May 2017

Address-based Sampling Atlas Series: Drop Points

White Paper

Prepared by

Ashley Amaya
RTI International
3040 E. Cornwallis Road
Research Triangle Park, NC 27709

Address-based Sampling Atlas Series: Drop Points

White Paper

May 2017

Prepared by

Ashley Amaya
RTI International
3040 E. Cornwallis Road
Research Triangle Park, NC 27709

CONTENTS

FIGURES

TABLES

[This page intentionally left blank.]

1. DROP POINTS DEFINITION

The Computerized Delivery Sequence (CDS) file contains listings for nearly all addresses in the United States, both residential and business.[1] Although most residential listings are city-style addresses (e.g., 123 Main St., Apt A, Anytown, MA 01234), other types of addresses exist. Among the alternative address types are **drop points**, a mail receptacle that is shared by multiple housing units (i.e., drop units). The **drop units** have a street number and name but are missing a unit or apartment number. Drop points make up 0.5% (n=737,196) of all unique residential addresses,[2] while drop units account for 1.5% (n=2,166,062) of all residential addresses.[3] Since each drop unit within a drop point has an identical street address (e.g., 123 Main St.) but is listed individually on the frame, the drop units for a given drop point appear as duplicates. Drop units can also be identified on the frame by a flag, DROP. If DROP is "2" or "Y" (depending on the coding scheme), the housing unit is a drop unit.[4,5]

2. CHALLENGES

In general, drop units pose a challenge to researchers because of their "one-to-many" relationship between address and housing unit: one unique address corresponds to many housing units. This relationship causes problems in different ways based on the mode of data collection.

---

[1] For simplicity's sake, we refer to the frame as the CDS. In practice, the frame used by researchers is obtained from a sample vendor and is based largely, but not exclusively, on the CDS. For more details on the distinction, please refer to Section 2.1 of the Address-based Sampling Task Force Report (American Association for Public Opinion Research, 2016).

[2] All data were pulled from the January 2017 CDS.

[3] Unique addresses have a unique combination of ADDRESS1, ADDRESS2, CITY, STATE, and ZIP on the frame. Residential addresses include all housing units, even if there are duplicate addresses. A drop point building with five units would count as one unique address but five residential addresses. A typical multi-unit building with five units would count as five unique addresses and five residential addresses. Both unique addresses and residential addresses include addresses listed as all residential, mostly residential, and mostly business.

[4] In rare cases, a building or street address may be mixed, containing both units that have a unit number and others that do not. This occurs when a single building has two (or more) mail receptacles, one used by a single housing unit and the other by two or more drop units. This phenomenon typically occurs when there are multiple entrances to a building: one entrance and mail receptacle may be used only by the basement unit, for example, while the other is used by all other units in the building (Amaya, LeClere, Fiorio, & English, 2014).

[5] There are a few frame errors regarding drop points of which statisticians should be aware. There are 38 drop points that only have one drop unit. These should be considered single units, not drop points, and sampled accordingly. There are also 806 duplicate addresses that are not flagged as drop units but make up only 296 unique addresses. These should be deduplicated and treated as typical city-style addresses.

2.1   In-person

For field surveys, the challenge lies with the interviewer. If a drop unit is selected, the interviewer will have no indication which unit at the selected address to interview. Without additional instruction, he/she may opt to place the case on hold. If the case is not adequately managed and reactivated, the overall survey response rate may needlessly suffer. Alternatively, the interviewer may select a unit for which a person is at home and cooperative. This may result in self-selection bias because the "at-home" household may be different than the other households at the address (Eckman & Koch, 2016). The interviewer's action will also make the selection probabilities unknowable.

One approach to address this challenge is to sample the entire drop point, eliminating the need for the interviewer to choose a unit. Unfortunately, sampling all units at a drop point may be tedious for field interviewers in some areas. Nearly all drop points contain two (80.0%) or three (14.9%) drop units (Figure 1). These addresses are typically houses that have been converted into apartments. Approaching all units at these addresses to take part in the interview should be of little trouble to interviewers. However, the remaining 5.1% of drop points have four or more units, and they can pose problems for two reasons. First, some are quite large—up to 999 units. Interviewing all units would be statistically unnecessary, costly, and logistically infeasible. Second, larger drop points are much more variable in structure type. Large drop points are often gated communities, high rises, trailer parks, alternative housing such as half-way housing, or may not be a housing unit at all (Amaya, LeClere, Fiorio, & English, 2014). Including a large drop point that is not a housing unit or nearly impossible to access may throw off sampling and/or response rate assumptions and require a second sample draw.

Figure 1.   Drop Points by Number of Drop Units They Contain
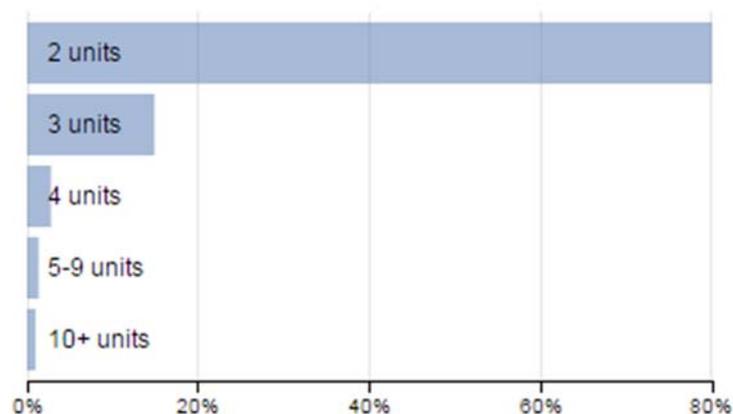
Sampling all drop units within a drop point also introduces challenges when the count of drop units identified by the interviewer does not match the count on the frame. Although the frame count of units per point has been demonstrated to be highly accurate, it is not without error (Kalton, Kali, & Sigman, 2014). To correct the frame errors, interviewers would typically be asked to enumerate the drop units at the drop point. They may perform a traditional listing in which they are not provided with the unit count. However, some drop points are very difficult to enumerate and interviewers may have a hard time correctly identifying the number of units at a drop point, undercounting and introducing undercoverage (Fiorio & Fu, 2012). Alternatively, a dependent listing may be used in which the interviewer is provided with the unit count and asked to confirm or update it. This approach introduces confirmation bias—interviewers confirm the frame count even though it is incorrect (Eckman & Kreuter, 2011). This results in both undercoverage (failure to add units not counted on the frame) and overcoverage (failure to delete duplicates on the frame).

Instead of sampling all units, one may opt to draw a subsample of units within a drop point. Although Kalton and his colleagues (2014) recommend this approach for buildings with four or more units, subsampling could be used regardless of drop point size. Coverage will be imperfect for the reasons discussed above. Units may be superficially numbered 1 through *X*. Central office may specify which units to sample and provide special instructions to the interviewer on how to identify the appropriate unit. Unfortunately, research on in-field sample selection suggests that interviewers have difficulty implementing these procedures either because of their complexity or the interviewer's carelessness (Eckman & O'Muircheartaigh, 2011). Interviewers may opt to bypass the correct procedure and select a unit in which someone is there. This results in self-selection bias and unknown selection probabilities. Additional training and monitoring protocols would be necessary to ensure correct implementation.

2.2   Mail

Mail surveys are difficult to administer to drop units because there is no way to ensure that the survey is delivered to the selected unit. Most survey mailings are addressed the "Resident of" a given address. Without a name or unit number, the researcher has no mechanism to ensure that the mailing is received by the occupants of the sampled unit. This may result in self-selection bias as the person who opens the survey may be different from the others at the drop point. Or, it is possible no one picks up the mailing since it is not addressed to her/his unit, resulting in lower response rates.

To overcome these challenges, one may once again opt to sample the entire drop point or select a subsample of units within the selected drop point. If the researcher opts to select all drop units within a drop point, he/she would identify the number of drop units based on the frame count and mail that many questionnaires to the address. Like with face-to-face

surveys, complications can occur if the number of units reported on the CDS is not correct. Mail response rates may also suffer even if the unit count is correct. Occupants who see that all units received the same package may be more likely to perceive it as a mass mailing and throw it out without opening it. This hypothesis is entirely untested, but data do show lower response rates for drop units (Link, Battaglia, Frankel, Osborn, Mokdad, 2008).

Subsampling is even harder to implement because it necessarily leaves building occupants to determine who should participate and may result in self-selection bias. It may also allow respondents to "pass-the-buck" and lead to lower response rates at these addresses (Link et al., 2008).

2.3   Telephone

Telephone surveys that use the CDS suffer from challenges different from field and mail surveys. To conduct a telephone survey using an address-based sample, the address has to be reverse-matched to a telephone number. Overall address-to-telephone matches and accurate matches to units in multi-unit buildings are significantly lower than single-unit addresses (Amaya & Skalland, 2010). Without a unit number, the match rate is likely near zero. As a result, most drop points will be finalized as unresolved (i.e., occupancy and eligibility are unknown) and nonresponsive units since no contact attempt can be made.

When phone number matches are made, they should be screened to ensure that the housing unit reached is the one intended. Since drop units are not unique addresses, mismatches are likely to be quite high. Mismatches and nonmatches ultimately lower the response rate and increase the potential for nonresponse bias.

3. APPENDING AUXILIARY DATA

For all modes, it may also be possible to append data to the frame. Some vendors may be able to match some drop units to names. For field surveys, this may aid the interviewer in identifying the correct unit (e.g., perhaps the name is on the doorbell). Name may also be used to target the mailing to a given unit, or it may be used to enhance address-to-telephone matching. However, it is impossible to guarantee that the names of matched individuals live in separate drop units. It is also impossible to ensure all units in a building are matched to a name. For field surveys, this is of little concern since the name is only a tool for interviewers, not critical for making contact. For other modes, if multiple names are matched to one unit and no names are matched to another, selection probabilities are altered and become unknown.

Unit numbers can be appended to 11.1% of drop units using the NoStat file.[6] In most cases, unit numbers are available for some units within a drop point, but not for all. The NoStat file can only be used to append numbers to all drop units in 0.5% of drop points. Drop points for which all unit numbers can be appended may be treated like any other multi-unit building. Kalton and his colleagues (2014) recommend that unit numbers be ignored for partially matched buildings as their presence may cause more logistical challenges. Finally, researchers have attempted to assign unit numbers to drop units using the neighboring addresses' numbering schema. If successful, field and mail efforts could be targeted to the sampled unit. Unfortunately, all of the assignment algorithms tested to date are flawed and have not provided accurate unit assignments (Amaya, Dekker, & LeClere, 2013).

4. THE EXCLUSION ALTERNATIVE

Given all of the above challenges and the fact that drop points and units account for a small proportion of the frame, one may argue for excluding these addresses altogether. The loss in overall coverage would be small (1.5%) and the logistical challenge of collecting data from these units would be avoided. Unfortunately, drop points/units are not evenly dispersed throughout the country (Figure 2). Drop points are clustered in older urban areas such as New York, Chicago, Boston, and Philadelphia.

Figure 2.   Drop Points by County

---
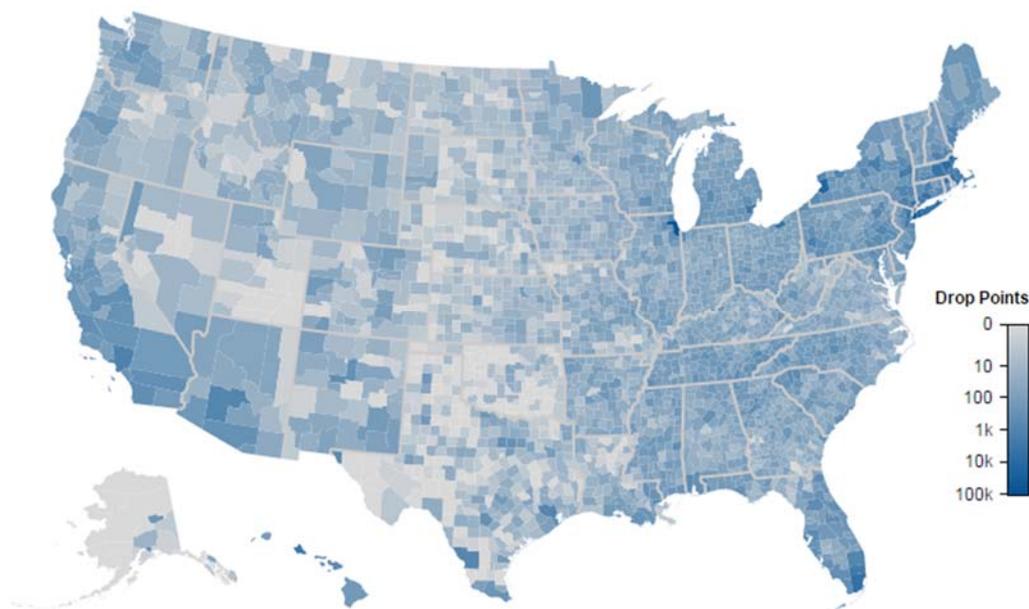
[6] NoStat is a U.S. Postal Service (USPS) database of supplemental data not included on the CDS. It includes rural vacant addresses, new growth (e.g., unfinished, new construction), and unit numbers for some drop units.

Whereas only 0.5% of all residential units are drop units in the state of Arkansas, for example, the percent rises as high as 27.0% in Queens County, New York (Table 1). A local survey with few drop units in the geography may be able to exclude drop units with little consequence, but their exclusion from other geographies could result in significant undercoverage.

Table 1.   Saturation of Drop Units in a Selection of Geographies

Area	Residential Addresses (N)	Drop Units (N)	Drop Units/Resid. Add. (%)
Arkansas State	1,514,376	7,385	0.5
New York State	8,611,306	799,880	9.3
New York City	3,560,697	638,162	17.9
Queens County	868,187	234,044	27.0

Drop units are further clustered within cities. As the proportion of drop units increases in a census community, the average age of the buildings, proportion of households with children, and the proportion of African-Americans and Latinos also increases, while the median income and owner occupancy rate decrease (Clark & Moul, 2003; Dekker, Amaya, LeClere, & English, 2012). Exclusion of all drop point units would likely bias estimates correlated with these demographic variables.

5. RECOMMENDATIONS

Different protocols may be appropriate for different types of surveys. We have attempted to list the most likely survey environments here and provide recommendations based on our knowledge as of January 2016.

- For small area surveys in which the drop units appear to be a relatively low proportion of the sample frame and do not appear to be clustered, they may be excluded from the frame with little risk of undercoverage or coverage bias. For all other surveys, drop units should be included on the frame.

- In all cases, we recommend appending unit numbers from the NoStat file (and other vendor-supplied sources, if available).

- To minimize self-selection bias, it may be appropriate to sample the entire drop point when the number of units is small (e.g., two to three units) but subsample medium and large drop points.

- After a sample has been drawn, it may be worthwhile to investigate exceptionally large drop points (via Google Maps, a phone call to the main office if one can be located, or an in-person visit), since these are less likely to be housing units. Large

drop points are rare and few (if any) are likely to be selected. So, this approach may be useful to minimize complications during data collection.

▪ Finally, an address-based sample using the CDS is not appropriate for all areas. Given the large proportion of drop points/units in New York state and the New York City metropolitan area (including Newark and Jersey City), we would advise using an alternative frame or enhanced listing procedure. To a lesser extent, the CDS may also be inappropriate in Philadelphia, Chicago, and Boston.

6. WORKS CITED

Amaya, A., & Skalland, B. (2010). What's in a Match? *Survey Practice*.

Amaya, A., Dekker, K., & LeClere, F. (2013). Using Imputation Procedures to Enhance the DSF Frame. *Presented at the Joint Statistical Meeting Annual Conference.*

Amaya, A., LeClere, F., Fiorio, L., & English, N. (2014). Improving the Utility of the DSF Address-based Frame through Ancillary Information. *Field Methods*, 70–86.

American Association for Public Opinion Research. (2016). *Address-based Sampling Task Force Report.*

Clark, J. R., & Moul, D. (2003). *Topic Report Series, No. 10: Coverage Improvement in Census 2000 Enumeration.* Bureau of the Census.

Dekker, K., Amaya, A., LeClere, F., & English, N. (2012). Unpacking the DSF in an Attempt to Better Reach the Drop Point Population. *Proceedings of the Joint Statistical Meeting, Section on Survey Research Methods*, (pp. 4596–604).

Eckman, S., & Koch, A. (2016). Are High Response Rates Good for Data Quality? Evidence from the European Social Survey. *Unpublished manuscript, currently under review*.

Eckman, S., & Kreuter, F. (2011). Confirmation Bias in Housing Unit Listing. *Public Opinion Quarterly*, 139–50.

Eckman, S., & O'Muircheartaigh, C. (2011). Performance of the Half-open Interval Missed Housing Unit Procedure. *Survey Research Methods*, 125–31.

Fiorio, L., & Fu, J. (2012). Modeling Coverage Error in Address Lists due to Geocoding Error: The Impact of Survey Operations and Sampling. *Proceedings of the American Association of Public Opinion Research Annual Conference*, (pp. 5588–96).

Kalton, G., Kali, J., & Sigman, R. (2014). Handling Frame Problems when Address-based Sampling is Used for In-person Household Surveys. *Journal of Survey Statistics and Methodology*, 283–304.

Link, M. W., Battaglia, M. P., Frankel, M. R., Osborn, L., & Mokdad, A. H. (2008). A Comparison of Address-based Sampling (ABS) versus Random-digit Dialing (RDD) for General Population Surveys. *Public Opinion Quarterly*, 6–27.