

# IMPROVING THE COVERAGE OF ADDRESSES IN THE 1990 CENSUS - PRELIMINARY RESULTS

Deborah H. Griffin, Joan M. Pausche, Emilda B. Rivers,  
Amy L. Tillman and James B. Treat

## 1. INTRODUCTION

Beginning with the 1970 Decennial Census, census-taking shifted from a tradition of canvassing, listing and enumerating to a mail methodology for census-taking. Prior to 1970, census workers personally visited each living quarters to list the address and to enumerate the household members. About 60 percent of the population in 1970 were mailed a census questionnaire and asked to complete it and return it by mail. Households for which a questionnaire was not returned by mail required a personal visit. This procedure was successful in 1970 and was expanded to include about 95 percent of the population in the 1980 Decennial Census.

Due to its effectiveness in 1970 and 1980, the 1990 Decennial Census relied on the mail census methodology to count most of the population. The mail census affects all states and the District of Columbia.

A key requirement for a successful mail census is the development of a complete list of addresses. The Census Bureau invests considerable time and resources into developing an accurate address list. But a list of addresses is of little value until each address is assigned to census geography. The combination of compiling addresses and assigning these addresses specific geographic codes results in the basic address files used to control data collection and processing. Different parts of the country require different address list development and "geocoding" strategies. This paper provides an overview of these methodologies and preliminary assessments of the impact of these programs on census address list coverage.

A successful mail census requires much more than a complete and accurate mailing list. It also requires accurate delivery of a census questionnaire to each address and timely mail response from the public. Low levels of mail response increase the time and staff resources required to correctly complete the data collection process. This paper provides preliminary assessments of the success of delivery and mail response in 1990.

## 2. SCOPE

This paper is limited to early results of the mail census and focuses on address list development activities. This paper does not cover the enumeration of persons and households in extremely rural areas or the enumeration of the homeless, persons on military bases or ships, persons in hospitals, halfway houses, and in other special living situations. Future papers will discuss coverage of persons and detailed analyses of the coverage of addresses in the census, looking at the costs as well as the errors in various components of the census.

## 3. BACKGROUND

The methods for census-taking by mail include three basic variations that are used in different types of areas across the country. These areas may be loosely labelled as urban, suburban and rural. In this paper rural areas refers only to the most rural mail census areas. The term, "rural" can also be used to describe the most rural, non-mail areas. The address list development and delivery methods used in urban, suburban and rural areas are tailored to the population density and the addressing systems used by the United States Postal Service (USPS). The urban methodology is used to identify and geocode the majority of the mail census addresses.

The urban areas are the largest metropolitan areas across the nation. While they cover only 2 percent of the land area, about 60 percent of the population live in these areas. In 1990, all states and the District of Columbia include at least some addresses that are compiled using urban area techniques. The highest concentration of urban areas were found in New York, New Jersey, Rhode Island and the District of Columbia. Addresses in these urban areas typically include a house number and a street name (for example, 801 Main Street). This type of address is termed, "city delivery." These areas include large and small apartment buildings.

The suburban mail census areas generally surround the major metropolitan areas and may contain smaller cities and towns. They are mostly concentrated in the Northeast, but others are scattered across the entire country. All states (excluding the District of Columbia) include areas where suburban methods were used to compile and geocode addresses. States with high concentrations of suburban areas include South Carolina, Maine and Indiana. Suburban areas have a lower population density than urban areas. A combination of city delivery addresses and rural addresses (for example, RR 4 box 28) are found in suburban areas. Post Office boxes are often used. Many addresses are single family homes.

Like suburban areas, the rural mail census areas have lower population density than the urban areas. Rural mail areas are found mostly in the eastern half of the nation, especially in Appalachia and the south. Addresses are characterized by rural route and box numbers. Some addresses are limited to descriptions such as "white house on the corner". As in suburban areas, many addresses are single family homes.

## 4. LIST COMPILATION AND GEOCODING

Compiling and geocoding of addresses can be accomplished either concurrently or sequentially. The methods used in identifying and geocoding addresses were designed to utilize existing automated geocoding capabilities and available address updating resources. Address compilation methods involve commercial vendors, the USPS and

census field visits. Geocoding refers to the assignment of a set of geographic codes to each address. Geocoding allows control during the collection process and aggregation of census data to various geographic levels (for example, state, county, block). During census field activities, staff use maps and address lists to identify and simultaneously assign a geographic code to an address. Addresses obtained from other sources (usually the USPS or a commercial vendor) require a separate geocoding activity.

The sections below detail the specific stages and preliminary results of address list development and geocoding activities in urban, suburban and rural mail census areas.

**4.1 Urban Areas** In urban areas relatively inexpensive commercial address lists can be purchased from various vendors. The USPS provides valuable assistance in updating the address list. Computer files exist that contain information that provides automated geocoding capabilities. The USPS delivers Census questionnaires.

**4.1.1 Initial list** In densely populated urban areas as well as areas surrounding these central cities, the Bureau of the Census purchased the initial inventory of addresses from a commercial vendor. The commercial address list for each urban area was selected through a competitive procurement process. To become a part of the census list each address had to be assigned to specific census geography. This assignment or "geocoding" included computer, clerical and field coding, as necessary. Addresses from the commercial vendor were used in areas of the United States where all of the following conditions were present:

- o a commercial address list existed
- o city type mail delivery (that is, delivery to house number and street name addresses) was provided by the USPS and
- o the Bureau of the Census had the ability to assign geographic codes by computer.

Only a portion of the vendor addresses could be used as the initial list in urban areas. A total of 69.3 million addresses were purchased in 1988 by the Bureau of the Census. These addresses represented the entire commercial vendor's inventory. The Census Bureau identified ZIP Codes that contained areas which receive city type mail delivery from the USPS. About 55.2 million addresses were identified in these ZIP Codes. These addresses covered 49 states (all but Hawaii) 1/ and the District of Columbia. The remaining addresses were in ZIP Codes without USPS city type mail delivery and therefore they were dropped from the census address file. Within the selected ZIP Codes, about 94.6 percent of the addresses (52.2 million) were identified by the Census Bureau as being complete city type addresses. The remaining addresses were dropped from the census address file because they were non-city delivery, duplicate or incomplete addresses.

The census address file after geocoding included approximately 51.6 million addresses. The Census Bureau was unable to geocode 0.6 million city type addresses. These rates only reflect computer and clerical geocoding and not the additional field geocoding procedures which were implemented prior to questionnaire delivery. Computer and clerical geocoding resulted in a

national geocoding rate of 98.9 percent. The state level geocoding rates ranged from 97.0 to 99.8 percent with a median value of 99.0 percent.

**4.1.2 First Update** Concurrent with geocoding of vendor addresses, the USPS conducted a postal check on these addresses. All 55.2 million vendor addresses were given to the USPS. The main objective of this postal check was to verify the accuracy and completeness of the purchased addresses. An address card was created for each mail list address. During the postal check, letter carriers and postal clerks sort or "case" address cards into slots for each address on their route, make corrections and identify duplicate and undeliverable addresses. If an address card was not provided for an address, the letter carrier identified the address as missing and provided an "add card" for each missing address.

As with vendor addresses, missing addresses were coded to census geography before being added to the census inventory. The Census Bureau attempted to computer geocode addresses identified by the USPS as missing. Missing addresses that were not computer geocoded were sent to field geocoding.

In 1989 the USPS conducted this initial postal check. The USPS identified 3.3 million potentially missing addresses. After editing and computer geocoding, 1.6 million addresses were added to the urban area address list. Approximately 0.4 million of these missing addresses were dropped from the census files because they fell outside of urban areas or represented duplicate addresses. About 1.3 million of these missing addresses could not be geocoded by computer. Along with the ungeocoded vendor addresses, efforts were made to geocode these addresses in a later office and field geocoding operation prior to questionnaire delivery.

These 1.6 million added addresses represent a 3.1 percent increase over the geocoded vendor address universe (51.6 million). These results varied by state ranging from a low of 1.5 percent in the District of Columbia to a high of 6.4 percent in Massachusetts. This was partially due to the state geocoding rates which ranged from 36.5 percent to 77.2 percent. High add rates may indicate areas where the initial vendor list was less complete, areas where the USPS and the Census Bureau have inconsistent perceptions of the mailing address, or areas where address conversions are taking place. In some parts of the country addresses are in the process of being renumbered and renamed. This is called, "address conversions." Low add rates can indicate an area with a very accurate vendor list, areas where housing growth is low or areas where the Census Bureau had trouble geocoding postal adds. Low add rates will also occur where the USPS did a poor job identifying missing addresses.

In addition, this check identified areas where the USPS detected that the census address list included high levels of undeliverable or duplicate addresses. Nationally, about 1.0 million city type vendor addresses (2.0 percent) were classified as either undeliverable or duplicate by the USPS. The variability in undeliverable rates can be especially noted at lower levels of geography, for example at the census district office level. Of the over 400 offices, most rates

were low with approximately 70 percent of the offices below a 2 percent rate. However, high undeliverable rates were noted in several offices, with over 3 percent having rates greater than 5 percent.

The offices with high undeliverable rates are the offices where we expected potential deliverability problems during the questionnaire delivery stage of the census. High undeliverable rates do not necessarily represent areas with poor address lists. The rates could be a result of the Census Bureau having a different version of an address than the address recognized by the USPS. High undeliverable rates also suggest areas with high commercial occupancy rates because the USPS is asked to classify these addresses as undeliverable. In addition, the USPS often classifies vacant units as undeliverable suggesting that areas with high undeliverable rates may represent areas with high vacancy rates.

**4.1.3 Second Update** After the geocoded missing addresses were added, lists of all addresses were printed in a sort by census geography. About 25,000 census workers travelled all streets in these areas to update the lists with missing addresses, make corrections to existing addresses, correct census geography, and identify duplicate, nonexistent and commercial addresses. Census workers used maps with census geography during canvassing. Missing addresses were geocoded during canvassing.

Almost 6 million addresses were added as a result of this review. This represents an 11.3 percent increase in the urban census address list nationally. These rates varied by state. The highest add rates occurred in Pennsylvania (26.6 percent), Mississippi (24.0 percent) and South Carolina (18.5 percent). High add rates may indicate areas where the vendor and USPS updates could not be geocoded. High rates will also result in areas undergoing address conversions. This was true in parts of Mississippi. Relatively low add rates were found in the District of Columbia (5.5 percent), Wisconsin (6.4 percent) and Louisiana (6.6 percent). Low growth areas, areas with relatively high geocoding success and areas with updated vendor files were expected to fall into this category.

During this update, corrections to census geography were made by adding addresses to the correct geography and deleting addresses from the incorrect geography. Geography was corrected on approximately 2.6 percent of the urban addresses. Most district offices had geocoding correction rates of less than 2 percent; however, over 8 percent of the offices had rates over 5 percent. The highest geographic correction rates were found in several offices in Pennsylvania. The extremely high rates in these areas of Pennsylvania led to a second geocoding operation to correct geocoding errors.

**4.1.4 Summary** The initial address list development activities in urban areas included compilation and geocoding of addresses from the following sources:

- o vendor
- o initial postal check
- o field check

These initial list development activities resulted in a count of approximately 59.2 million

addresses in urban areas. Although the vendor lists were the source of most urban addresses, significant additional gains in address list coverage resulted from update activities.

The relative contribution of these update programs can be especially noted at lower levels of geography such as the district office level. In over 9 percent of the district offices the initial postal check and the field check combined to increase the coverage of addresses by more than 25 percent.

**4.2 Suburban Areas** In suburban areas, the absence of computer files that would provide automated geocoding capabilities led the Census Bureau to compile its own list of addresses. Despite some differences in the USPS and Census Bureau addressing definitions, the USPS assistance is valuable in updating address lists in suburban areas and delivering the census questionnaires.

**4.2.1 Initial list** From June 1988 through January 1989, the Census Bureau used about 30,000 field listers to create an initial list of addresses in suburban areas. They travelled every passable road in their assigned area looking for houses and other places where people could live. They listed vacant and occupied living quarters. For each address they found, they attempted to obtain a complete mailing address and the householder's surname, if available, and placed a spot on the map showing the geographic location of the address. This combined address compilation and geocoding activity produced the initial address lists in suburban areas. This activity was known as prelist.

The field staff recorded address information for nearly 27.8 million addresses during this initial listing. Approximately 76.0 percent of the addresses listed had house number and street name. About 19.0 percent were rural type addresses, such as rural route/box number, post office box or general delivery addresses. The remaining 5.0 percent were incomplete addresses.

During the initial listing, some clusters of addresses were inaccessible to the field staff, due to such situations as washed out roads or locked gates. These clusters were often in seasonal resort areas. The listers were unable to determine the actual number of living quarters in these clusters and could not verify the correct addresses. They estimated the number of addresses and recorded this information so that the addresses could be listed and geocoded in subsequent operations.

**4.2.2 First and second updates** In suburban areas the initial listing was updated in a two stage procedure designed to ensure that the address list was complete and accurate. In 1989 the USPS conducted the first coverage check on the list of addresses compiled during the initial listing operation. Updates that were received from the USPS required field reconciliation by Census staff. Addresses for some parts of the country were reviewed by the USPS in February and reconciled in June. In other areas the USPS conducted the check in April with field reconciliation occurring in August.

The purpose of the postal check was to have the USPS review the addresses from the initial listing in preparation for the final mailing list. The

postal carriers were instructed to check the deliverability of each address, separate any duplicate addresses and identify any residential addresses on their routes that were missing.

The most important reason for conducting the postal check was the improvement of address coverage in the listed areas. In addition, the address information obtained during the initial listing did not always agree with the USPS information. For these cases, the USPS was asked to make corrections to improve the deliverability of the addresses. Incomplete addresses that were known to be undeliverable following the initial listing were not sent to the postal check. Similarly, addresses in ZIP Codes that covered both urban and suburban areas were not included in this postal check.

Approximately 20.1 million addresses, or 72.1 percent of the listed addresses, were reviewed by the USPS. The carriers considered 78.0 percent of these addresses deliverable as they were addressed, 5.0 percent deliverable with corrections, 14.3 percent undeliverable and 2.7 percent duplicates of other addresses. After completing the postal check and processing the results, an updated listing was prepared by the Census Bureau.

The second update to the initial list was a field check designed to locate each address identified as missing by the postal carriers to verify that it was not already accounted for in the census files. The Census Bureau made an attempt to obtain better mailing addresses for all incomplete addresses and for addresses returned as undeliverable by the USPS. In addition, this field check included each address classified as a duplicate by the USPS and attempted to obtain addresses for clusters of housing units that were inaccessible during the initial listing.

Of the initial listing, the field check corrected 11.3 percent of the addresses, identified 1.5 percent of the addresses for deletion and left the remaining 87.2 percent as they appeared in the updated address listing. Of the addresses classified as duplicates by the USPS, only 12.0 percent were considered true duplicates by the field staff and were deleted from the census files.

During the field check, the staff added nearly 1.2 million addresses that were not already on the suburban address list. These adds were, in some cases, addresses that had been identified as missing by the USPS. This represents an increase of 4.2 percent. These results varied by Regional Census Center. with the Los Angeles region achieving an add rate of nearly 10 percent. High add rates may suggest areas of growth or areas with a less accurate initial listing.

**4.2.3 Summary** The initial address list development activities in suburban areas included compilation and geocoding of addresses from the following sources:

- o prelist
- o initial postal check
- o field reconciliation

Although the initial lists were the source of most suburban addresses, as in urban areas, significant additional gains in address list coverage resulted from update activities. Future papers will document detailed levels of coverage gain.

**4.3 Rural Areas** This section describes the methodology used in rural mail areas. Note that this methodology is not applied in extremely rural areas where the mail census is not used. In rural mail census areas, computer files are not available that could facilitate automated geocoding. These areas contain non-city delivery addresses for postal delivery such as RFD (Rural Free Delivery), route and box numbers and General Delivery. Research has shown that inconsistencies exist between census and USPS addressing definitions and that often the USPS will be unable to deliver accurately to these areas. Address lists in these areas were developed without the involvement of the USPS. Census questionnaires were delivered by the Census Bureau.

In June and July of 1989, the Census Bureau hired about 10,000 census takers to compile an initial list of addresses in these areas. As in suburban areas, they travelled every passable road in their assigned area looking for houses and other places where people could live. For each house they found, they recorded the address and the householder's surname, if available. In addition, they placed a spot on their map showing the geographic location of the address. This combined address identification and geocoding activity produced the initial address lists in rural areas. Preliminary results show that approximately 10.1 million addresses were identified during the initial listing. Section 5.2 describes the final list update in rural areas.

## **5. FINAL UPDATING AND DELIVERY**

**5.1 Urban Areas - Final Geocoding** As mentioned earlier, an additional office and field geocoding operation was conducted prior to questionnaire delivery. This operation resulted in 0.5 million geocoded addresses being added to the census address list. These addresses originated from either the vendor list or the first postal check.

**5.2 Urban, Suburban and Rural Areas - Final Updating** The final update for the census address list involved approximately 96.2 million geocoded addresses. These addresses had been compiled using urban, suburban and rural methods.

In urban and suburban areas, the USPS performed the final address check in February 1990, just before census day. Implementing a procedure similar to the earlier postal check, the USPS identified missing addresses as well as undeliverable addresses which included duplicates. Missing addresses were geocoded before being added to the census address list. This check also provided information on areas where the USPS was more likely to have problems with questionnaire delivery.

In rural mail areas, field staff verified and updated the initial address lists and hand-delivered census questionnaires. In conjunction with updating address information and delivering questionnaires, the field staff updated the original maps. This was the only update to the census address list in rural mail areas.

After geocoding, unduplication and editing, approximately 2.7 million addresses were added in urban, suburban and rural areas from these final updates. Mailing packages were addressed and

delivered to these households. This represents an approximate 2.8 percent increase to the 96.2 million addresses that went into the final updates. These updates were designed to identify areas of new construction and change from earlier updates. It was expected that low growth areas would therefore show relatively low add rates.

These results varied at the state level with a low add rate of 0.7 percent for the District of Columbia and a high add rate of 5.0 percent for Kentucky. Since this was the third update in urban and suburban areas and the first update in rural areas, the areas with the greatest increase tended to be rural. Five out of seven of the states with high rates of coverage gain (4.1 to 5.0 percent increase) were states where an average of 41 percent of the total addresses were compiled using rural methods.

**5.3 Delivery** Two different delivery methods were used in mail areas. In some areas the USPS delivers questionnaires while in other areas this delivery is conducted by the Census Bureau. The choice of methodology is based on the ability of the USPS to recognize and accurately deliver to addresses compiled by the Census Bureau.

Mailing packages with census questionnaires were provided to the USPS for delivery. A mailing package was labelled for all addresses sent to the USPS for the final postal check and for most of the geocoded adds resulting from that postal check. Beginning on or about March 23, the USPS delivered questionnaires.

Questionnaires that were undeliverable were returned to the Census Bureau. The Census Bureau reviewed the undeliverable questionnaires and attempted to deliver them in those areas where it appeared that the USPS had delivery problems. We determined that some of the major reasons for non-delivery included vacancies and incorrect addresses (specifically apartment designations and ZIP Codes).

## **6. MAIL RESPONSE**

Starting in late March, respondents completed and returned Census questionnaires by mail. Within a month, approximately 65 percent of the estimated mailout was returned. This rate is calculated as the ratio of responses to the mailout. Included in the mailout however, are addresses for vacant units and units that will eventually be deleted. The most meaningful measure of respondent cooperation to the Census is the rate of response from occupied households. Based on recent counts of vacant and deleted addresses, the rate of response from occupied households can be estimated to be approximately 73.1%.

The highest response rates were found in Wisconsin, Iowa and Minnesota. The lowest rates of response were found in the District of Columbia, South Carolina, and Maine. A general trend by region is noted with the middle of the United States having higher than average response rates and both coasts and the south having relatively lower rates of response.

The 1990 state level response rates parallel the estimated 1980 state level response rates with an approximate 11 percentage point drop in response. Figure 6.2 depicts this trend indicating that relative response rates at the

state level did not change significantly from the 1980 census to the 1990 census. The lowest areas of 1980 response remained the lowest areas of response in 1990.

Despite lower than expected national response rates, over 26 percent of the district offices had response rates exceeding 1990 projections.

## **7. CONCLUSIONS**

Preliminary results indicate that varied address list development techniques are both required and effective in different parts of the country. Further analyses of these results will enable the Census Bureau to better target areas for specific coverage improvement activities. The use of the mail for delivery is still an effective procedure for most of the country. As exhibited by our methods in rural areas and our experience in some suburban and urban areas, for some types of addresses alternatives to USPS delivery are needed. Further research is needed to understand the issues related to effective USPS delivery.

Response to the mail census appears to be dropping from 1980 but until actual response measurements for occupied households are known, it is premature to speculate on the degree of decline. Additional research into the causes of census and mail survey nonresponse will allow the Census Bureau to assess the implications for future censuses.

Our evaluation of the 1990 coverage improvement program will cover many aspects of these early activities as well as detailed evaluations of special coverage improvement programs. Future papers will address costs and comparative coverage gains, sources of census coverage errors, and identify areas requiring research for the 2000 Decennial Census.

1/ Special procedures were used to obtain and geocode vendor addresses for Hawaii. The results for Hawaii are not included in this documentation.

## **8. REFERENCES**

- Harding, Dan N. "Final Report - Weekly Status Report for Geocoding Operation for Week Ending 1-April-89" 1990 Decennial Census Geocoding System Memorandum Series, Chapter 3, Document 10-45, dated 4/3/89, Washington, D.C.: U.S. Bureau of the Census
- Hazard, Joan "Results of the Advance Post Office Check for the 1988 Dress Rehearsal - Prelist Area" 1988 Preliminary Research and Evaluation Memorandum Series #4, dated 1/6/88, Washington, D.C.: U.S. Bureau of the Census
- Hazard, Joan "Results of the 1988 Dress Rehearsal Advance Post Office Check Reconciliation" 1988 Preliminary Research and Evaluation Memorandum Series #6, dated 3/21/88, Washington, D.C.: U.S. Bureau of the Census
- Miskura, Susan M. "Updated Major Program Components Document" 1990 Decennial Census Budget Memorandum #89-31, dated 10/30/89 Washington, D.C.: U.S. Bureau of the Census

Tillman, Amy "1990 Advance Post Office Check I Results at the District Office Level" 1990 REX Memorandum Series #E-4, dated 3/22/90, Washington, D.C.: U.S. Bureau of the Census

Tillman, Amy "Preliminary 1990 Advance Post Office Check I and Vendor File Geocoding Results." 1990 Decennial Census Preliminary Research and Evaluation Memorandum Series #24, dated 5/30/90, Washington, D.C.: U.S. Bureau of the Census

Tillman, Amy "Preliminary Results from the Precanvass Operation in the 1990 Decennial Census." 1990 REX Memorandum Series #G-3, dated 1/18/90 Washington, D.C.: U.S. Bureau of the Census

Treat, James B. "Evaluation of Commercial Address Lists for the 1990 Decennial Census" American Statistical Association 1988 Proceedings

Woltman, Henry F., 1990, "Comparing 1980 and 1990 Mail Response Rates" Washington, D.C.: U.S. Bureau of the Census

Woltman, Henry F., 1990, "Preliminary Address List Development Results." 1990 REX Memorandum Series #A-1. Washington, D.C.: U.S. Bureau of the Census