking. The address list for Pasco, which was developed from six reputable sources, failed to provide adequate coverage for multi-family units and manufactured housing.

This case underscores the difficulty of developing address lists that are truly representative of the housing or population in any community and also highlights how GIS can facilitate evaluating potential address lists for sampling. Using GIS technology to evaluating the survey option for the City of Pasco readily identified sampling problems that led to other solutions for improving the city's population estimate.

¹ These were houses that were identified in the building permit records as being completed and ready for occupancy between April 2, 2002 and April 1, 2003.