

Tracking and Evaluating Changes to Address-Based Sampling Frames over Time

Jennifer Unangst¹, Joseph McMichael¹

¹RTI International, 3040 Cornwallis Road, Research Triangle Park, NC 27709

Abstract

Address-based sampling frames created from the USPS Computerized Delivery Sequence (CDS) file are updated by the USPS twice monthly. With each update, addresses are added, deleted, revised and/or moved between the CDS and No-Stat files. To assess the quality and coverage of the ABS frame, it is important for researchers to understand the frequency and scope of these changes. This paper tracks more than 130 million addresses from the national ABS frame over time from 2012 to 2014. We detail where and when changes were made, as well as the type of changes (e.g., whether a change was an addition, deletion, or vacancy status change). This analysis has never been available to the public and will aid researchers in study planning where ABS frames are being considered.

Key Words: Address-Based Sampling, ABS

1. Introduction

Address-based sampling (ABS) frames are residential household frames suitable for in-person, mail contact, and multimode surveys. ABS frames are based on lists products managed by the United States Postal Service (USPS) – primarily, the Computerized Delivery Sequence (CDS) file, which contains the set of delivery points receiving mail. The supplemental No-Stat file, another list product managed by the USPS, contains delivery points which are not enumerated on the CDS file. It is expected that the universe of residences in the U.S. will change over time as housing units are built, torn down, or change vacancy status. These changes can impact the desired sampling population for a study (e.g., a researcher may want to exclude households expected to be vacant from the pool of households that are eligible to be sampled). Furthermore, changes in the residential household population which occur in the time period between frame construction and data collection can contribute to wasted resources and potential over- or under-coverage. Thus, it is desirable to understand and evaluate the types of changes that occur in an ABS frame over time. This paper includes the results of an exploratory exercise seeking to characterize some of the changes occurring in the files.

1.1 Address Types in the USPS Files

The CDS file contains over 130 million delivery points in the U.S., while the No-Stat file includes just over 8 million delivery points. Delivery points in the USPS lists can be categorized into several types of addresses.

- *City-style addresses*: These delivery points include a street address, city, state, and zip code, so they can be located in the field for an in-person or mail survey.
- *Drop Points*: These are delivery points where mail is received on behalf of multiple associated residences (e.g., gated communities, apartments). Drop points pose a challenge for mail surveys, since individual residences are not distinguished.
- *P.O. Boxes*: Post-office boxes are delivery points located at the local post office and do not provide information on the location of the associated household; thus, these addresses are not locatable for an in-person survey. P.O. boxes can be the “only way to get mail (OWGM)” for a residence, or they can be a supplemental delivery point linked to a city-style address already on the USPS lists. The CDS file includes a flag to distinguish OWGM addresses.
- *Rural route (RR) and highway contract (HC) boxes*: These are delivery points which are not curbside to the associated household and are not locatable for an in-person survey.
- *Simplified addresses*: These delivery points may only include the name of the resident, city, state, and zip code and would not be locatable for mail or in-person surveys.

When building an ABS frame, it is unlikely that a researcher will want to include all address types in these lists. Exclusions for an ABS frame are study dependent and can be influenced by cost and data collection mode. Iannacchione (2011) provides a more detailed discussion of the relationship between survey mode and address types on the CDS and No-Stat files.

2. Methods

RTI International receives an updated commercial version of the CDS and No-Stat files monthly from Compact Information Systems (CIS). For this paper, we created ABS frames at quarterly time points between October 2012 and October 2014, which represents a recent history of the files. The ABS frames used in this analysis are derived exclusively from the CDS file, except in the discussion of “new growth” addresses on the No-Stat file in Section 3.3.

Each frame was constructed by excluding addresses flagged as non-residential (i.e., businesses), P.O. boxes affiliated with a city style address (i.e., not “Only Way to Get Mail” addresses), and simplified addresses. After applying these exclusions, our baseline frame in October 2012, derived from the CDS file, contained 98.7% city-style addresses, 1.2% highway contract and rural route boxes, and 0.1% OWGM addresses.

We then matched delivery points across the frames by their full address after accounting for zip code changes and splits.¹

¹ Full address includes street address, city, and zip code.

3. Changes Observed in the Frame

The types of changes which may be observed between frames include the addition/deletion of addresses, vacancy status changes, changes in the number of household units associated with drop points, and movement of addresses between the CDS and No-Stat files. This analysis focuses on three types of changes in the frames: (1) the types of addresses being added and deleted, (2) changes in vacancy status, and (3) the movement of addresses flagged as “new growth” on the No-Stat file.

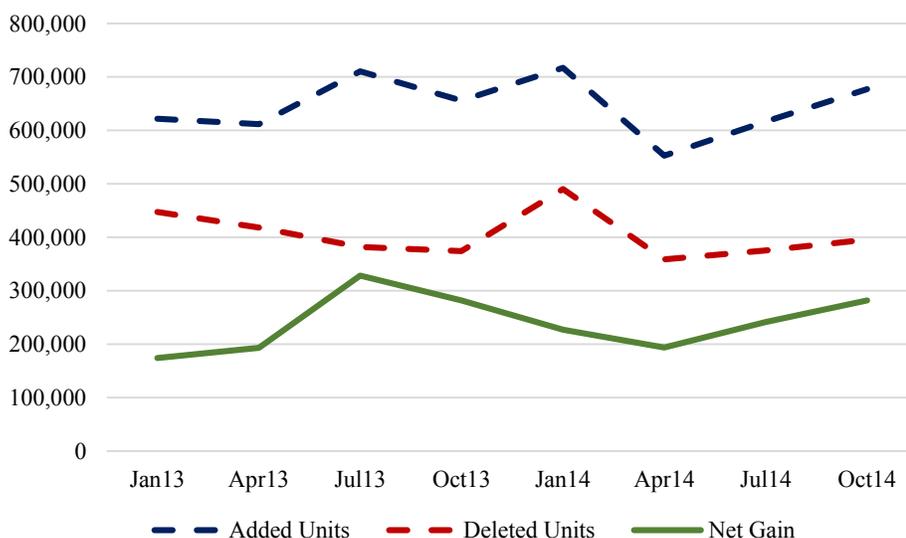
In the following sections, counts of address units are weighted such that drop point addresses on the CDS file represent the number of associated drop units. Drop points will not be distinguished from city-style addresses for the duration of this paper.

3.1 Added and Deleted Address Units

The processes of residential construction and teardowns lead us to expect that each frame will demonstrate both new and lost delivery points relative to prior frames. Between the October 2012 and October 2014 frames, which are based exclusively on the CDS file, we observed a net gain of 1.92 million address units. This increase represents a meager 1.54% increase given the size of the frame. Thus, despite delivery points being added and deleted at each quarterly time point, the number of added units tended to exceed the number of deleted units over time.

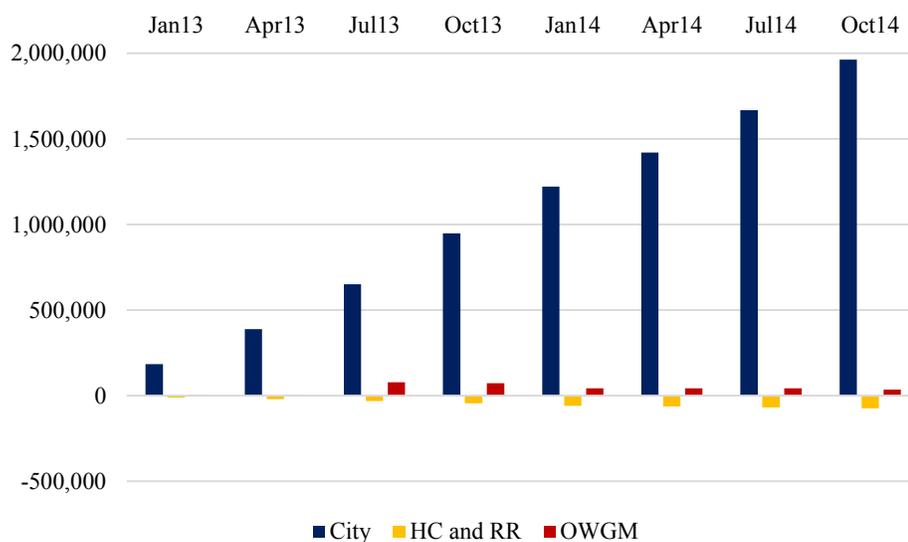
Figure 1 illustrates the variability in the number of address units added and deleted between each quarterly time point. The first time point shown in the figure, January 2013, displays the number of units added and deleted relative to the baseline October 2012 frame. The second time point shown, April 2013, displays the number of units added and deleted relative to the prior time point, January 2013, and so on. Given the scale of the frame, these additions and deletions are minor; however, these frame-to-frame changes contribute to the overall trend of an increasing number of units on the frames.

Figure 1: Number of Added and Deleted Address Units between Frames



We further studied whether the composition of the frame by address type is changing over time. Figure 2 shows the cumulative change in the number of address units relative to the baseline October 2012 frame. The changes are displayed by address type: city-style (city), highway contract and rural route boxes (HC and RR), and P.O. boxes that are the “only way to get mail (OWGM).” Of the 1.92 million units gained between October 2012 and October 2014, majority were city-style addresses. This is expected given that 98% of the frame is city-style addresses. Furthermore, despite the net gain of city-style addresses, we observed a loss of highway contract and rural route boxes. We hypothesize that the households associated with the lost HC and RR addresses are in fact still represented on the frame among the set of gained city-style addresses. This may be a result of e911 address conversion efforts designed to improve ability to locate households in an emergency. This finding bodes well for survey researchers as the percentage of locatable address types on the ABS frame appears to be improving over time.

Figure 2: Cumulative Change by Address Type Relative to October 2012 Frame



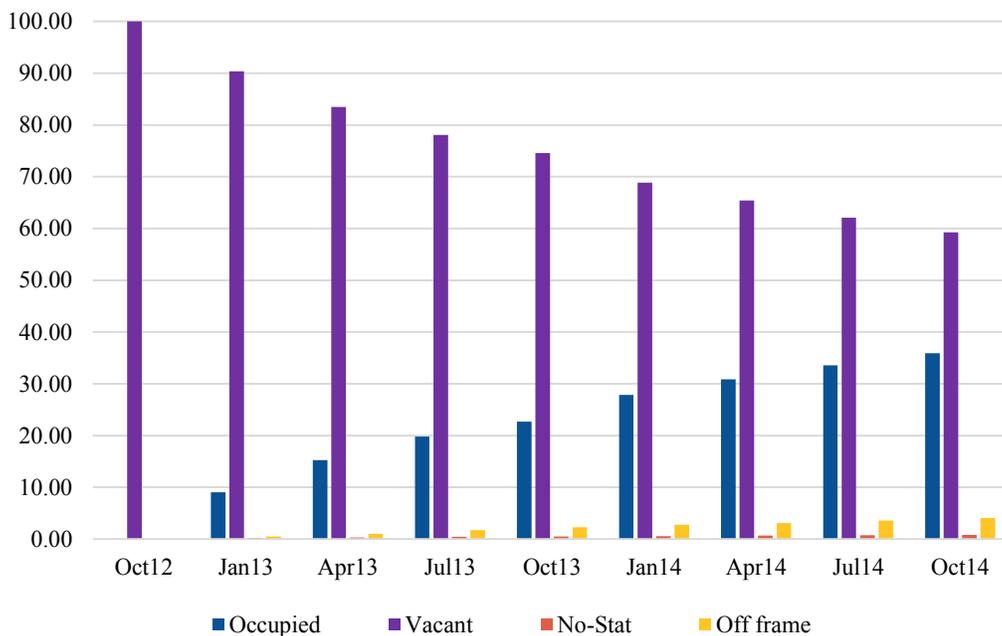
3.2 Vacancy Status Changes

The *Vacant Indicator* on the CDS file is used to identify addresses which have been vacant for at least 90 days (USPS 2013). Across the quarterly time points considered in this paper, the rate of vacancy in the ABS frame ranged from 3.2% to 3.5%, which represents 4.0 million to 4.3 million address units, respectively. When researchers are creating an ABS frame, they must decide whether or not to include vacant addresses in the frame. Cost and mode will, of course, influence this decision. Nevertheless, understanding how vacant addresses are expected to change over time may provide some insight.

Figure 3 follows vacant addresses from the October 2012 frame across time and shows the percentage of units which remain vacant, become occupied, move to the No-Stat file, or are deleted from the frame. Note that each of the time points in the figure represents the same set of address units – the addresses flagged as vacant on the October 2012 frame. In October 2012, 100% of the address units are vacant, since we’ve conditioned on this set of units.

As expected, we generally observed that more vacant units become occupied as time passes, although the rate of this change may be slow. In the October 2013 frame, only 23% of the addresses flagged as vacant a year earlier had become occupied. Also, interestingly, of the 59% which are shown as vacant in October 2014, 95% of those remained vacant at every intermediate quarterly time point, suggesting no change in vacancy status for those units across two years.

Figure 3: Changes to Vacant Addresses from the October 2012 CDS-based Frame over Time



3.3 Movement of “New Growth” from the No-Stat File

Shook-Sa et al (2013) has suggested that supplementing the CDS file with active addresses from the No-Stat file provides modest coverage gains for the ABS frame. The *New Growth Indicator* on the No-Stat file is used to identify units under construction which are expected to begin receiving mail in the near future (USPS 2013). New growth addresses are not actively receiving mail and thus would fall outside the scope of Shook-Sa’s recommendation; however, they may represent households which should be accounted for if there is a significant delay between frame cleaning and field work (i.e., new construction may become occupied and become in-scope for the study).

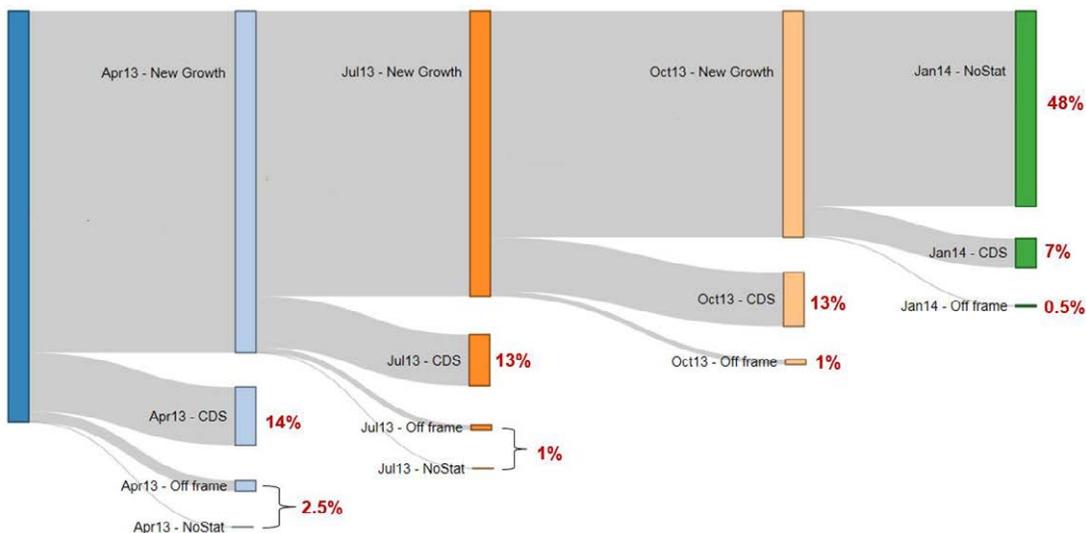
Across the quarterly time points we considered, the percent of addresses on the No-Stat file flagged as new growth varied from 0.9% to 6.4%, which represents 70,000 to 500,000 units, respectively. Given the magnitude of our CDS-based frame which has over 125 million units after exclusions, any coverage impacts from excluding new growth on the No-Stat file would likely be negligible. Nevertheless, we studied the movement of new growth addresses over time to further our understanding of this set of units.

Figure 4 follows the addresses flagged as new growth in the January 2013 No-Stat file over time. Each segmented column in the figure represents a quarterly time point with the first dark blue column representing all records flagged as new growth in January 2013. Possible

changes to the new growth address units include moving to the CDS file (CDS), remaining on the No-Stat file but losing the new growth flag (NoStat), and being removed from both the CDS and No-Stat files (Off frame). An address which does not undergo one of these changes remains on the No-Stat file flagged as new growth (New Growth). In the figure, once an address experiences a change, it resides in the terminal node that reflects its change. Thus, the set of address units at each time point is a subset of those which existed at the prior time point. The sum of the percentages displayed at the terminal nodes yields 100%.

As Figure 4 shows, of the 70,000 addresses flagged as new growth in January 2013 (i.e., represented by the first dark blue column), 14% moved to the CDS file by April 2013; 27% moved to the CDS file by July 2013; 40% moved to the CDS file by October 2013; and 47% moved to the CDS file by January 2014. Thus, in a year's time, nearly half of these units might have been included in a CDS-based frame. Furthermore, 48% of the new growth remained on the No-Stat file after a year but were no longer categorized as new growth. This suggests that new growth addresses on the No-Stat file tend to undergo a change such as movement to the CDS file, deletion from the USPS lists, or losing the new growth flag within a year.

Figure 4: Changes to Addresses Flagged as New Growth on the January 2013 No-Stat File



4. Discussion

Changes in the composition of address types on the ABS frame, such as the reduction of unlocatable highway contract and rural route boxes, suggest that address quality is improving for use by survey researchers. However, researchers must still make decisions when creating their frame. For example, the trade-offs between coverage and cost must be considered when determining whether to include vacant addresses or to use addresses on the No-Stat file to supplement a CDS-based frame. Excluding vacant addresses may be an option for surveys where cost constraints outweigh coverage concerns, given the high rate of vacant addresses still flagged as vacant two years later. Furthermore, the small number of addresses originally flagged as new growth on the No-Stat file which are moved to the

CDS file within a year may not justify the additional processing time and costs of including these records in an ABS frame.

One aspect of this analysis which remains unclear is the potential impact of clustering on the changes observed in the USPS files. Surveys with a concentrated target population (e.g., a specific state in the U.S.) may experience greater coverage loss from decisions to exclude vacant or unlocatable address types from the frame.

References

- Iannacchione, V. 2011. "The Changing Role of Address-Based Sampling in Survey Research." *Public Opinion Quarterly* 75:556-575.
- Shook-Sa, B., Currivan, D., McMichael, J. P., Iannacchione, V. 2013. "Extending the Coverage of Address-Based Sampling Frames." *Public Opinion Quarterly* 77:994-1005.
- United States Postal Service [USPS]. 2013. "CDS User Guide. https://ribbs.usps.gov/cds/documents/tech_guides/CDS_USER_GUIDE.PDF