Recent Developments in Address-Based Sampling (ABS)

Mansour Fahimi
Dale Kulp and David Malarek

FedCASIC Workshop
March 17-19, 2009
Emerging Alternatives in Survey Administrations

Issues with the “Old Methods”

Need for More Flexible/Innovative Methods

Using DSF for Sampling Purposes

Potential Issues with DSF as a Sampling Frame

Possible Enhancements of DSF
Address-based sampling (ABS) methodologies are gaining popularity for several reasons:

- Evolving coverage problems associated with telephone-based samples
- Eroding rates of response to single modes of contact along with the increasing costs of remedies to reduce nonresponse
- Recent improvements in the databases of household addresses available to researchers
A growing number of households are becoming cell-only or cell-mostly

According to NCHS more than 3 out of 10 adults in the U.S. receive all or nearly all calls on cell phones

Cell-only and cell-mostly individuals have different characteristics than the general public – younger and more mobile

If these individuals are not included in surveys results can be biased
Coverage Problems for Telephone Surveys
(Composition of the Landline Frame)

Distribution of the 100-Series Telephone Number Banks
NPA-NXX-XX00 to NPA-NXX-XX99

- 0-Listed HH
- 1+ Listed HH
- Non-Genesys

Years:
- 2000
- 2004
- 2008
- 2009
In 1995 Westat and MSG estimated the percentage of households in 0-listed banks to be only 3.7%

Results from the 2008 study:

<table>
<thead>
<tr>
<th>Disposition</th>
<th>1+Listed 100-Banks (n = 9,062)</th>
<th>0-Listed 100-Banks (n = 20,000)</th>
<th>Remaining POTS (n = 8,937)</th>
<th>Total</th>
</tr>
</thead>
<tbody>
<tr>
<td>Residential</td>
<td>80.5%</td>
<td>14.5%</td>
<td>5.0%</td>
<td>100%</td>
</tr>
<tr>
<td>Business</td>
<td>35.7%</td>
<td>51.2%</td>
<td>13.1%</td>
<td>100%</td>
</tr>
<tr>
<td>Nonworking</td>
<td>23.9%</td>
<td>49.1%</td>
<td>27.0%</td>
<td>100%</td>
</tr>
</tbody>
</table>
Eroding Rates of Response to Telephone Surveys

Response Rate for the BRFSS Surveys

45\% 50\% 55\% 60\% 65\% 70\% 75\%

Need for More Flexible/Innovative Methods

- Researchers are struggling with the “old” methods of survey administration:
  - Evolving coverage problems of telephone surveys
  - Prohibitive costs of in-person surveys
  - Growing rates of nonresponse to single mode methods

- Multi-mode methods are gaining popularity because different modalities can be combined effectively to:
  - Improve coverage
  - Boost response rates
  - Reduce cost
Pros and Cons of Multi-Mode Alternatives

In comparison to single-mode methods of data collection multi-mode methods can (Link 2006, 2007, 2008):

- RDD – Improve response and coverage rates
- In-person – Reduce cost and time significantly
- Mail – Improve response rates

Addressed-based sampling methods provide a convenient framework for multi-mode alternatives
Pros and Cons of Multi-Mode Alternatives

- There are concerns about systematic differences when collecting similar data using different modes (Dillman 1996)

- There is a greater likelihood for socially desirable responses to sensitive questions in interviewer-administered surveys (Aquilino 1994)

- The rate of missing data is higher in self-administered surveys as compared to interviewer-administered surveys (Biemer 2003)
Pros and Cons of Multi-Mode Alternatives

- Is it feasible to untangle the convoluted interactions between the mode, interviewer, respondent, and survey contents (Voogt & Saris 2005)?

- Is mode effect simply a reflection of respondents’ preference or comfort level with different modes of survey administration?
  - Maybe “techie” respondents are more comfortable (prefer) a web-based method
  - Maybe older respondents prefer an interviewer
  - Maybe sensitive questions should be asked via IVR
Pros and Cons of Multi-Mode Alternatives

- Whatever a respondent’s preference might be, it is better to have them than to lose them.

- To reduce mode effect in multi-mode surveys, we need to:
  - Minimize differences in survey instruments for each mode of administration.
  - Devise effective weighting adjustments to account for differences in the profile of respondents to each mode.
The Delivery Sequence File (DSF) of the USPS is a database that contains all delivery point addresses. The first generation of DSF included over 125 million records with the following delivery features:

- Address validation and standardization
- ZIP+4 and carrier route coding
- Delivery sequence
- Detection of addresses that are potentially undeliverable
- Delivery-type code that indicates business or residential
- Seasonal delivery information
Improvements in Databases of Household Addresses

- With more than 135 million addresses the second generation of DSF is the most complete address database available.

- By providing the most current delivery information and improved *address hygiene* this system helps reduce cost and improve efficiency by:
  - Reducing the number of *undeliverable-as-addressed* mailings
  - Increasing the speed of delivery

- Given daily feedback from thousands of letter carriers, the database is updated on a nearly continuous basis.
Start with an address-based sample down to ZIP+4:
- Stratified or random across the entire domain
- 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 include attempts for:
- Survey administration at the point of initial contact
- Recruitment for participation via other modes

Once contact has been established follow-up attempts can take place in any order or combination of modes.
Available Data Items
(DSF File Layout)

- Zip
- Zip+4
- Walk Sequence Number
- Route Type
- PO Box Throwback
- House Number
- Pre Directional: NE and W
- Street Name
- Street Suffix, Ave, Blvd
- Post Directional: NE and NW
- Secondary Unit Descriptor
- Apt Number
- Delivery Type Code
- Vacant Code
- Drop Indicator
- Drop Count
- Seasonal Code
- Carrier Route
- Delivery Point
- Delivery Point Check Digit
- City Code
- State Code
- County
- Tract
- Block
- Normalized address
- City Name
- State Name
**Topology of the DSF**

*(Delivery Point Type Indicator)*

- **Business:** Indicates the delivery point is a business address.
- **Central:** The delivery point is serviced at a mail receptacle located within a centralized unit.
- **CMRA (Commercial Mail Receiving Agency):** A private business that acts as a mail-receiving agent for specific clients.
- **Curb:** The delivery point is serviced via motorized vehicle at a mail receptacle located at the curb.
- **Drop:** A delivery point or receptacle that services multiple residences such as a shared door slot or a boarding house in which mail is distributed internally by the site.
- **Educational:** Identified as an educational facility such as colleges, universities, dormitories, sorority or fraternity houses, and apartment buildings occupied primarily by students.
**Topology of the DSF**
*(Delivery Point Type Indicator)*

- **NDCBU (Neighborhood Delivery Collection Box Unit):** Services at a mail receptacle located within a cluster box
- **No-Stat:** Indicates address is not receiving delivery and is not counted as a possible delivery point for various reasons
- **Seasonal:** Receives mail only during a specific season and the months the seasonal addresses are occupied are identified
- **Throwback:** Address associated with this delivery point is a street address but the delivery is made to the customer’s PO Box address
- **Vacant:** Was active in the past, but is currently vacant (in most cases unoccupied over 90 days) and not receiving delivery
Potential Issues When Using DSF for Sampling

- Certain households have a higher likelihood of not being included as a delivery point (simplified addresses void of delivery information such as street or P.O. Box number):
  - The coverage rate diminishes with population density in areas where home delivery of mail is unavailable (Staab & Iannacchione 2003)
  - When comparing on-site enumerated addresses to those from DSF the rate of mismatches may be high in rural areas (Dohrmann & Mohadjer 2006)
  - A minor source of non-coverage is due to households that request that their addresses not be sold (O’Muircheartaigh 2003)
- Rural area addresses go through the 911 address conversion to acquire a city-style format and become un-simplified
Potential Issues When Using DSF for Sampling

(Counts of Simplified Addresses by Year)
Potential Issues When Using DSF for Sampling
(Percent Simplified Addresses by State)
Possible Enhancements of DSF

- “Raw” DSF contains very little information to be suitable for complex sample surveys
- Many list suppliers simply offer basic extracts from the DSF without any enhancements
- Possible enhancements include appendage of:
  - Simplified address resolution
  - Geographic information
  - Household demographic information
  - Name and telephone number retrievals
Possible Enhancements of DSF (Simplified Address Resolution)

- A carrier route consists of 100 to 2,500 households served by an individual mail carrier within a five-digit ZIP Code area.

- There are approximately 570,000 carrier routes in the U.S.
  - Simplified
  - Box Route
  - Rural Route
  - City Route
  - Highway Contract Route
  - General Delivery Route

- DSF provides only counts of addresses (physical or P.O. Box) in simplified routes.
Possible Enhancements of DSF
(Augmentation of DSF for Simplified Addresses)

- DSF contains all addresses in all non-simplified carrier routes (Box, Rural, and City).

- Can obtain a list of all simplified carrier routes and counts of active simplified addresses in each route.

- There are legitimate city-style addresses in simplified carrier routes available via commercial databases such as: Experian, infoUSA, and Axiom.

- Such addresses can be identified using the various databases available to MSG and added to DSF.
There are about 134 million residential addresses of all types:

- The latest DSF contains 852,723 simplified addresses
- MSG can augment about 718,121 addresses
- Augmented sampling frame covers over 99% of all residential addresses in the U.S.
Possible Enhancements of DSF
(Appending Information)

- Geographic Information Enactments:
  - Census geographic domains
  - Marketing and media domains

- Household Information Enhancements:
  - Direct household data from commercial databases
  - Molded household statistics at various levels of aggregation

- Name and Telephone Number Retrievals:
  - Append names to addresses (about 85%)
  - Retrieve telephone numbers (about 60%)
Single-mode methods of data collection are problematic for response rate, coverage, and cost reasons.

Telephone surveys based on landline RDD samples are subject to non-ignorable coverage bias.

Multi-mode methods of data collection can reduce some of the problems associated with single-mode methods.

DSF provides a natural and efficient framework for design and implementation of multi-mode surveys.

Available enhancements for the DSF can significantly improve its coverage and expand its utility for design and analytical applications.


