

## Using Ancillary Information to Facilitate Address-Based Sampling

Ashley Amaya, NORC at the University of Chicago  
Ned English, NORC at the University of Chicago  
AAPOR Annual Conference  
May 2011



---

---

---

---

---

---

---

---

---

---

### Background – Overview of DSF Coverage

- The DSF is not perfect
  - Overcoverage
    - 136 million residential addresses on frame<sup>1</sup> vs.
    - 112.6 million households in U.S.<sup>2</sup>
  - Undercoverage
    - Rural areas as low as 49% housing units<sup>3</sup>
    - Urban Census tracts as high as 100% incomplete addresses



---

---

---

---

---

---

---

---

---

---

### Background – Overview of ABS Sampling

- Different surveys use different mechanisms
  - Include / Exclude lines based on frame flags
    - Nielsen<sup>4</sup> vs. NORC
  - Stratify based on appended flags
    - REACH U.S.
  - Enhance with lists
    - NHES ABS Pilot<sup>5</sup>
  - Enhanced / Traditional listing
    - RECS 2010



---

---

---

---

---

---

---

---

---

---

**Background – Research Question**

- Where are the households?
  - Investigate DSF coverage to identify:
    - Overcoverage
      - Vacant housing units
      - Seasonal housing units
      - Group quarters
    - Undercoverage
      - Drop points
      - Simplified addresses
      - Include rural households

---

---

---

---

---

---

---

---

---

---

**Methods**

- Evaluate ability of frame flags to identify households / non-households
  - Vacant
  - Seasonal
  - College
  - Drop points
- Not included are
  - Simplified Addresses
  - Rural Addresses

---

---

---

---

---

---

---

---

---

---

**Methods (cont'd)**

- Two part investigation
  - Qualitative Field Observations
    - 8 zip codes in Illinois, Indiana, and Michigan
    - Area layout and characteristics
    - Comparisons of flagged vs. unflagged addresses
  - Quantitative Analyses
    - 2 large, phone-based ABS projects
    - Descriptive statistics of observed resolution and residency rates

---

---

---

---

---

---

---

---

---

---

**Methods – Descriptive Information of Selected Zip Codes**

ZIP Code	City, State	Residential Addresses (N)	SEASON (%)	DROP (%)	COLLEGE (%)
60624	Chicago, IL	16,047	0.0	33.4	0.0
46590	Winona Lake, IN	3,010	0.0	33.2	33.2
49125	Sawyer, MI	1,447	24.1	0.6	0.0
49007	Kalamazoo, MI	5,235	0.0	1.9	30.1
49401	Allendale, MI	7,899	0.0	22.4	45.6
48818	Crystal, MI	1,342	13.8	12.0	0.0
49636	Glen Arbor, MI	657	15.8	1.1	0.0
49635	Frankfort, MI	2,520	18.1	2.8	0.0

**Overcoverage**

- Vacancy
- Seasonal
- College

**Results – Vacant Housing Units**

- What would happen if you excluded addresses flagged as “vacant”?
  - An effort to minimize overcoverage
- Analysis
  - Quantitative descriptive statistics
- The real world
  - 14.3% of U.S. housing units are vacant<sup>6</sup>

**Results – Vacant Housing Units (cont'd)**

- The DSF's world
  - 6.5% residential addresses are flagged vacant
  - Can exclude without any false-positives
  - Remaining 7.8% will be false-negatives

**Results – Comparison of Expected vs. Observed Vacancy Rates**

Project Observations	Expectations Based on DSF Flag					
	Not Flagged as Vacant			Flagged as Vacant		
	Phone Match	No Phone	Subtotal	Phone Match	No Phone	Subtotal
N	58,854	42,191	101,045	1,248	4,140	5,388
Undeliverable (%)	6.2	11.6	8.4	37.8	73.4	65.2
Occupied (%)	50.2	24.4	39.4	22.0	4.8	8.8
Unresolved (%)	43.6	64.0	52.2	40.2	21.8	26.0

**Results – Vacant Housing Units (cont'd)**

- Trusting the DSF flag eliminates
  - 6.5% of residential addresses
  - Up to 2.0%\* of households on the frame
- Using the DSF flag within a phone + flag model eliminates
  - 3.9% of residential addresses
  - Up to 1.2%\* of households on the frame

\* Equation: (% of sample flagged as vacant \* (100 - % identified as vacant)) / ((% of sample flagged as vacant \* (100 - % identified as vacant)) + (% of frame not flagged as vacant \* (100 - % identified as vacant)))

**Results – Seasonal Addresses**

- What would happen if you excluded addresses flagged as “seasonal”?
  - An effort to minimize overcoverage
- Analysis
  - Quantitative descriptive statistics
  - Qualitative observation
- In the real world
  - Don't know how many seasonal addresses

---

---

---

---

---

---

---

---

---

---

**Results – Seasonal Addresses (cont'd)**

- In the DSF's world
  - Addresses flagged seasonal if away for 3+ months/year
  - 0.6% residential addresses are flagged seasonal
  - Can exclude without any false-positives

---

---

---

---

---

---

---

---

---

---

**Results – Comparison of Expected vs. Observed Seasonal Addresses**

Project Observations	Expectations Based on DSF Flag					
	Not Flagged as Seasonal			Flagged as Seasonal		
	Phone Match	No Phone	Subtotal	Phone Match	No Phone	Subtotal
N	59,831	46,211	106,042	271	120	391
Undeliverable (%)	7.1	17.0	11.4	54.3	58.3	55.5
Occupied (%)	49.2	22.7	37.7	22.9	12.5	19.7
Unresolved (%)	43.7	60.3	51.0	22.9	29.2	24.8

---

---

---

---

---

---

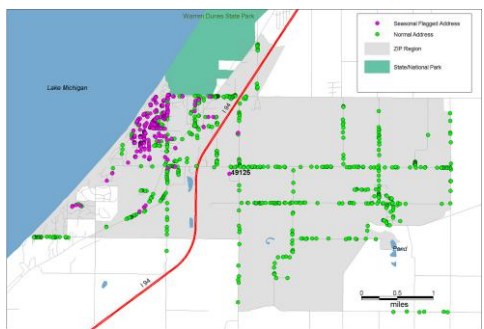
---

---

---

---

### Results – Seasonal and Non-Seasonal Housing Units as Flagged



---

---

---

---

---

---

---

---

---

---

### Results – Seasonal Addresses (cont'd)

- Trusting the DSF flag eliminates
  - 0.6% of residential addresses
  - Up to 0.2% of households on the frame
- Using the DSF flag to impute entire block groups eliminates
  - Approximately 1.1% of residential addresses
  - Need to evaluate accuracy

---

---

---

---

---

---

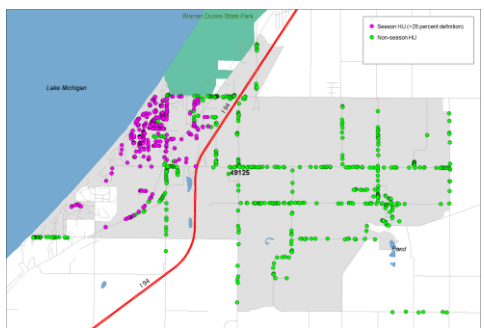
---

---

---

---

### Results – Seasonal and Non-Seasonal Housing Units as Imputed



---

---

---

---

---

---

---

---

---

---

### Results – College Addresses

- What would happen if you excluded addresses flagged as “educational”?
  - An effort to minimize overcoverage
- Analysis
  - Qualitative observation
- In the DSF’s world
  - Addresses flagged for student housing
    - Varies by postal carrier
  - 0.3% residential addresses are flagged as educational / college
  - Can exclude without any false-positives

---

---

---

---

---

---

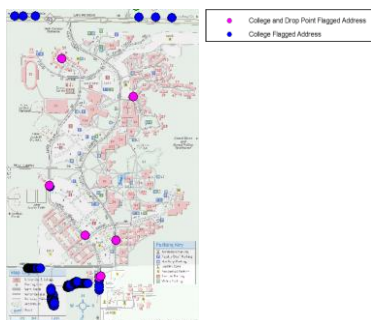
---

---

---

---

### Results – Grand Valley State University (49401)



---

---

---

---

---

---

---

---

---

---

### Results – College Addresses (cont’d)

- Trusting the DSF flag eliminates
  - 0.3% of residential addresses
  - Up to 0.3% of households on the frame
- Using the DSF flag to stratify
  - Maximize / Minimize 18-25 year olds on frame

---

---

---

---

---

---

---

---

---

---

### Undercoverage



#### Drop Points

22

---

---

---

---

---

---

---

---

### Results – Drop Points

- How can we include them?
  - Minimize undercoverage
- Analysis
  - Qualitative observation
- In the DSF's world
  - 1.9% of residential addresses are drop points
  - Randomly scattered throughout U.S.
    - Similar to other housing units (i.e. no bias)
  - Could assign unit numbers 1 through  $n$



23

---

---

---

---

---

---

---

---

### Results – Drop Point and non-Drop Point Housing Units as Flagged



24

---

---

---

---

---

---

---

---





### Summary (cont'd)

- Next Steps
  - Test accuracy of these procedures
    - Field a sample of imputed seasonal and drop point addresses
  - Evaluate
    - Bias
    - Scalability
  - Begin to tackle other coverage problem areas
    - Simplified addresses
    - Rural exclusions

### Works Cited

1. Lattemer, Michael, Ned English, & Colm O'Muircheartaigh. "Capturing Change in Address Lists: a Spatial-Temporal Analysis." *Presentation at the American Association of Public Opinion Research Annual Conference*, 2010.
2. U.S. Census Bureau. 2005-2009 American Community Survey. [http://factfinder.census.gov/servlet/STTable?\\_bmsy&geo\\_id=01000US&q\\_1\\_name=ACS\\_2009\\_5YR\\_G00\\_S25048-ds\\_name=ACS\\_2009\\_5YR\\_G00](http://factfinder.census.gov/servlet/STTable?_bmsy&geo_id=01000US&q_1_name=ACS_2009_5YR_G00_S25048-ds_name=ACS_2009_5YR_G00) . Accessed: February 17, 2011.
3. O'Muircheartaigh, Colm, Ned English, Stephanie Eckman, Heidi Upchurch, Erika Garcia, & James Lepkowski. "Validating a Sampling Revolution: Benchmarking Address Lists against Traditional Listing." *Presentation at the Joint Statistical Meetings*, 2006. <http://www.amstat.org/sections/srms/proceedings/y2006/Files/US12006-000009.pdf>.
4. Link, Michael W., Gail Daily, Charles D. Shuttles, Tracie L. Yancey, & H. Christine Bourquin. "Building a New Foundation: Transitioning to Address Based Sampling After Nearly 30 Years of RDD." *Presentation at the American Association of Public Opinion Research Annual Conference*, 2009. <http://www.amstat.org/sections/srms/proceedings/y2009/Files/409029.pdf>.
5. Montaguilla, Jill M., Douglas Williams, & Daifeng Han. "An Application of a Two-Phase Address Based Approach to Sampling for Subpopulations." *Presented to the Washington Statistical Society*, March 1, 2011.
6. U.S. Census Bureau. 2010 Annual Housing Vacancies and Homeownership Statistics (CPS:HVS). <http://www.census.gov/hhes/www/housing/hvs/annual10/ann10ind.html>. Accessed: May 4, 2011.

Ashley Amaya  
Amaya-Ashley@norc.org

Thank You!