# Improving Coverage of Residential Address Lists in Multistage Area Samples 

Sylvia Dohrmann, Daifeng Han, Leyla Mohadjer<br>Sylvia Dohrmann, Westat, 1650 Research Boulevard, Rockville, Maryland 20850


#### Abstract

Address lists originating from the United States Postal Service (USPS) can be used as area sampling frames in place of on-site enumerations of dwelling units. While it has become clear that purchased USPS lists are less costly than the process of enumeration, it is less clear whether these lists are adequate as substitutes for them. We will evaluate the coverage of purchased lists for a selection of primary sampling units, (PSUs) differing in size and composition. We will do this by forming second stage units (SSUs) using Census geography and determining which of the USPS addresses fall into those SSUs using on-site enumeration. We will then focus on ways to improve the coverage of these lists and discuss the impact of discovering missed units on field operations. We will also examine practical ways of assigning selection probabilities to the missed units.


Keywords: Area samples, USPS, address lists
Commercially available address lists originate from the United States Postal Service's (USPS) Address Management System (AMS). The AMS contains all residential and commercial addresses along a mail delivery route, excluding government and military addresses. The advantages of using these lists in place of address listings collected manually by survey field staff are numerous. First and foremost is the cost. Address lists range from $\$ 15$ to $\$ 25$ per 1,000 addresses. While the cost could accumulate quickly in urban, densely populated areas, it is still low compared to the cost of the traditional listing process, including hiring, training, and supervising field staff. We have estimated that traditional listing is at least twice the cost of using purchased address lists. If costs for staff travel and lodging are included, traditional listing is at least four times the total cost. Staab and Iannacchione (2003) estimated traditional listing in their studies to be 10 times the cost of using purchased lists.

Speed and quality are other advantages. The address lists are in electronic format, and, once purchased, can be transmitted quickly from vendors. This saves the time and cost of converting addresses to electronic format, obviating the possibility of key-entry errors in cases of traditional paper and pencil listing. Finally, purchased addresses all have street numbers and names, which may
be difficult to obtain through traditional listing operations when they are not visible.

However, there are other considerations to be made before deciding to use the lists as a sampling frame for an area sample. Previous research has shown (Dohrmann, et al., 2006, Staab and Iannacchione, 2003, and O'Muircheartaigh, 2002) that the coverage of the lists may not be adequate in some areas, and especially for samples based on Census geography, units may not be properly placed into their second stage units (SSUs) sampled within selected geographic areas (or Primary Sampling Units or PSUs). In this paper we provide guidance in purchasing address lists for improved coverage (Section 1), and evaluate the coverage of purchased lists for a selection of PSUs differing in size and composition (Section 2). For the areas with adequate coverage, we further consider how residential address lists may be used and improved upon as frames for area samples of dwelling units based on Census geography (Sections 3 and 4).

## 1. Purchasing Address Lists

Address lists cannot be purchased directly from the USPS, but must be purchased through a third-party vendor. Generally a full ZIP Code is the smallest area for which an address list can be purchased; however, the purchase can be limited either to residential or commercial units as appropriate. Some vendors may provide the option of purchasing residential addresses within a particular Census geography of interest (block, block group, tract). Some vendors may offer to draw samples, but the vendors' sampling methodology should be reviewed thoroughly to ensure it meets the stated objectives. It may also be possible to sample from a vendor's list and then purchase only the sampled records. However, as discussed in Section 3.3, whenever only a portion of the DU frame is purchased, coverage issues may be exacerbated.

Coverage and accuracy are of major concern when purchasing address lists. When purchasing a list of addresses for a ZIP Code, several questions should be asked of the vendor to ensure that the vendor's list is both complete and current. In our previous work (Dohrmann, et al, 2006), we showed that coverage was best for
vendors licensed with the USPS to use the Computerized Delivery Sequence File (CDS) and not simply the Delivery Sequence File (DSF2). In order to have its list updated by the USPS, a vendor with this type of license must "own" (i.e. have at least $90 \%$ but no more than $110 \%$ of the addresses in) a particular ZIP Code. Also, its list may be updated every two months. So when purchasing, it is important to enquire whether the CDS licensed vendor owns the ZIP Codes of interest and has updated their lists within the last two months.

The residential lists generally include Post Office Boxes. However, most people who receive mail at a Post Office Box also live along a mail delivery route (Iannacchione, 2003). Also, Post Office Box numbers are not locatable within a geographic area, so including them in a purchase will not improve coverage. Seasonal and vacant units are flagged on the USPS files; it is recommended that these be included in a purchase since they may be occupied or become primary residences by the time the sample is in the field.

## 2. Coverage

There are several reasons why purchased residential address lists may not provide complete coverage for an area, especially in the context of area sampling. Since the AMS only contains residential addresses along a mail delivery route, those addresses for which mail service is not available will be excluded from the list. Residents living in these areas must collect their mail either at a general delivery facility or a Post Office Box. City-style addresses may not be available for housing units along rural routes. If not converted to city-style addressing, these units will be addressed similarly to postal boxes and are not locatable from the address alone.

In addition, purchased address lists may not include residential facilities within a business establishment or
group quarters such as dorms, assisted living facilities, halfway homes, and shelters. Whether these units are on the lists depends on how their residents receive mail. For example, residents of a dorm may receive their mail through a central campus post office via Post Office Boxes, thus not locatable. People living in other facilities may receive their mail through the business address and will not appear on a residential list. Group quarters could be included if business addresses are also purchased. However, this would significantly increase the amount of screening.

Rural areas are likely to have the least coverage (Dohrmann, et al, 2006, Staab and Iannacchione, 2003, and O'Muircheartaigh, 2002). This is expected to improve as more areas are converted to city-style addresses for purposes of E-911 location ${ }^{1}$. It is unclear, though, when that conversion will be completed. The USPS will not provide a timeline on conversion, but does periodically provide a list of ZIPs with converted addresses (http://www.usps.com/ncsc/addressservices/address qualityservices/lacsystem.htm). If a ZIP Code of interest appears on this list, it does not mean that the conversion has been completed in the ZIP Code, only that some addresses in that area have been converted.

### 2.1 Coverage Evaluation

In order to review the coverage of residential address lists, we examined six U.S. counties of varying size and urban population levels. These counties defined six PSUs from which a total of 161 SSUs were formed. Within the SSUs, 21,653 addresses were listed in the traditional manner. Characteristics of these counties are shown in Table 1.

[^0]Table 1. Characteristics of counties evaluated*

| PSU | Approximate <br> population | Percent population <br> living in urban area | Percent population <br> living in rural area | Population per <br> square mile |
| :--- | :---: | :---: | :---: | :---: |
| Urban/S. Atlantic | 700,000 | 99.6 | 0.4 | 2,483 |
| Urban/West | $>1$ Million | 98.9 | 0.7 | 2,344 |
| Urban/NE | 800,000 | 97.4 | 2.6 | 1,651 |
| Mixed/S. Central | 600,000 | 91.8 | 6.6 | 363 |
| Rural/S. Atlantic | 75,000 | 0.0 | 36.9 | 199 |
| Rural/Midwest | 30,000 | 0.0 | 71.1 | 28 |

[^1]The first three counties are quite large, urban, densely populated areas. The last two areas are much smaller and sparsely populated. The area labeled "Mixed/S. Central" is moderately sized, with over 90 percent of the population living in an urban area, but it is quite sparsely populated.

We reviewed the USPS website mentioned above containing ZIP Codes converted to city-style addressing
to determine if the ZIP Codes covering our SSUs had undergone some conversion. Table 2 shows the amount of address conversion for each of the six areas evaluated. All the areas studied had some amount of conversion with the Rural/Midwest area experiencing the most. As a result, we saw very few addresses in our purchased lists with rural route addressing. In fact, the only significant number of rural route addresses was found in the Mixed/S. Central area.

Table 2. Amount of rural route address conversion in evaluated areas

| PSU | Percent ZIPs purchased <br> with some conversion* | Ratio of converted to <br> purchased addresses in <br> converted ZIPs* | Percent RR addresses in <br> purchased ZIPs |
| :--- | :---: | :---: | :---: |
| Urban/S. Atlantic | 56.5 | $0.77 \%$ | 0.00 |
| Urban/West | 44.4 | $0.04 \%$ | 0.00 |
| Urban/NE | 5.3 | $0.01 \%$ | 0.00 |
| Mixed/S. Central | 100.0 | $12.17 \%$ | 5.87 |
| Rural/S. Atlantic | 100.0 | $13.91 \%$ | $<0.01$ |
| Rural/Midwest | 100.0 | $45.25 \%$ | 0.00 |

*Number converted is from USPS and may not be consistent/from the same time period as the purchased list.

### 2.2 Evaluation Procedures

We purchased address lists at the ZIP Code level containing all residential addresses in all ZIP Codes covering the sampled SSUs. We also purchased ZIP Codes bordering the sampled SSUs to ensure we were not missing addresses as a result of incorrect ZIP Code assignment. The lists were purchased from a single vendor holding a license to the USPS CDS file.

In order to determine the coverage rate of the purchased lists, we compared the traditional listings prepared by field staff canvassing the SSUs to purchased lists, assuming the traditional listings were closer to the truth. While the field staff compiling the traditional listings may make errors, they perform this work as a full-time occupation and are extremely well-trained and qualified.

We matched the enumerated lists with the purchased lists for each of the three areas in several ways. First, the enumerated and purchased lists were merged together by all address fields (street number, street name, pre- and post-direction, unit number, and ZIP Code). Any enumerated addresses that failed to match on all fields were merged again to the purchased list by the same fields, excluding the unit number/designator. Any enumerated addresses failing this match were merged a third time by geocoded latitude and longitude (obtained by our internal GIS), where possible. These second and third matching steps were conducted to overcome any
differences in unit designations (such as apartment A, B, C vs. 1, 2, 3), or spelling (such as Ft. Meyer Blvd vs. Fort Meyer Blvd). Any remaining unmatched records were investigated manually.

Some addresses listed traditionally were found on the purchased lists only after manual investigation. In many cases, the listers had misspelled the street name or street suffix, or these attributes were in a different format than on the purchased lists ("St. Thomas" rather than "Saint Thomas" or "TERR" rather than "TER"), and the cases could not be matched by geocoordinates. Another cause for automated match failure was that the listers often attached the incorrect ZIP Code to an address.

### 2.3 Evaluation Results

Table 3 shows the match rates for the six PSUs evaluated. The match rate in the second column of the table is defined as the number of matches between the two sources divided by the total number of addresses traditionally listed. These rates exclude addresses known to be in group quarters. Such units were identified in the Urban/West and Urban/NE PSUs. Not surprisingly, the urban areas had the highest match rates, all exceeding 97 percent. The rural areas had rates below 75 percent. The Mixed/S. Central area, even though most of the population lives inside an urban portion of the county, had a rate equivalent to that of the rural areas. We know this area has some rural route addresses on the purchased list.

However, less than 2 percent of the traditionally listed addresses were along rural routes, so it is unlikely that this alone accounts for the poor match rate.

In rural areas, some street names were simply not found on the purchased lists. This could be the result of E-911 conversion street name changes occurring in the area, but not yet incorporated into the AMS. In the Rural/S. Atlantic PSU over 22 percent of the addresses on the traditional lists were on streets not found on the USPS list.

In some cases, the street name was on the purchased list, but not a particular street number. A search on the USPS website indicated that some of these addresses were "nondeliverable." Such addresses are likely newly constructed, but not yet occupied units. In such cases, the listers are instructed to include these addresses on their lists since they might be occupied during the field period. These "nondeliverable" addresses could also be extended vacancies that are no longer maintained on the AMS.

Table 3. Match rates between traditional and purchased lists for the selected areas

| PSU | Match rate* | Match rate among <br> "matchable" cases only |
| :--- | :---: | :---: |
| Urban/S. Atlantic | $98.3 \%$ | $98.6 \%$ |
| Urban/West | $97.2 \%$ | $98.4 \%$ |
| Urban/NE | $99.1 \%$ | $99.8 \%$ |
| Mixed/S. Central | $75.0 \%$ | $87.5 \%$ |
| Rural/S. Atlantic | $65.4 \%$ | $72.8 \%$ |
| Rural/Midwest | $74.9 \%$ | $77.0 \%$ |

* Match rates exclude group quarters which were found in Urban/West and Urban/NE PSUs. The match rates with group quarters for these PSUs are 94.8 percent and 79.1 percent respectively.

In rural areas especially, it may not be possible for the listers to capture the entire address as the street number or street name are not apparent. In such cases, the listers provide a description of the unit so that interviewers may locate it at the time of interviewing. We cannot expect that these units would be matched to the USPS lists. If we remove such cases from the denominator of the match rate and re-calculate the rate as the number of matches divided by the total number of "matchable" addresses listed, we obtain the match rates in third column of Table 3. These are improved from the earlier match rates. It appears that the poor match rate for the Mixed/S. Central area was partly due to the listers inability to locate adequate address information. However, in the more rural areas, the rates are still well below 80 percent.

Given the high match rates in the urban, densely populated areas, it seems that purchased address lists could serve as adequate sampling frames in these areas. The next section will discuss how these lists might be used in such areas for this purpose.

## 3. Purchased Address Lists in Practice

Since the address lists are based on ZIP Code, this would seem a logical sampling unit and was used as such in a national household survey conducted in 2002 by Staab and Iannacchione (2003). However, ZIP Code-based sampling does not fit well into the multi-stage area sampling framework usually used for in-person area surveys. ZIP Codes are not spatial or geographic entities but simply categories for grouping mailing addresses. ZIP Codes themselves are also quite large. Table 4 shows the relative size of ZIP Codes as compared to Census blocks, the more usual basis for building SSUs.

As can be seen in the table, ZIP Codes would form excessively large SSUs, so as to be completely impractical for field activities. The more detailed ZIP+4 designation, such as 20850-3195, may seem to be the solution to large ZIP Code-based SSUs. The four-digit extension is usually defined as a block face and may contain very few units. Since the numbering system for the +4 suffix is not sequential, that is, adjacent homes may not have sequentially numbered ZIP+4 categories, combining ZIP+4 categories to form efficient SSUs would also prove difficult. ZIP Codes can also cross county lines, so PSUs in such a design must also be defined in terms of ZIP Codes. All these reasons tend to make ZIP Code-based sampling impractical for most purposes.

Table 4. Average number of dwelling units in ZIP Codes and Census blocks

| County population <br> density (number of <br> persons per square mile) | Average number <br> of dwelling units* |  |
| :--- | :---: | :---: |
|  | ZIP Code | Census block |
| $5,000-9,999$ | 15,821 | 86 |
| $1,000-4,999$ | 8,734 | 43 |
| $400-999$ | 6,612 | 36 |

* Data calculated from 2000 Census SF1 file using Census ZIP Code Tabulation Areas (ZCTAs) as an approximation for ZIP Code. Excludes Census blocks with no population.

A more common method for forming SSUs for the purposes of area sampling uses Census geographic boundaries. To use the purchased address lists as a sampling frame with such geographically based SSUs,
one could purchase addresses for the entire PSU, form the SSUs, and then place the purchased addresses into the entire SSU frame; however, this method would be very costly. Alternatively, one could first draw a sample of SSUs, purchase addresses in ZIP Codes covering the sampled SSUs, attach geocoordinates to each address, and then assign each address to the selected SSUs based on their geography. If the relationship between selected SSUs and ZIP Codes is uncertain, such as SSUs appear to lie close to the border of a ZIP Code, more ZIP Codes should be purchased to assure coverage.

Regardless of the ordering of these steps, there are two issues that must be addressed before overlaying the address lists onto geographic SSUs: not all addresses can be geocoded and not all geocoded addresses may be placed into the correct SSUs. We discuss each of these in turn below.

### 3.1. Geocoding and Geographic SSUs

The geocoding process is not an absolute one. It cannot be thought of in terms of "rooftop" geocoding, with latitude and longitude attached to an address at the exact position that the unit appears on a street. Rather, the geocoding process requires matching a particular address to a geographic information system database containing a mapping of the street network geographic coordinate space. In the database, portions of each street (usually called a "street segment") are attributed with street number ranges. When an address is geocoded, it is matched to a street and street segment with the street number range containing the address of interest, rather than matching to a specific address in the database. The position of the address is then determined by interpolating within the range along the street segment.

As a result, in many cases, especially those where the addresses are not evenly dispersed along a street, the specific geographic coordinates assigned to an address may not be at the precise location of the structure with that address. The structure may lie several yards down the street or even across the street.

Once an address has geocoordinates attached to it, it can be matched to other geogrpahic entities such as the Census geographic designations usually used to form the SSUs. However, since the geocoding process is not exact, neither is the matching to the Census geography. As a result, not every geocoded address will be placed into the correct SSU. However, every geocodable unit will receive a chance of selection as long as SSUs are defined by where the units are geocoded.

### 3.2. Nongeocodable Addresses

An address will not be geocodable if its street name or number is not found in the geocoding database. This is simply the result of the geocoding database not being as current as the address list, or vice versa. For example, units in a new housing development on newly constructed streets may not be found in the geocoding database. Newly constructed housings units along an existing street may also not be found. Rural routes recently converted to city-style addressing may not be incorporated into the database.

In urban areas, where the address lists appear to have the best coverage, the percentage of addresses that are not geocodable is likely to be quite small. The number of units, however, may be quite large relative to the sample. Table 5 contains the percentage and number of purchased addresses that were not geocodable in the urban areas evaluated.

Table 5. Purchased addresses not geocoded in selected urban areas

| PSU | Percent not <br> geocodable | Number of <br> units |
| :--- | :---: | :---: |
| Urban/S. Atlantic | 4.9 | 14,251 |
| Urban/West | 0.2 | 1,190 |
| Urban/NE | 1.4 | 3,251 |

If we cannot geocode some DUs, we cannot place them into a particular SSU and give them a chance of selection. Including all these units into the sample is clearly impractical. Rather, we must include them in the DU sampling frame and give them an appropriate chance of selection so we can bring a representative portion of them into the sample.

As discussed in Section 3.1, all geocodable DUs assigned to sampled SSUs can be sampled within that SSU. As shown in Equation (1), the probability of sampling a geocoded DU is the multiplication of the probability of SSU selection and the probability of sampling the DU within the sampled SSU.

$$
\begin{align*}
\operatorname{Pr}(\text { Geocodable DU })= & \operatorname{Pr}\left(\mathrm{SSU}_{i}\right) * \operatorname{Pr}(\text { Geocodable } \\
& \left.\mathrm{DU} \mid \mathrm{SSU}_{i}\right) \tag{1}
\end{align*}
$$

For the nongeocodable DUs, the finest geographic information we can obtain is their ZIP Codes, so these cases will be brought into the sample through the ZIP Codes overlapping the sampled SSUs. Similar to

Equation (1), we can calculate the selection probability of these cases using the knowledge of conditional probability:
$\operatorname{Pr}($ Nongeocodable DU$)=\operatorname{Pr}\left(\mathrm{ZIP}_{j}\right.$ coming into the sample $)$

* $\operatorname{Pr}$ (Nongeocodable DU | $\mathrm{ZIP}_{j}$ coming into the sample)
where the probability of $\mathrm{ZIP}_{j}$ coming into the sample equals to the probability of at least one of the SSUs overlapping $\mathrm{ZIP}_{j}$ being selected, as shown in Equation (3)

$$
\begin{equation*}
\operatorname{Pr}\left(\mathrm{ZIP}_{j} \text { coming into the sample }\right)=1-\prod_{i}\left(1-\operatorname{Pr}\left(\mathrm{SSU}_{i}\right)\right) \tag{3}
\end{equation*}
$$

where $i$ is any SSU that overlaps $\mathrm{ZIP}_{j}$.
To obtain a self-weighting sample, we would set (1) = (2). This will allow us to solve the conditional probability of selecting the nongeocodable DUs within $\mathrm{ZIP}_{j}$. By applying this conditional probability to the total number of nongeocodable cases in $\mathrm{ZIP}_{j}$ we can obtain the expected number of nongeocodable DUs that will come into the sample through each ZIP Code.

The result is a portion of the sample consisting of individual, likely non-neighboring dwelling units. These units may be more difficult and time consuming for field staff to locate. However, based on the probabilities above, we estimate that only a small number of these will be selected.

### 3.3. Missed Units

The remaining issue regards units existing within the geographic boundaries of sampled SSUs that are not on the purchased lists. As discussed in Section 2, the lists have quite good coverage in urban areas, but do not provide complete coverage. Units not on the vendor lists will be either entire structures, such as single-family homes or apartment buildings, or units within a known structure (whose address is on the list), such as individual apartments or trailers in a trailer park.

A quality check in the field is necessary to capture units not on the purchased list. Field staff must be able to identify any missed units as either truly missed or elsewhere on the purchased lists and simply placed outside the SSU as a result of the inherent geocoding difficulty. In terms of locating the missed structures, organizations use different procedures some of which were discussed in Dohrmann, et al (2006). Regardless of the specific method, we suggest that the field staff perform the check, respecting the Census or geographic boundaries. To do this, staff will need a list of units that were part of the geocoded DU frame for that SSU, and some means to check the entire address list frame.

Consider the geographic SSU indicated by the white area in Figure 1. If field staff find a structure such as the one indicated by the + , they will note that it is neither on the DU frame for the SSU nor anywhere on the purchased list, so it needs to be added to the sample. When they come across units at the southern boundary of the SSU, they will find that, while not on the DU frame for the SSU, the units are elsewhere on the list. Thus, the units were given a chance of selection, either by being geocoded into another SSU, or if not geocodable, by the separate sampling for these units (as described in Section 3.2). Staff will find the unit near the western-most edge of the SSU physically outside the boundary, but part of the SSU in terms of its geocoordinates. The SSU boundary is then more accurately represented by the dashed lines in Figure 1.

Missed units within known structures would be brought into the sample in the same manner using the application of sampling procedures currently used by the organization. If missed units are too numerous some subsampling may be required, since cost, staffing, or time concerns may preclude including all the additional units. Subsampling the missed units introduces weight variability. On the other hand, adding all missed units into the sample increases clustering. The appropriate course of action will be determined by the differing effects of the clustering or the subsampling on the variance.


Figure 1. Example of missed structure check in one SSU

It is important to note that a quality check such as the one described above is not possible if only a selected sample of addresses is purchased, or if lists are purchased by Census geography, since we know that units along the boundaries of Census geography are likely to be geocoded into another SSU.

## 4. Lists as Sampling Frames for Urban Areas: An Illustration

To illustrate the use of purchased address lists for urban areas, we combined the three urban areas evaluated into one fictitious PSU and stepped through the process described in Section 3. Together, the 3 areas would form a sampling frame of over 27,000 SSUs based on geography from which 81 are selected. Table 6 shows the distribution of the combined purchased list at each of the sampling steps.

Purchased lists covering all ZIP Codes for the combined PSUs totaled approximately 1.1 million addresses. Most of these cases are geocodable; only less than 2 percent of the lists were not geocodable. As shown in Table 5, this results in a large number of addresses that are not geocodable. While most of the address list is geocodable, less than 1 percent of the addresses geocode into the sampled SSUs, again illustrating the large difference in the size of the ZIP Codes and Census geographic SSUs.

After setting aside the addresses which geocode outside the SSUs, the result is a DU sampling frame consisting of 2.6 percent of the purchased list. It is interesting that even though over 97 percent of this list is not used for sampling, the method is still less costly than traditional listing. Most of the DU frame (close to 70\%) consists of units we cannot geocode. Using the method described in Section 3.2, we estimate that only 4.2 percent of the sample will consist of nongeocodable units.

## Section on Survey Research Methods

Table 6. Distribution of the purchased list addresses across sampling steps

|  | Attach <br> geocoordinates | Assign each <br> address to <br> sampled SSUs | DU Frame | DU Sample |
| :--- | :---: | :---: | :---: | :---: |
| Geocoded | $98.2 \%$ |  |  |  |
| Geocoded outside sampled SSUs |  | $97.4 \%$ |  |  |
| Geocoded into sampled SSUs | $1.8 \%$ | $0.8 \%$ | $30.8 \%$ | $95.8 \%$ |
| Not Geocodable |  | $1.8 \%$ | $69.2 \%$ | $4.2 \%$ |
| Missed units |  |  | $1.1 \%$ | $11.6 \%$ |

Fifty-one of the 81 SSUs (63\%) in the urban PSUs, included missed units. On average, there were three missed units per SSU. Most of these missed units were in structures that were included in the purchased lists. Either there were no subunits associated with the address, or there were subunits, but not that particular unit. SSUs with missing subunits in known structures had the most missed units, with one SSU missing a total of 27 such units. While these cases are just as problematic as other types of missed units, it may be helpful to recognize the potential for a greater number of missed units in SSUs with multi-unit structures.

## 5. Summary and Conclusions

The benefits of using the purchased address lists are clear. The reduced cost, ready access to quality electronic addresses, and the promise of obtaining the sample in a much shorter period of time promote use of these lists as DU frames. However, it is important to consider the limitations of these lists and the consequences of using them as sampling frames for DUs especially in area samples.

Our research has shown that lists purchased at the ZIP Code level from a CDS licensed vendor can provide quite good coverage in urban, densely populated areas. In practice, the lists must be geocoded and then placed into the sampled SSUs. Since some cases cannot be geocoded, a separate sampling step must be performed for these units so that they may be represented in the sample. In addition, a quality check in the field, with respect to the SSU geography, must be performed to include units in the SSUs not found on the purchased list.

The limited coverage of the lists is their greatest shortcoming. Their use is generally restricted to urban, densely populated areas. An additional complication for area sampling is that only a portion of the list will be linkable to the geography used to form SSUs. Purchased lists could be incorporated into area samples if these two
hindrances are addressed by giving the nongeocodable units a chance of selection and performing a concerted missed unit effort in the field. Some of the time saved by using the lists, instead of creating lists manually, may be spent locating the nongeocodable units (as they will not be clustered inside the samlpled SSUs) and checking for potentially missed units. As a rule, it is quite likely that the resources needed for these tasks will not exceed those needed for traditional listing.

Recent research into the use of purchased address lists as sampling frames has created opportunities for some studies that may not have been feasible otherwise. The lists are well suited for mail surveys, for example, in cases in which ZIP Codes appear to be efficient SSUs and when missed units may be easily incorporated into the sample. Whether these lists provide the makings of a sampling revolution in the field of area samples remains to be seen, as these lists become more widely used and evaluated.

## References

Dohrmann, S., Han, D., and Mohadjer, L. (2006). Residential Address Lists vs. Traditional Listing: Enumerating Households and Group Quarters. Proceedings of the Survey Research Methods Section of the American Statistical Association, 2959-2964.
Iannacchione, V.G., Staab, J.M., and Redden, D.T. (2003). Evaluating the Use of Residential Mailing Addresses in a Metropolitan Household Survey. Public Opinion Quarterly, 67, 202-210.
O’Muircheartaigh, C., Eckman, S., and Weiss, C. (2002). Traditional and Enhanced Field Listing for Probability Sampling. Proceedings of the Social Statistics Section of the American Statistical Association, 2563-2567.
Staab, J.M., and Iannacchione, V.G. (2003). Evaluating the Use of Residential Mailing Addresses in a National Household Survey. Proceedings of the Survey Research Methods Section of the American Statistical Association, 4028-4033.


[^0]:    ${ }^{1}$ E-911 location refers to the ability of 911 emergency service vehicles to locate physical locations based on street addresses.

[^1]:    * Data are from the 2000 decennial census. Population percentages may not sum to 100 percent since a third Census Bureau category, population living inside urban clusters, is omitted from this table.

