Marketing Systems Group

GENESYS SAMPLING

PROS AND CONS OF ADDRESS BASED SAMPLING

Overview

In recent years, survey and market researchers have been reconsidering address-based sampling (ABS) methodologies to reach the general public for data collection and related commercial applications. Essentially, there are three main factors for this change:

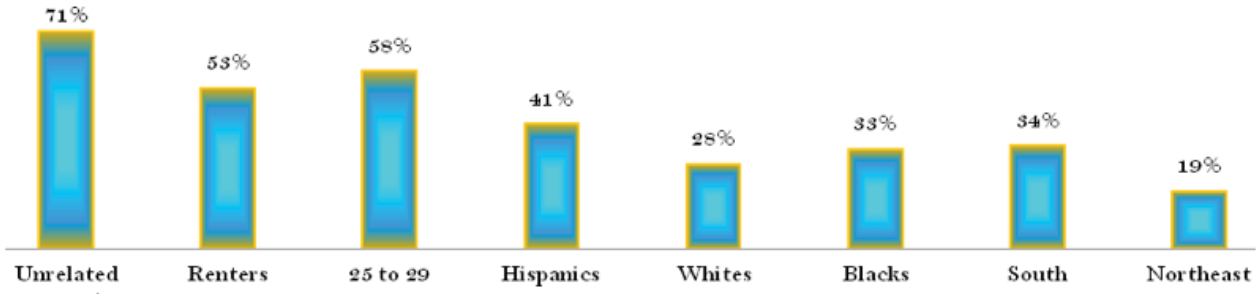
- Evolving challenges of telephone-based sampling methods, particularly those relying on random digit dialing (RDD);
- Eroding rates of response to single modes of contact along with the increasing costs of remedial measures to counter nonresponse; and on the other hand
- Recent improvements in the databases of household addresses available to researchers.

This note provides an assessment of the above factors, evaluates pros and cons of ABS as an alternative, and discusses specific enhancements that can establish this emerging methodology as a practical solution for many survey and market research applications. In particular, such enhancements include amelioration of some of the known problems associated with ABS frames through augmentations with geodemographic and other supplementary data items. While enabling researchers to develop more efficient sampling designs, such enhancements broaden their analytical possibilities by providing an expanded set of covariates for nonresponse bias analysis and weighting, as well as hypothesis testing and statistical modeling tasks.

Challenges of RDD-Based Surveys

For nearly two decades, the traditional sampling methodology of list-assisted landline RDD has served as the survey research workhorse for population-based studies. In recent years, however, virtually all RDD surveys rely on dual-frame techniques in an attempt to improve coverage. Primarily, this change is due to the growing number of households that are abandoning their landline phones in favor of cellular services. The following figure shows the geodemographic composition of adults living in households without wireline services, and hence the potential coverage bias that can result should such individuals be excluded from sample surveys. Consequently, the dual-frame RDD technique has become the standard practice whereby landline telephone numbers are supplemented with cellular numbers to produce representative samples of households – including those dubbed cellphone-only (CPO).

Geodemographic composition of adults living in CPO households (Source: CDC 2011)



While a dual-frame RDD methodology can offer an effective remedy for improving coverage of the traditional landline samples, the current practice of this alternative is subject to technical and operational inefficiencies. On the one hand, the unavailability of current estimates for the number of CPO households creates inconsistencies for sample selection and weighting applications. On the other hand, even with the addition of cellular telephone numbers there are other households that remain

unreachable via the current dual-frame RDD methodology. This includes landline households whose telephone numbers are assigned outside of the traditional landline RDD frame. It is worth noting that MSG has developed innovative solutions to address these challenges, including producing quarterly estimates for CPO households at the county level, as well as creating a new landline frame that can capture virtually all existing assigned landline numbers. For more information about these solutions you can download a copy of our whitepaper on *Improvements for Dual-Frame RDD Sampling and Weighting Methodologies*.

Eroding Rates of Response to Single Modes of Contact

Biener et al. (2004) and Curtin et al. (2005) point out that the rate of response to telephone surveys has been on a decline. More recent investigations by Fahimi et al. (2007a) reveals that response rates to well-funded government surveys follow this trend as well. For instance, BRFSS that is the world's largest RDD survey has suffered a drop of nearly 20 percentage points in response rates during the course of the past decade. Given that nonresponse is highly differential and can vary significantly across different geodemographic subgroups, it is of a great concern when a larger percentage of sample households opt not to respond to a survey. Even when sophisticated nonresponse adjustment procedures are employed to reduce bias, this gain is always exercised at the expense of diminished precision of survey estimates, since weighting inflates variance of survey estimates (Fahimi et al., 2007b). Beyond statistical techniques many researchers resort to other tactics to improve response rates to surveys. As reported by Fahimi et al. (2004) the offer of incentives can significantly increase response rates, however, such gains are often achieved at a high cost as practical nonresponse conversion strategies are labor intensive and require exceedingly larger amounts of incentives to be effective.

Improvements in Databases of Household Addresses

Recent advances in database technologies along with improvements in coverage of household addresses have provided a promising alternative for surveys and other commercial applications that require contacts with representative samples of households. Obviously, each household has an address and virtually all households receive mail from the U.S. Postal Service (USPS). The Computerized Delivery Sequence File (CDSF) of the USPS is a database that contains all delivery points, with the exception of general delivery where carrier route or P.O. Box delivery is not available and mail is held at a main post office for claim by recipients.

With more than 135 million delivery points on file, the latest generation of the CDSF is the most complete address database available. As such, it is safe to assume that if an address cannot be matched against the CDSF it is most likely an undeliverable address. What is more, by providing validation services for both correctness and completeness of addresses the CDFS can significantly enhance the address hygiene. Consequently, this system helps reduce the number of undeliverable-as-addressed mailings, increase the speed of delivery, and reduce cost. Also, with daily feedback from tens of thousands of letter carriers the database is updated on a nearly continuous basis. The following table provides counts for the main groups of delivery types, a complete listing of which and their operational definitions is provided in the appendix.

Distribution of the CDSF delivery type points as of April 2012

Delivery Point Type	Count
City Style	116,022,545
P.O. Box	15,674,989
Seasonal	859,948
Educational	99,707
Vacant	3,657,683
Throwback	277,832
Drop Points	732,230
Drop Units	1,869,633
Drop Unit Augments (by MSG)	404,817
City Style Augments (by MSG)	119,946
P.O. BOX Augments (by MSG)	29,450

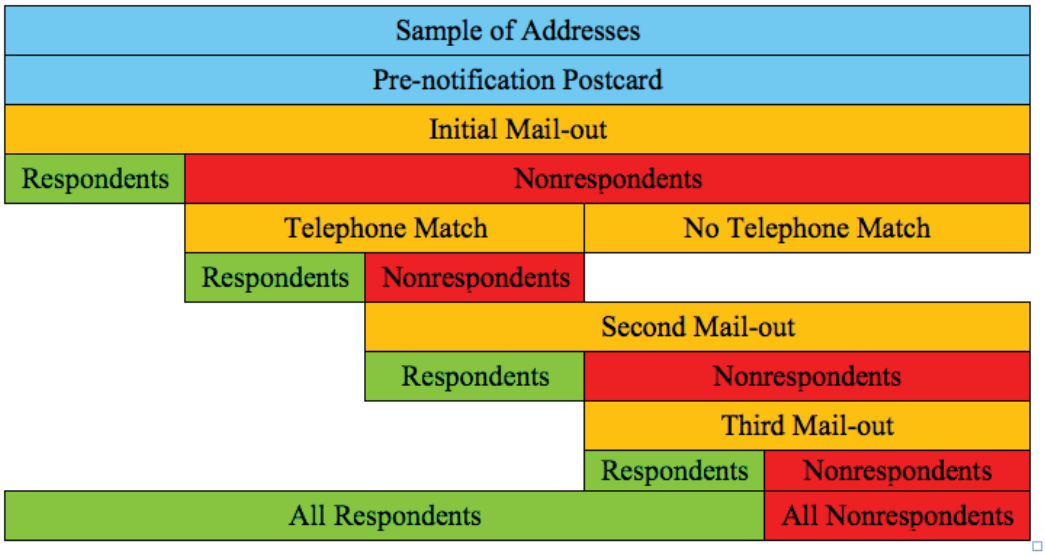
Using CDSF for Sampling Purposes

As mentioned earlier, many researchers are now considering the use of CDSF for sampling purposes to avoid some of the complications associated with the RDD alternatives. Moreover, the growing problem of nonresponse – which is not unique to any individual mode of survey administration or country (de Leeuw & de Heer 2002) – suggests that more innovative approaches will be necessary to improve survey participation. These are among the reasons why multi-mode methods for data collection

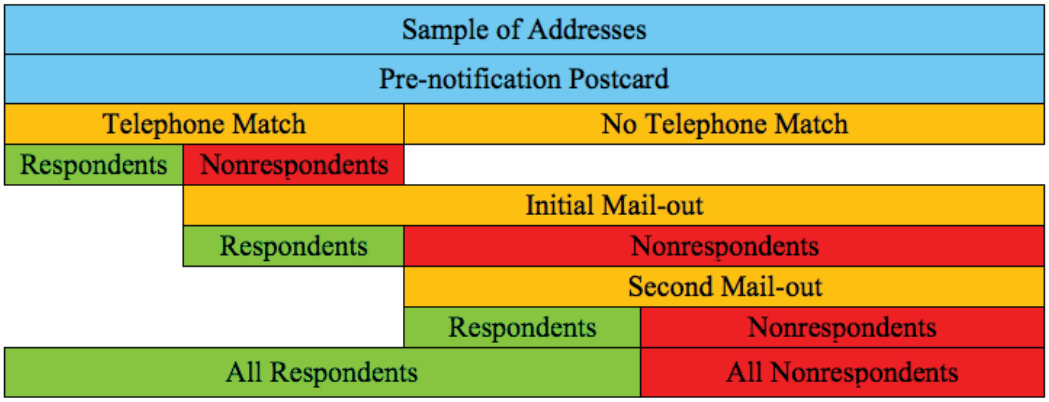
are gaining increasing popularity among survey and market researchers. It is in this context that ABS designs provide a versatile framework for creative methods of survey administration that employ multi-mode alternatives for data collection.

Considering that through reverse-matching the telephone numbers for many addresses can be obtained, different strategies for a multi-mode survey administration can be developed to accommodate the timing, budgetary, and response rate needs of a survey. One such strategy could start with the selection of a CDSF-based probability sample of households in the geographic domain of interest. This sample may be selected across the entire domain, or clustered in an area probability fashion if in-person attempts are contemplated as part of the design. Initial contacts can be by phone and/or mail and can include attempts for survey administration at the same time. Alternatively, this first contact can serve as a recruitment effort to invite potential respondents to participate in the survey using one of the many options, including web, dial-in numbers for live interviewing, an IVR system, or even a regular mail. Once the nexus of contact modes has been developed for each respondent, further contacts and reminders for survey completion can take place in any order or combination of modes that meets the project needs and is best suited for the given respondent. The following schematics display two options for multi-node survey administration under an ABS design.

First Possible Multi-mode Protocol for Survey Administration under an ABS Design



Second Possible Multi-mode Protocol for Survey Administration under an ABS Design



Cognizant of the potential implications of combining different modes of data collection, the emerging conclusions from many studies seem to suggest that different contact modalities can often be combined effectively to boost response rates (Gary 2003). In comparison to an RDD-only approach, in particular, an address-based design using multiple modes for data collection can provide response rate improvements, cost savings, as well as better coverage for households that are completely uncovered by landlines (Link 2006). In comparisons with in-person and mail-only modes of data collection, needless to say, the former is too costly to be practical for many applications while the latter requires expensive nonresponse follow-up efforts to produce creditable data (Groves 2005). What seems critical, however, is for researchers to minimize differences between survey instruments associated with each mode. Moreover, effective weight adjustment techniques might be needed post data collection to account for the observed differences in the profile of respondents to each mode.

Potential Issues When Using the CDSF for Sampling Purposes

Fundamentally, the CDSF is a database for mail delivery and not a sampling frame. As such, the raw CDSF needs refinements in several aspects before it can qualify as a credible tool for survey sampling. First and foremost, this database does not include geodemographic indicators for effective sample stratification – an issue of critical importance for complex designs. Moreover, certain households have a higher likelihood of not being included as a delivery point on the CDSF. Staab and Iannacchione (2003) estimate that approximately 97% of all US households have locatable mailing addresses, however, this prevalence may diminish with population density and in areas where home delivery of mail is not readily available. Dohrmann and Mohadjer (2006) report that when comparing lists of on-site enumerated addresses to the CDSF generated listings of households for the same geography, in rural areas the rate of mismatches can be over 23%. However, as rural area addresses go through the 9-1-1 address conversion and acquire a city-style format the coverage of CDSF-based lists in rural areas is rapidly improving. As will be discussed later, in 2004 more than 7% of all addresses were undeliverable (simplified) yet today this percentage has all but disappeared.

Beyond coverage issues, when CDSF generated samples are used in surveys with a multi-mode approach for data collection one has to be prepared to address concerns about mode effects. While somewhat academic in nature, concerns have been raised about systematic differences that can be observed when collecting similar data using different modes (Dillman 1996). On the one hand, several studies have shown a greater likelihood for respondents to give socially desirable responses to sensitive questions in interviewer-administered surveys than in self-administered surveys (Aquilino 1994). On the other hand, the rate of missing data is often significantly higher in self-administered (mail or web) surveys as compared to interviewer-administered (telephone or in-person) surveys (Biemer et al., 2003). While roots of differences in data quality and response rates between various modes of data collection deserve further investigations, it can be argued that certain shortfalls of one method can be mitigated when other methods of data collection are made available to the respondents as well. Ultimately, however, it might be impossible to untangle the immeasurable interactions between the mode, the interviewer, the respondent, and the survey content (Voogt & Saris 2005).

Lastly, there are survey situations where data are to be collected in-person. In such cases reliance on delivery information may not be adequate as the exact location of all sample dwellings must be known. This is of particular complication when a P.O. Box is the only means of delivery for a household. Also, there are households that have both residential addresses as well as P.O. Boxes. Ignoring this problem leads to frame multiplicity, since such households will have multiple chances of selection. These are additional refinements that may be added to the CDSF before it can evolve from a delivery database into a sampling frame.

Available Enhancements for the CDSF

As mentioned above, one of the attractions of the ABS methodology is that it can be used to select probability-based samples of addresses in finely defined areas down to ZIP+4. Unfortunately, since there is not a one-to-one correspondence between the USPS and Census geographic definitions, this creates a problem as in most surveys the Census geographic definitions are used for sampling. However, this gap can be bridged by geo-coding each address to a unique Census block. While accommodating the geographic needs of sample surveys, this enhancement also allows appendage of many ancillary data items to each address, including those available from the Census and commercial sources. This is the crossroad where basic list suppliers – those that simply offer raw extracts from the USPS – are differentiated from Marketing Systems Group that provides enhanced version of the CDSF by supplementing this basic delivery database with a variety of data items researchers need to develop efficient sampling designs. These include:

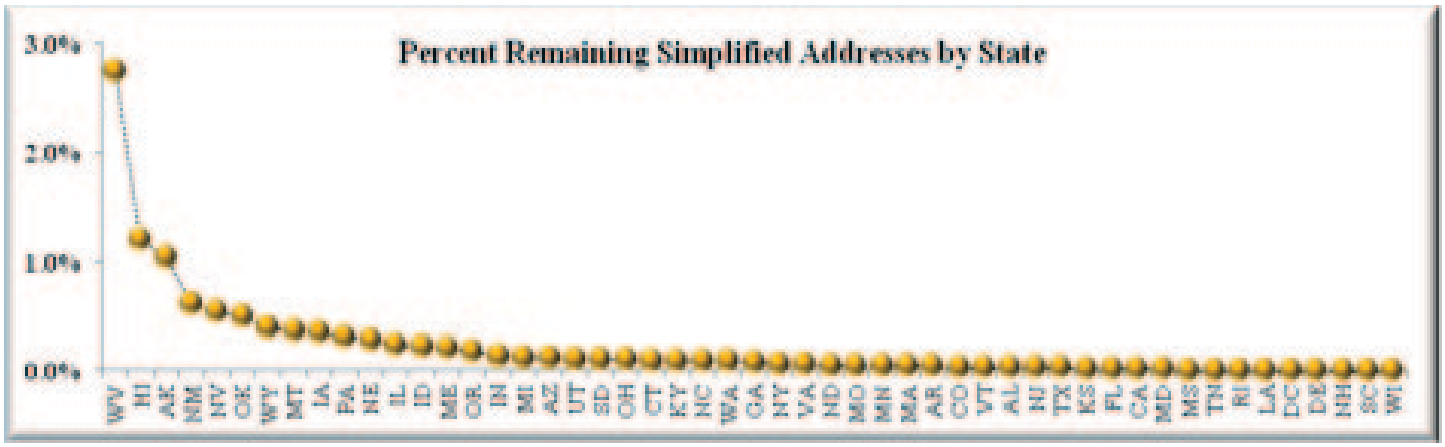
- Detailed geodemographic information;
- Name and telephone number;
- Simplified address resolutions;
- Indicators for areas with potential coverage problem; and
- Frame multiplicity reduction.

Detailed Geodemographic Information – While the CDSF can provide basic delivery details about an address, oftentimes, researchers require detailed geodemographic data for sample design and allocation. Such needs can be accommodated by appending Census-based localities and population figures at the available geographic levels. Moreover, by accessing several commercial databases that contain various data items for households it is possible to enhance the CDSF for targeted sampling applications. While many of such data items correspond to individual households, there are also modeled characteristics that are available at different levels of aggregation. Starting from the ZIP+4 level, which typically consists of only a handful

of households, the resulting information can then be rolled up to higher levels. This includes all Census geographic domains (Block, Block Group, Tract, County, MSA, State, and Region); marketing geographic domains (Media Markets, ZIP Areas, etc.); as well as custom areas (Retail Trading Areas and specific geographies based on distance or radius).

Name and Telephone Number Retrieval – Customizing the initial mailings to sample households is known to improve response rates and reduce cost. Given the plethora of junk-mail that households receive on daily basis where the packets typically carry generic contact information, research suggests that the rate of response can increase when the name of survey recipients appear on the mailed material (Dillman 1991). Moreover, with multi-mode survey applications one can reduce nonresponse through follow-up phone calls to those not responding to the initial mail-out. Taking advantage of multiple databases, MSG makes it possible to retrieve names and telephone numbers associated with many of the CDSF addresses. Done correctly, more than 90% of addresses can be name-matched and about 55% can be linked to a landline telephone number, although match rates can vary depending on the geography.

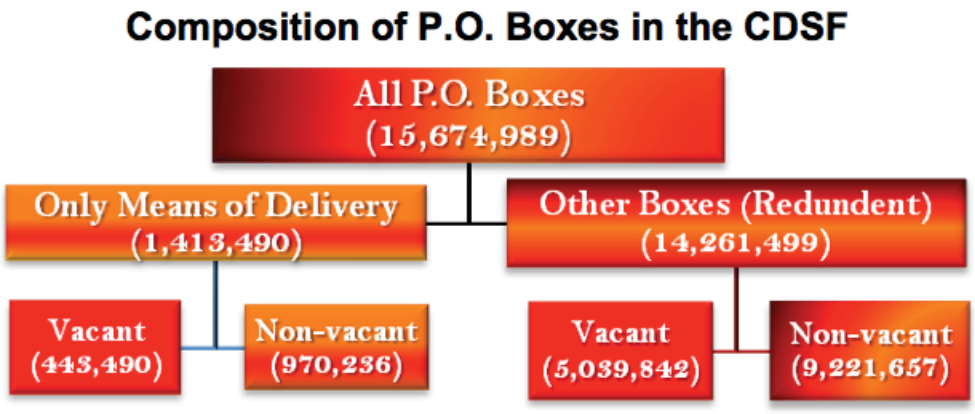
Simplified Address Resolutions – Since the CDSF only provides counts of undeliverable (simplified) addresses that are void of street numbers or other pertinent delivery information, resolution of such cases provides an important enhancement for sampling purposes. While the number of such addresses is rapidly decreasing as they go through the 9-1-1 address conversion, there are still a number of simplified addresses in the CDSF. As seen from the following chart, the distribution of simplified addresses varies across states with West Virginia topping the rank with about 3% of its addresses considered to be simplified. Again, by accessing several large databases that contain different information for households MSG obtains the missing information for virtually all simplified addresses. Subsequent to this resolution, all other informational data that exist for addressed households become available for sample design and data collection purposes.



Indicators for Areas With Potential Coverage Problem – There are delivery points that are reachable only via P.O. Boxes. Also, in certain areas there are newly constructed dwellings that are currently not registered with the Postal Service. In such cases the physical location of the corresponding households maybe unknown and not included in the CDSF. Yet, there are surveys for which visit to sample households is part of the data collection protocol for in-person interviewing or gathering of physical measurements. For such instances contacting households by mail or telephone (if obtainable) may not be a viable alternative for survey administration – although, mail or telephone can always be used to recruit such households and obtain their residential addresses. When the physical location of a household is unviable, it might become necessary to resort back to the traditional method of onsite enumeration.

Given the significant cost of such endeavor, however, researchers have developed creative options for assessing the need for onsite enumerations so that only those areas poorly covered by the CDSF may require onsite enumeration. For example, using regression models based on specific characteristics of area segments it is possible to predict the quality of the CDSF coverage (McMichael et al. 2010 and Montaquila et al. 2010). Relying on various commercial and public databases MSG can provide the needed covariates for use in such models.

Frame Multiplicity Reduction – The CDSF includes about 15.7 million addresses that are in the form of a P.O. Box, of which about 1.4 million are the only way of getting mail (OWGM). In all likelihood, the majority (if not all) regular boxes correspond to households that are represented in the CDSF via regular residential addresses as well. Moreover, a large percentage of boxes (regular or OWGM) are currently vacant. Consequently, by eliminating vacant OWGM and all regular P.O. Boxes it is possible to remove virtually all duplicate listings in the CDSF. Before selection of samples, MSG provides the counts of these and all other delivery types so that researchers can determine the exact composition of the sampling frame for their surveys.



Concluding Remarks

All single-mode methods of data collection are subject to growing rates of coverage and participation difficulties. Surveys that rely on telephone for data collection, in particular RDD-based surveys, suffer from pronounced coverage problems unless supplemented with cellular numbers. In-person surveys are typically too costly to be practical as the only mode of data collection in many instances, and mail surveys alone are often too slow and secure too low of a response rate to produce reliable results. It is against this background that multi-mode methods of data collection are gaining popularity as alternatives that can reduce many of the problems associated with single-mode methods. In this regard, address-based sampling provides a convenient framework for development of effective sampling designs and creative protocols for implementation of surveys that employ multi-mode alternatives for data collection.

The Computerized Delivery Sequence File of the USPS can provide a powerful tool for sample surveys, however, in its raw form the CDSF is simply a database for delivery of mail. It is only through proper enhancements that the CDSF can evolve into an effective sampling frame for selection of probability-based samples with surgical precisions. Enhancements provided by MSG aim to achieve this critical objective by significantly improving the coverage of the CDSF and expand its utility for complex sampling designs and analytical applications.

References

1. Aquilino, W. (1994). Interview mode effects in surveys of drug and alcohol use: a field experiment. Public Opinion Quarterly, 58, 210-40.
2. Barron, M., J. Kelly, R. Montgomery, J. Singleton, H. Shin, B. Skalland, X. Tao, and K. Wolter. (2010). More on the Extent of Undercoverage in RDD Telephone Surveys Due to the Omission of 0-Banks. Survey Practice, April 2010.
3. Biener, L., Garrett, C.A., Gilpin, E.A., Roman, A.M., & Currivan, D.B. (2004). Consequences of declining survey response rates for smoking prevalence estimates. American Journal of Preventive Medicine, 27(3), 254-257.
4. Biemer, P.P. & Lyberg, L.E. (2003). Introduction to Survey Quality, New York: John Wiley & Sons, Inc.
5. Blumberg, S. and Luke, V. (2007). "Early Release of Estimates from the National Health Interview Survey."
6. Brick, J. M., J. Waksberg, D. Kulp, and A. Starer. 1995. "Bias in List-Assisted Telephone Samples." Public Opinion Quarterly, 59: 218-235.
7. Curtin, R., Presser, S., & Singer, E. (2005). Changes in telephone survey nonresponse over the past quarter century. Public Opinion Quarterly, 69, 87-98.
8. de Leeuw, E. & de Heer, W. (2002). Trends in household survey nonresponse: a longitudinal and international comparison. In R. M. Groves, D. A. Dillman, J. L. Eltinge (Eds.), Survey Nonresponse (pp. 41-54). New York: John Wiley & Sons, Inc.
9. Dillman, D. A. 1991. The Design and Administration of Mail Surveys, Annual Review of Sociology, 17, 225-249.
10. Dillman, D., Sangster, R., Tanari, J., & Rockwood, T. (1996). Understanding differences in people's answers to telephone and mail surveys. In Braverman, M.T. & Slater J.K. (eds.), New Directions for Evaluation Series: Advances in Survey Research. San Francisco: Jossey-Bass.
11. Dohrmann, S., Han, D. & Mohadjer, L. (2006). Residential Address Lists vs. Traditional Listing: Enumerating Households and Group Quarters. Proceedings of the American Statistical Association, Survey Methodology Section, Seattle, WA. pp. 2959- 2964.
12. Groves, R.M. (2005). Survey Errors and Survey Costs, New York: John Wiley & Sons, Inc.
13. Fahimi, M., M. W. Link, D. Schwartz, P. Levy & A. Mokdad (2008). "Tracking Chronic Disease and Risk Behavior Prevalence as Survey Participation Declines: Statistics from the Behavioral Risk Factor Surveillance System and Other National Surveys." Preventing Chronic Disease (PCD), Vol. 5: No. 3.
14. Fahimi, M., D. Creel, P. Siegel, M. Westlake, R. Johnson, & J. Chromy (2007b). "Optimal Number of Replicates for Variance Estimation." Third International Conference on Establishment Surveys (ICES-III), Montreal, Canada.
15. Fahimi, M., Chromy J., Whitmore W., & Cahalan M. Efficacy of Incentives in Increasing Response Rates. (2004). Proceedings of the Sixth International Conference on Social Science Methodology. Amsterdam, Netherlands.
16. Fahimi, M., Kulp, D. & Brick, J. M. (2009). A Reassessment of List-Assisted RDD Methodology. Public Opinion Quarterly, Vol. 73, No. 4, pp. 751–760.
17. Gary, S. (2003). Is it Safe to Combine Methodologies in Survey Research? MORI Research Technical Report.
18. Iannacchione, V., Staab, J., & Redden, D. (2003). Evaluating the use of residential mailing addresses in a metropolitan household survey. Public Opinion Quarterly, 76:202-210.
19. Link, M., M. Battaglia, M. Frankel, L. Osborn, & A. Mokdad. (2006). Addressed-based versus Random-Digit-Dial Surveys: Comparison of Key Health and Risk Indicators. American Journal of Epidemiology, 164, 1019 - 1025.
20. McMichael, J., V. Iannacchione, B. Shook-Sa, J. Ridenhour, K. Morton, and J. Chromy (2010). Predicting the Coverage of Address-Based Sampling Frames Prior to Sample Selection. Joint Statistical Meetings, Vancouver Canada.
21. Montaquila, J., S. Shore, and V. Hsu (2010). An Investigation of the Presence or Absence of Households at Addresses Obtained by Field Listing and from USPS Lists. Joint Statistical Meetings, Vancouver Canada.
22. O'Muircheartaigh, C., S. Eckman, and C. Weiss (2003). Traditional and enhanced field listing for probability sampling. Proceedings of the American Statistical Association, (CD-ROM), Alexandria, VA, pp.2563- 2567.
23. Staab, J.M., & Iannacchione, V.G. (2004). Evaluating the use of residential mailing addresses in a national household survey. Proceedings of the American Statistical Association, Survey Methodology Section (CD-ROM), Alexandria, VA, pp.4028- 4033.
24. Tucker, C. Lepkowski, J. M. (2008). Advances in Telephone Survey Methodology. New York: John Wiley & Sons, Inc.
25. Voogt, R. & Saris, W. (2005). Mixed mode designs: finding the balance between nonresponse bias and mode effects. Journal of Official Statistics. 21, 367-387.
26. Weiss, A., M. Battaglia, J. Boyle, A. Hyon, and D. Kulp. 2009. "Unlisted Banks in New York City: Coverage Error and Bias in Urban Areas from RDD Samples Based on Hundreds Banks with Listed Numbers." Presented at the American Association for Public Opinion Research, Hollywood, FL.
27. Wilson, C., Wright, D., Barton, T. & Guerino, P. (2005). "Data Quality Issues in a Multi-mode Survey" Paper presented at the Annual Meeting of the American Association for Public Opinion Research, Miami, FL.

APPENDIX
Data Element Definitions for the CDSF

Address Vacant Indicator: A field that specifies whether a delivery point has been unoccupied for 90 days or more:

1) Y = Vacant 90 days or more
2) N = Not vacant

Business Delivery Active Count: The total number of active business deliveries in the CDS ZIP Code Header Record. This field contains:

1) The total number of active business deliveries for a given ZIP Code
2) The total number of active business deliveries for a carrier route (in the CDS ZIP Route Header Record)

Business Delivery Possible Count: The number of possible deliveries in the CDS ZIP Code Header Record.

1) The field contains the number of possible business deliveries for a given ZIP Code
2) The number of possible business deliveries for a given carrier route

Carrier Route ID: A 4-byte code assigned to a given mail delivery or collection route within a 5-digit ZIP Code area. The first character of this identification is alphabetic; the last three are numeric:

1) Bnnn = PO Box
2) Hnnn = Highway Contract
3) Rnnn = Rural Route
4) Cnnn = City Delivery
5) Gnnn = General Delivery

Delivery Address Number: The component of an address preceding the street name – often referred to as the house number. Delivery address numbers that are preceded by significant leading zeroes are identified by a hyphen preceding the address number. All numeric data is right-aligned with leading zeroes, and all alphanumeric data is left-aligned.

Delivery Point Barcode (DPBC) Digits: The last two digits of a primary street number (post office box number, rural route box number, or highway contract route number). Delivery point codes for multi-delivery points are calculated using the secondary number.

Delivery Point Barcode Check Digit: The last number in a barcode; it is calculated by the following formula: 10 – last digit of the sum of (ZIP5+ZIP4+DPBC). As an example, if ZIP5 = 94497, ZIP4 = 9200, and DPBC = 00, then, 9+4+4+9+7+9+2+0+0+0+0 = 44 10 – 4 = 6.

Delivery Point Business Family Served Count: The number of potential deliveries for a drop site. If the Delivery Point Drop Indicator equals Y, the Business Family Served Count field contains the number of businesses or families served at that drop site. This number is greater than one (except for general delivery records, in which it can equal zero.) If the Delivery Point Drop Indicator equals N, the Business Family Served Count equals zero. If the Delivery Point Drop Indicator equals C, the Business Family Served count is greater than zero.

DPBC Check Digit = 6: If the last digit of the sum of ZIP5 + ZIP4 + DPBC is 0 (zero), the check digit will be 0 (zero).
Delivery Point Drop Indicator: A 1-byte field that specifies whether a delivery point is a drop site whereby the carrier delivers the mail to one point from which the company distributes it to specific boxes. A drop is a single delivery point or receptacle that services multiple residences, such as a single door slot shared by two residences, a box on a wall for duplexes, or a boarding house or fraternity in which mail is delivered to the door for subsequent distribution. Mail for drop sites is distributed internally by the site. A commercial mail receiving agency (CMRA) holds mail or forwards it to an addressee. Each CMRA must be registered with the post office responsible for delivery to the CMRA.

1) Y = Drop point
2) N = Not drop point
3) C = CMRA

Delivery Point Record Count: The total number of delivery points for a given area. In the CDS ZIP Code Header Record, this field contains the total number of delivery point records for a given ZIP Code; in the CDS Carrier Route Header Record, it contains the total number of delivery point records for a given carrier route.

Delivery Sequence Number: A numeric code indicating the position of an address within a carrier's walk path. The order in which a carrier delivers mail can be determined by listing all address records for the carrier sorted by the delivery sequence number.

Delivery Type Code: A numeric sequence indicating the category of delivery point and its type of service:

A. Residential Curb: A delivery point for a residence with a mail receptacle located at the curb.
B. Residential Cluster Box Unit (CBU): A delivery point consisting of residential cluster boxes.
C. Residential Central: A delivery point within a building that has two or more ZIP+4 codes assigned to a bank of boxes (e.g., Arrow locks).
D. Residential Other: A delivery point serviced other than by curb, central, or NDCBU. Examples of this delivery type include door-to-door (walking route) or door-slot delivery.
E. Residential Facility Box: A residential customer's P.O. Box located in a USPS facility.
F. Residential Contract Box: A residential customer's P.O. Box located in a contract unit.
G. Residential Detached Box: A residential customer's P.O. Box that is not located in a post office building but for one which the USPS collects box rent.
H. Residential Non-Personnel Unit (NPU): A residential self-service postal center that furnishes essential mail services such as collection and delivery of ordinary mail and sale of stamps. At a residential NPU, mail is delivered to mail boxes (similar to PO boxes). Residential NPU deliveries are non-staffed, self-service and are most often found in rural areas.
I. Business Curb: A delivery point for a business with a mail receptacle located at the curb.
J. Business CBU: A delivery point consisting of business cluster boxes.
K. Business Central: A delivery point for a business serviced by receptacles that are within a delivery center or mail-room.
L. Business Other: A delivery point serviced other than by curb, central, or NDCBU. Examples of this delivery type include door-to-door (walking route) or door-slot delivery.
M.Business Facility Box: A business's PO Box whose PO Box section is located at a USPS facility.
Business Contract Box: A business's PO Box whose PO Box section is located at a USPS contract unit office building, but which is rented from the USPS. (Business Detached Box: A business's PO Box whose detached box section is not located in a post)
N. Business Non-Personnel Unit: A self-service postal center that furnishes essential mail services, such as collection and delivery of ordinary mail and sale of stamps. At a business NPU, mail is delivered to mail boxes, similar to PO Boxes. Business NPU's are non-staffed, self-service units and are most often found in rural areas.
O. General Delivery: An alternate delivery service that allows customers with proper identification to pick up mail at post offices. Provided primarily at offices without letter carrier delivery or for transients and customers who do not have a permanent address or who prefer not to use PO Boxes.

Drop Count: The total number of drop sites. In the CDS ZIP Code Header Record, this field contains the total number of drop sites for a given ZIP Code; in the CDS ZIP Route Header Record, it contains the total number of drop sites for a given carrier route. It is NOT the number of customers served by those drop sites. The number of customers served by these drops can be found in the Resident/Business Served Drop counts.

File Version Month: The month for this edition of the file.

File Version Year: The last two digits of the year for this edition of the file.

Mailer ID: The CDS customer ID number.

Plus4 (+4): A 4-digit sequence consisting of the sector and segment numbers. The +4 further subdivides the ZIP Code area.

PO Box Throwback Count: The total number of PO throwbacks for an area. In the CDS ZIP Route Header Record, the PO Throwback field contains the total number of PO throwbacks for a given carrier route and the total number of PO throwbacks for a given ZIP Code.

PO Box Throwback Indicator: The address of this delivery point is a street address, but actual delivery is made to the customer's PO Box address. T = PO Box Throwback

Preferred Last Line Locale Key: This field contains the locale key of the preferred last line of a particular delivery point and is used to locate that delivery point's preferred city name in City State Product.

Residence Business Served Drops Count: The total number of families or businesses served by a drop delivery or other multi-service delivery point for a given area. In the CDS ZIP Route Header Record, this field contains the number of families or businesses served by a drop delivery or other multi-service delivery point for a given carrier route; in the CDS ZIP Code Header Record, it contains the number of families or businesses served by a drop delivery or other multi-service delivery point for a given ZIP Code.

Record Type Code: An alphabetic value that identifies the type of data in the record.

1) P = PO Box
2) G = General
3) R = Rural Route
4) S = Street
5) F = Firm
6) H = High-rise

Residential Delivery Active Count: The total number for active residential deliveries for a given area. In the CDS ZIP Route header Record, this field contains the number of active residential deliveries for a given carrier route; in the CDS ZIP Code Header Record, it contains the number of active residential deliveries for a given ZIP Code.

Residential Delivery Possible Count: The total number of possible residential deliveries for a given area. In the CDS ZIP Route Header Record, this field contains the number of possible residential deliveries for a given carrier route; in the CDS ZIP Code Header Record, it contains the number of possible residential deliveries for a given ZIP Code.

Route Count: The total number of routes for a given area. For the CDS ZIP Header Record, this field gives the number of routes for a given ZIP Code; in the CDS Volume Header Record, it contains the number of routes for a given volume.

Seasonal Delivery Indicator: A 1-byte field that specifies whether a given address receives mail only during a specific season (e.g., a summer-only residence).

1) Y = Delivery point has seasonal delivery
2) N = Not applicable

Seasonal Count: The total number of seasonal deliveries for a given area. In the CDS ZIP Code Header Record, this field gives the number of seasonal deliveries for a given ZIP Code; in the CDS ZIP Route Header Record, it contains the number of seasonal deliveries for a given carrier route.

Secondary Address Abbreviation: A descriptive code that identifies the type of secondary address range, e.g., APT and BLDG. When these abbreviations are used, a secondary address number is not required. See the latest version of Publication 28, Postal Addressing Standards, for a complete list of secondary address abbreviations. If any discrepancies exist between this document and Publication 28, the most current version of Publication 28 always takes precedence.

Secondary Address Number: The alpha and/or numeric sequence that specifically identifies a unit at an address. The number may represent an apartment, room, suite, floor, space, or other similar addressable unit. Apartment or room numbers that are preceded by significant leading zeroes are identified by a hyphen preceding the number. All numeric data is right-aligned with leading zeroes, and all alphanumeric data is left-aligned.

Street Name: The official name assigned to a street by a local governing authority. This field contains only the street name and does not include directional or suffix components. This element may also contain literals (e.g., PO Box, General Delivery, USS, PSC, or UNIT). Numeric street names that have numeric components of four characters (or less) are aligned so the low-order digit of the number is in the fourth position of the field. This shift is made so that the numeric street names appear in numeric sequence.

Street Suffix Abbreviation: A standard USPS code for a word frequently appearing as a trailing designator in street addresses.

Street Post-Direction Abbreviation: A geographical direction following a street name, e.g., E (East) and W (West), etc.

Street Pre-Direction Abbreviation: A geographic direction preceding a street name.

Total Record Count: The total number of delivery point records for a given area.

Urbanization Locale Key: The locale key of an urbanization for a given delivery point; it is used only with addresses in Puerto Rico. This field is used to locate the urbanization name in City State Product.

Vacant Code: The total number of delivery points unoccupied 90 days or longer for an area. In the CDS ZIP Code Header Record, this field gives the number of delivery points unoccupied for 90 days or more for a given ZIP Code; in the CDS Carrier Route Header Record, the total number of delivery points unoccupied for 90 days or more for a given carrier route.

Volume Sequence Number: A number representing the position of any given volume within the entire set of tapes/cartridges.

ZIP Code: ZIP is an acronym for Zone Improvement Plan. It is a 5-digit code that identifies a specific geographic delivery area. ZIP Codes can represent an area within a state, an area that crosses state boundaries (unusual condition) or a single building or company that has a high mail volume.

ZIP Count: The total number of ZIP Codes for a given volume.

ZIP Sector Code: The first two digits of the 4-digit ZIP add-on code. This code represents a smaller geographic area within a ZIP Code area. ZIP sector boundaries are established so as not to cross state or county lines.

ZIP Segment Code: The last two digits of the 4-digit ZIP add-on code, which represent a smaller geographic area within a ZIP Sector area. Geographically, ZIP segments represent areas such as one side of a city block between intersections; both sides of a street, including cul-de-sacs; a company or building; a floor or group of floors within a building; a cluster of mailboxes; sections of post office boxes; or other similar delivery groups.