



# Uses of and Experiences with Address-Based Sampling

Jill M. Montaquila  
Westat and the Joint Program in  
Survey Methodology

- Motivations for using address-based sampling (ABS)
- Comments on Westat's recent experiences with ABS
- The United States Postal Service (USPS) Delivery Sequence File (DSF): Issues and considerations
- Discussion

# Motivations for Using Address-Based Sampling

- Declining random digit dial (RDD) telephone survey response rates and coverage rates (landline) from the late 1990s into this decade

*Use address-based sampling frames in place of RDD (with change in mode).*

- As coverage of these address lists improves, they may be considered as a cost-effective alternative to traditional listing in multi-stage area samples

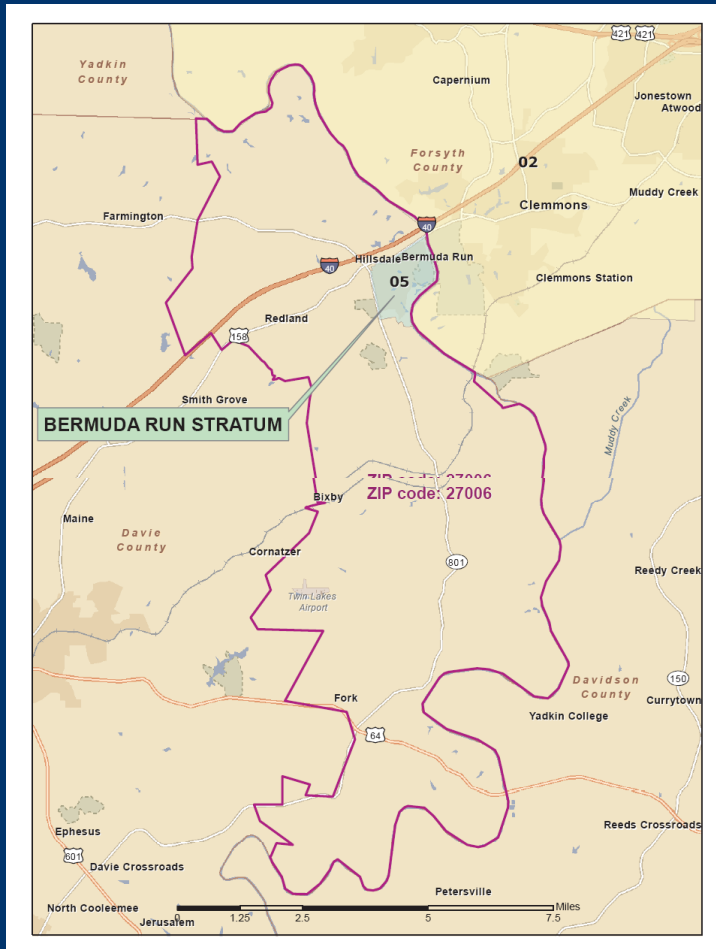
- Westat's experiences with the USPS-based address lists include:
  - Using USPS-based lists as sampling frame--in some cases, in lieu of RDD
  - Multi-stage samples:
    - As dwelling unit frame in lieu of traditional listing
    - Quality control of the traditional listings
    - To obtain counts used to update measures of size

- National Children's Study (NCS) Vanguard Study
  - USPS-based address lists used for quality control of traditional listing
- Health Information National Trends Survey (HINTS) 2007
  - USPS-based address lists used as sampling frame
  - Strictly mail
  - One-phase administration (no screening; survey all adults in household)
  - 31% response rate (vs. 24% for independent landline RDD)

- National Survey of Veterans (NSV), 2009 Pilot
  - USPS-based address lists used as sampling frame
  - Primarily mail with web option
  - Two-phase administration (screen for veterans; survey all veterans in household)

# Comments on Westat's Recent Experiences with ABS

- National Household Travel Survey (NHTS)



- National Household Education Surveys Program (NHES), 2009 Pilot Study
  - USPS-based address lists used as sampling frame
  - Nationally representative, n = 10,200 addresses
  - Primarily mail with telephone follow-up to a subsample (experiment vs. mail follow-up)
  - Two-phase (screen for children; sample one child)
  - Overall rates (with higher rate for certain conditions tested):
    - Screener response rate: 58.5% (vs. 52.8% in NHES:2007 RDD)
    - Topical (extended) response rate (Screener mail completes only): 73.9% (vs. 74-77% in NHES:2007 RDD)



In summary, Westat has used ABS:

- In place of RDD
  - HINTS 2007
  - NSV 2009
  - NHES 2009 Pilot Study
  - NHTS (Bermuda Run area)
- As dwelling unit frame in lieu of traditional listing
- For quality control of the traditional listings
  - NCS Vanguard Study
- To obtain counts used to update measures of size

# The USPS Delivery Sequence File

- USPS-based address lists are maintained by vendors (MSG, CIS, InfoUSA, etc.); quality of these lists and services provided vary
- Can be used to obtain lists of residential addresses, nationally or for restricted areas
  - State
  - ZIP code
  - Census tract (not all vendors)

- *Geocoding* is the process of attaching geospatial coordinates (latitude, longitude) to an address
- The accuracy and completeness of the geocoding process depends on
  - Engine and databases used for geocoding
  - Features of the address
- If census geography (tracts, blocks) is used to define sampling units, address lists must be geocoded

- Issues/considerations include:
  - Coverage
  - P.O. box, rural route (RR), and highway contract (HC) addresses
  - Drop point (multi-drop) addresses
  - Households with multiple addresses
  - Geocoding errors
  - Appending other information
- In some cases, these are inter-related

- Factors affecting coverage:
  - Lag time between USPS updates and vendor updates
  - Vendor “ownership” of ZIP codes
  - For scenarios requiring geocoding (e.g., local studies, samples with census geography-based sampling units):
    - Geocoding errors
    - Inability to geocode:
      - Non-city-style addresses
      - Incompleteness of street database

- May affect coverage due to geocoding issues
- In the NHES 2009 selection, about 11 percent of addresses were P.O. box or rural route
- Special considerations depending on mode:
  - Ability to locate for in-person interviews
  - Issues for special delivery service (e.g., FedEx) in mail surveys
- In NHES, P.O. boxes had significantly lower residency and response rates than city-style (about 20% lower residency and 5% lower Screener response)

## Issues/Considerations: Drop Point (Multi-Drop) Addresses

- A *drop point* is a single address that serves as a delivery point for more than one residence.
- Drop points may be flagged, and the number of drops indicated, on files provided by vendors.
- In a recent ABS selection:
  - Fewer than 1 percent of addresses were drop point addresses;
  - But the number of drops was as high as nearly 400 (with a mode of 2).
- Preliminary indication (based on NHES:2009 Pilot) that multi-drop households are less likely to respond.

- Handling of drop points:
  - Account for number of drops in determining probability of selection of the address
  - Need approach for subsampling
    - In-person interviews: Can specify approach for subsampling
    - Telephone interviews: Matching of telephone numbers may be ambiguous
    - Mail: Recipients self-select



- Households may have multiple chances of selection, e.g.
  - Households with summer/winter homes
  - Households that receive mail at both street address and P.O. box
- To accurately compute household's probability of selection, need question to ascertain means by which household receives personal mail
- With seasonal homes, could apply residency rules (implicitly or explicitly)

- Consider two scenarios:
  - A. Target population is defined by specific geographic area (e.g., Rockville, MD)
  - B. Areas serve as secondary/tertiary/etc. sampling units in a multi-stage sample
    - Consider using sampling units that do not require geocoding (e.g., ZIP codes rather than census blocks)
    - Issues with using units such as census blocks:
      - Databases used for geocoding may be incomplete/inaccurate
      - Inability to correctly geocode may affect coverage or operational efficiency of sample

- When constructing frames for a small geographic area (e.g., based on census blocks), need to decide on approach:
  - Treat any address that geocodes into the area as eligible; addresses that geocode to a location outside the area are ineligible
  - Cast a wider net and keep only those addresses that are truly located within the designated area

## Issues/Considerations: Appending Other Information

- For an ABS sample, vendors may be able to append:
  - Telephone number (our experience with national samples has been that phone number can be appended to about 60 percent of addresses)
    - Need to confirm address: In 2007 study, about 28 percent of matched cases were found to be associated with nonworking/nonresidential phone number or incorrect address.
    - Effectiveness of phone vs. mail

## Issues/Considerations: Appending Other Information

- For an ABS sample, vendors may be able to append:
  - Name
    - Mail may be undeliverable if name for mailing is mismatched
    - If mail is deliverable, named person might not be the “right” household respondent
  - Other demographics, etc.

- Factors leading to the advance of ABS:
  - Decline in RDD response rates and landline RDD coverage rates
  - Availability of USPS-based address lists through vendors
  - Improved geocoding databases
  - Roll-out of E-911 addressing
- Experience with ABS (with mail as primary mode) has proven effective as an alternative to RDD
- “The devil is in the details”
- Still much that is unknown; need for methodological research

## Contact Information

Jill M. Montaquila

1600 Research Blvd., Room RE 482

Rockville, MD 20850

[jillmontaquila@westat.com](mailto:jillmontaquila@westat.com)