



# Estimating Net Coverage of ABS Frames

Rachel Harter, Katherine B. Morton,  
Ashley Amaya

RTI International, 3040 Cornwallis Road, Research Triangle Park, NC 27709

# Acknowledgement

- This work was partially supported by the Energy Information Administration and used frame and sample data from the 2015 Residential Energy Consumption Survey (RECS).
- The 2015 RECS was funded by Energy Information Administration (EIA), Department of Energy under 2015 RECS Contract Nos. DE-EI-0000515.
- The views expressed in this presentation do not necessarily reflect the official policies of the EIA, Department of Energy, nor does mention of trade names, commercial practices, or organizations imply endorsement by the U.S. Government.

# ABS and Coverage Basics

# Address-Based Sampling (ABS)

- ABS uses an address frame derived from U.S. Postal Service data.
- Selects a sample of housing units (HUs).
- See AAPOR (2016) and Iannacchione (2011).

Here we assume group quarters are not eligible.

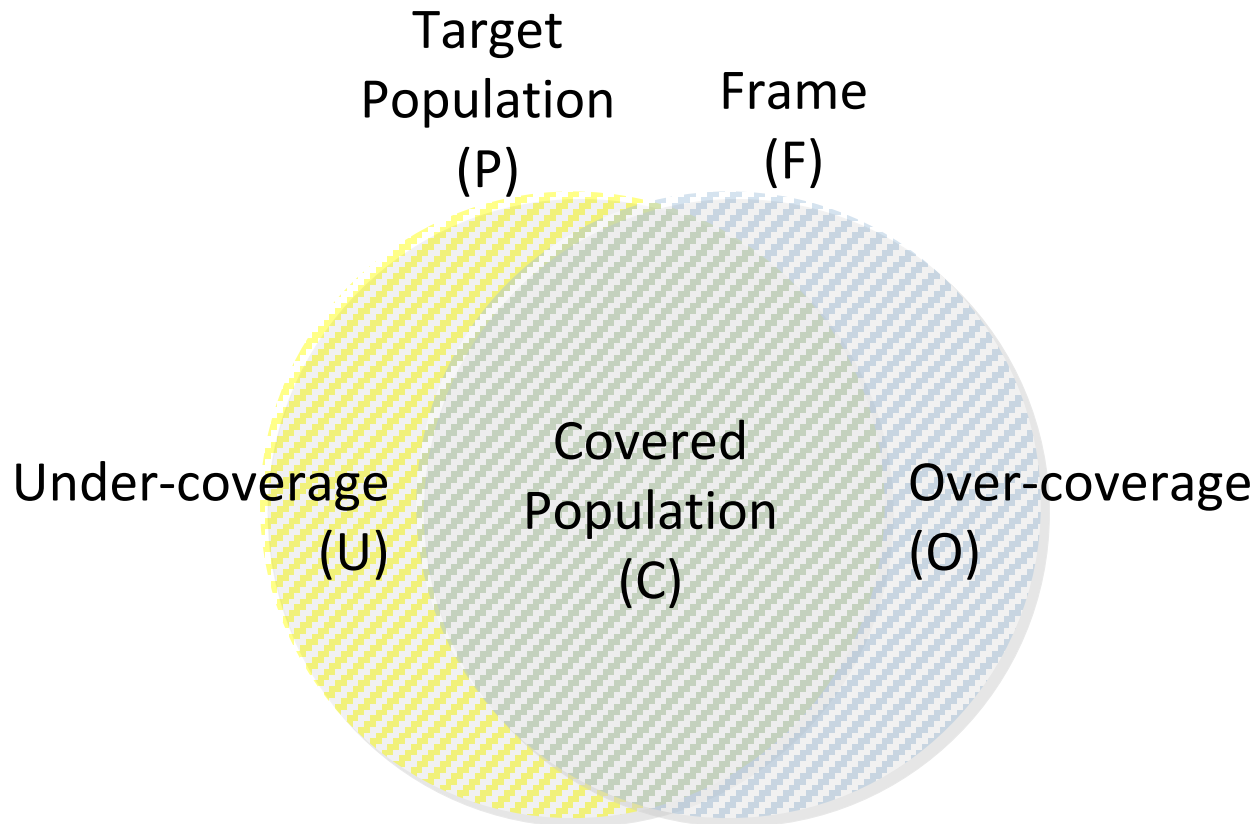
# Known Coverage Problems with ABS Frames

- **Unlocatable addresses** such as PO boxes
- Geocoding error
- Group quarters
- Mobile homes
- Simplified addresses
- AIAN lands
- Units within drop points
- Irregular homes
- Vacant and newly constructed homes

In general, coverage is more likely to be a problem in rural areas.

(Dorhmann et al. (2006, 2007); Dohrmann and Sigman (2013); Eckman and English (2012); Iannacchione et al. (2012); Kennel and Li (2009); McMichael (2015); O'Muircheartaigh et al. (2006, 2007); Shook-Sa et al. (2010); Unangst and McMichael (2015); Zandbergen, (2011))

# Coverage Concepts



$$\text{Coverage rate} = \frac{C}{P}$$

$$\text{Net coverage rate} = \frac{F}{P}$$

(Kish 1965)

# Research Questions and RECS Data for This Research

# 2015 Residential Energy Consumption Survey (RECS)

- Purpose - estimate energy costs and usage
- Target Population – Occupied Primary Housing Units (HUs) in U.S.
- Stratified 3-Stage Design
- Within 2<sup>nd</sup> stage units (segments), create HU frames



# 3 Types of Segments in RECS

- ABS frame (547 **ABS** segments)
- ABS frame supplemented with field searches (213 **SUP** segments)
- Field Enumerated frame (40 **FE** segments)

ABS frames available for all 3 types of segments, even if used only to determine how to classify each segment.

# Research Questions

- How well did our net coverage estimates categorize segments for RECS?
- What should we consider when preparing net coverage estimates?

# Estimating ABS Coverage for Field-Enumerated Segments in RECS After the Fact

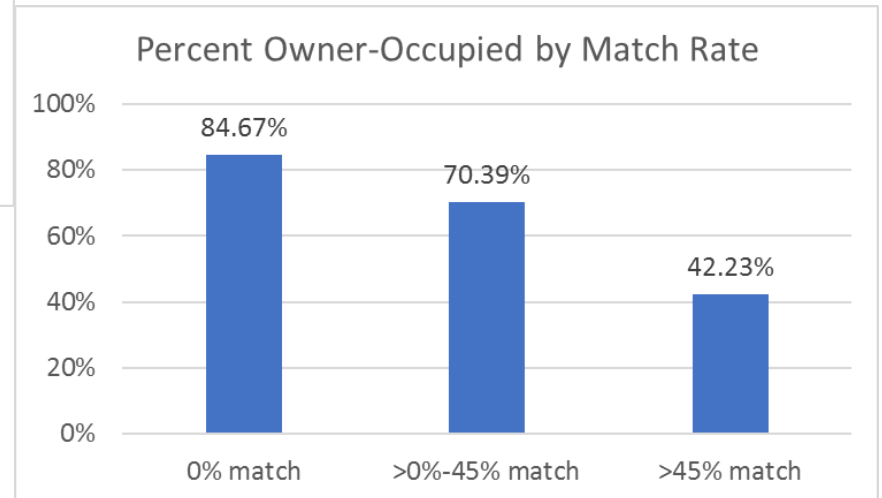
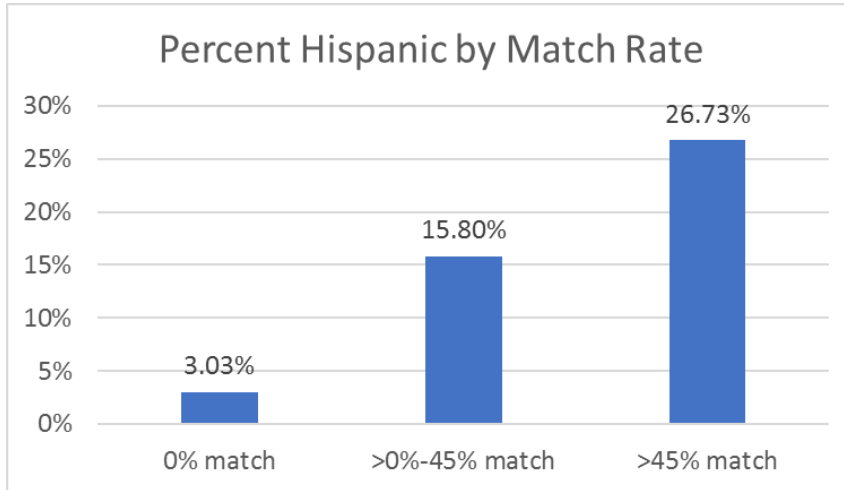
# Coverage of 2015 RECS FE Segments: Analysis Steps

- All FE segments initially estimated to have 56% net coverage or less.
- For FE segments, match addresses found in the field to the ABS frames.
  - Assumes FE listings are “truth”, which is not necessarily a good assumption. (Eckman and Kreuter 2013, Cunningham et al. 2006)
- Summarize match rates, undercoverage, and overcoverage.
- Summarize characteristics of segments by coverage rates.

# RECS Address Match Rates, 40 FE Segments

<b>Coverage and Match Rates for 40 FE Segments Combined</b>		
<b>Class of HUs</b>	<b>Count</b>	<b>Percent</b>
Matched Addresses (coverage lower bound)	2,839	27%
FE Matched outside segment (geocoding error)	1,762	17%
FE Only (other undercoverage and match error)	5,844	56%
Total FE HUs	10,445	100%
ABS only (overcoverage upper bound)	2,219	21%
Total ABS (net coverage, coverage upper bound)	5,058	48%

# RECS FE Segment Characteristics By Match Rate

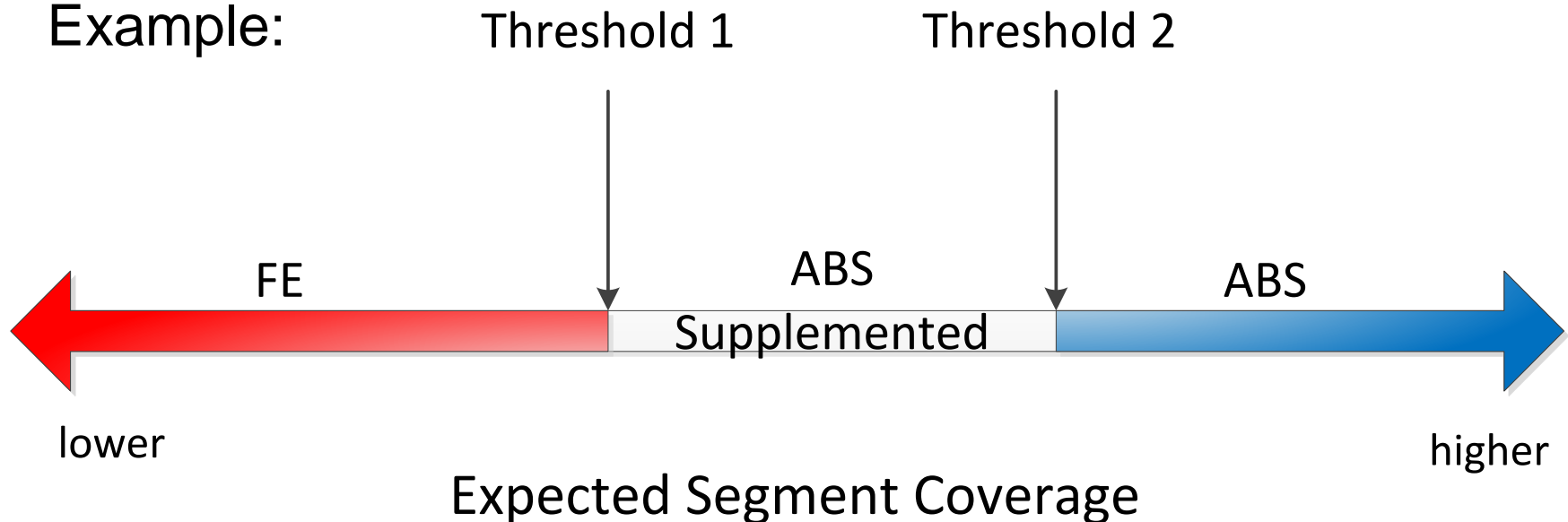


# Predicting Segment Coverage (at least relative to a threshold)

# Goal: Predict Coverage of Individual Geographic Segments

Alternatively, predict if coverage is above or below a specified threshold value.

Example:



A study typically has 1 or 2 thresholds, depending on the number of frame construction methods.



# Prior Work in Predicting Coverage

Models for estimating coverage or classifying segments relative to thresholds:

- O'Muircheartaigh et al. (2007, 2009): regression trees.
- Hsu et al. (2010) and Montaquila et al. (2011): linear regression model.

Must have prior FE listings (estimates of truth) to fit models.

Must have auxiliary variables at the segment level.

# No Standards for Estimating Coverage or Net Coverage

(No standards for establishing thresholds, either.)

## **Models:**

Become less reliable over time, and cannot be updated without new “truth”.

## **Net coverage ratios (frame count / “truth”):**

Quick and easy.

No auxiliary variables required except “truth”.

Masks the mix of overcoverage and undercoverage.

# Estimating Net Coverage Using All RECS Segments

- Compute various net coverage ratios for all 800 RECS segments.
- Compare classifications for different threshold values.
- Examine mismatches between two net coverage ratios.

# Options for Net Coverage Ratios

- Numerator - ABS frame exclusions
- Denominator - Choice of “truth” for comparison
- Purpose of net coverage ratio

# ABS Frame Counts for Net Coverage Numerators

Remove PO boxes that are not an HU's only way to get mail.

Optionally add active addresses from No-Stat file to increase coverage. (Shook-Sa et al. 2013)

## Face-to-Face Surveys

- Remove unlocatable addresses (all PO boxes, etc.)

## Mail Surveys

- Remove drop points and their units

# “True” Counts for Net Coverage Denominators

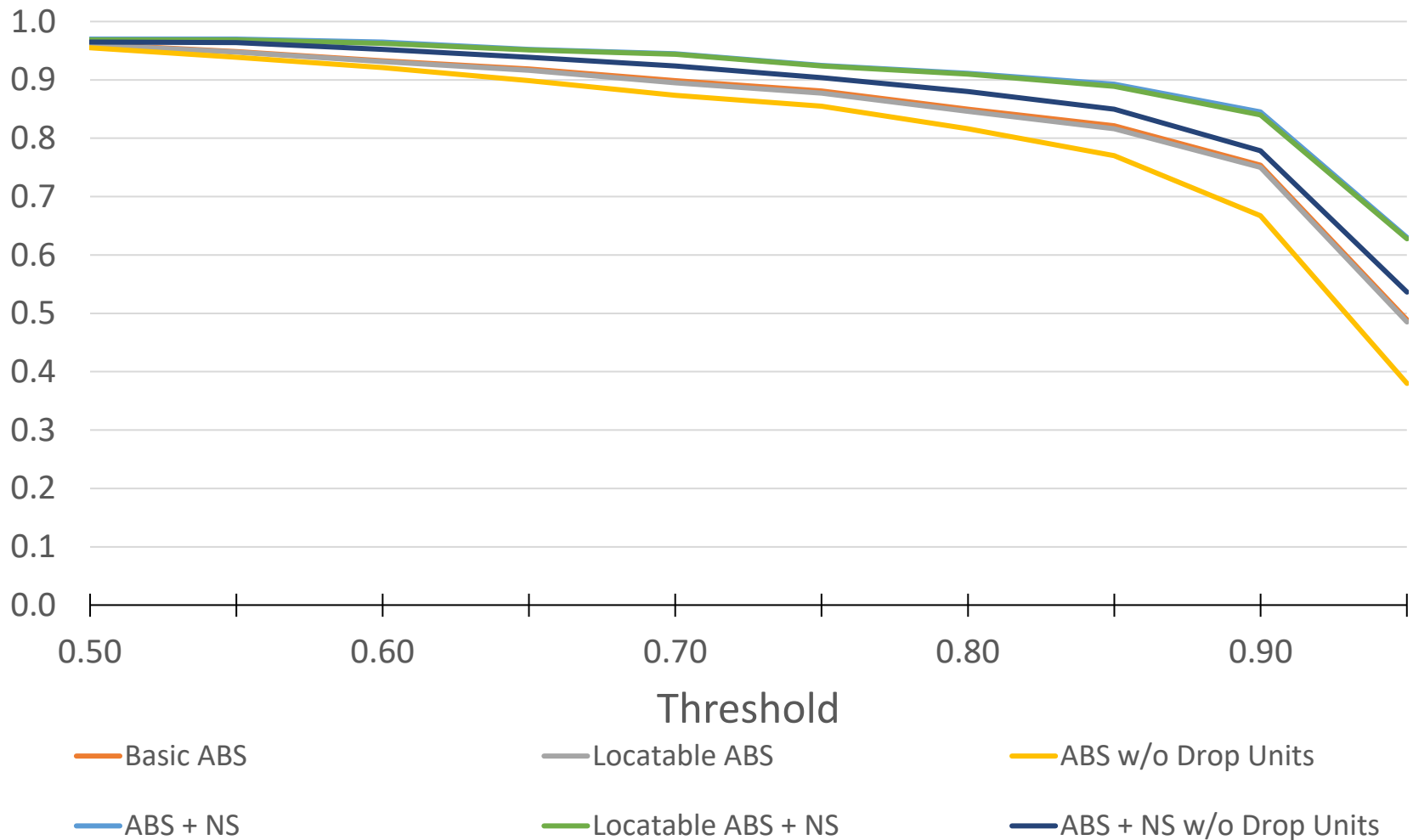
## Source

- Decennial Census (Census)
- American Community Survey (ACS)
- Claritas (CL)

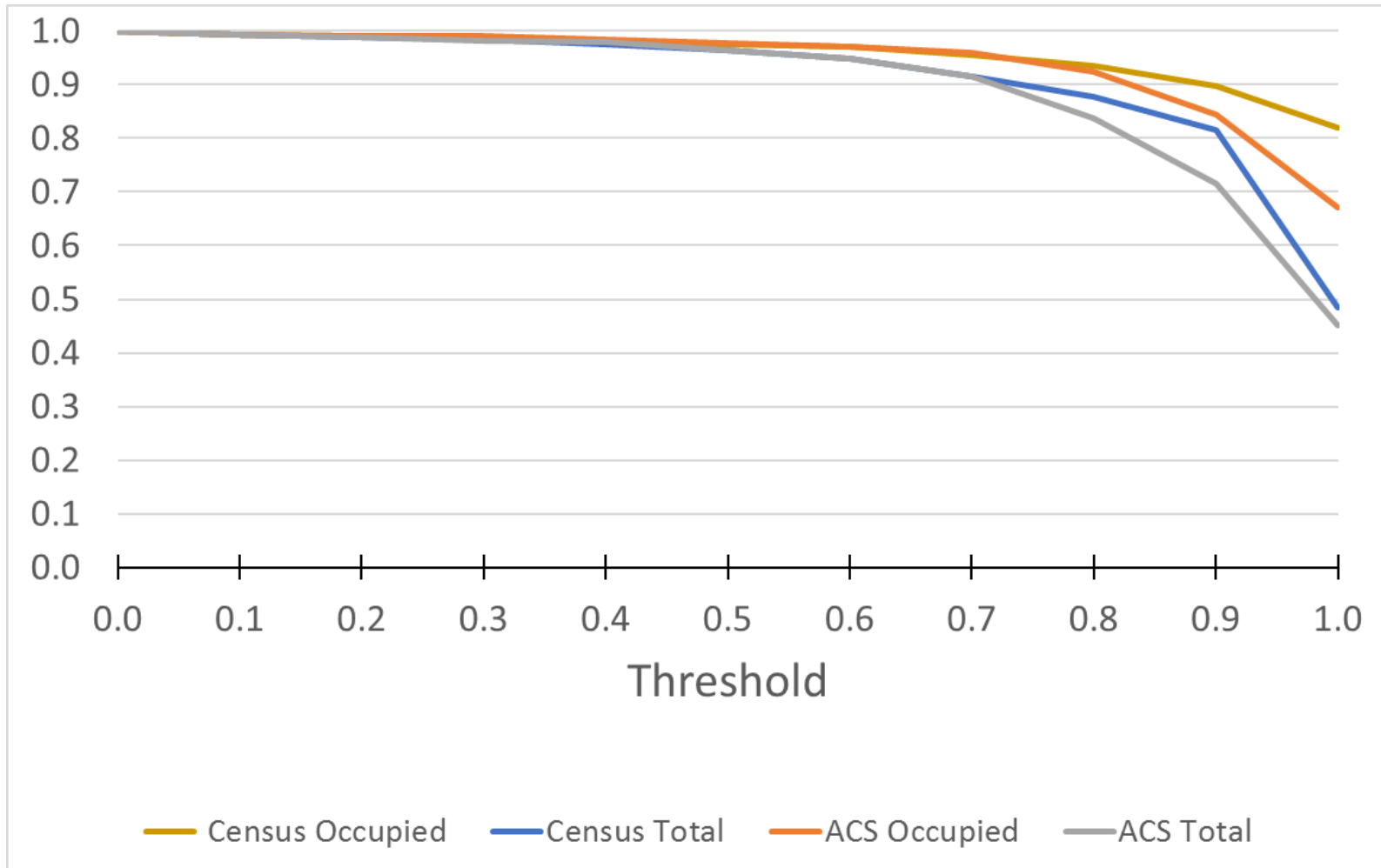
## Measure

- Total HUs (Tot)
- Occupied HUs (Occ)

# Proportions of Segments Exceeding Threshold: Ratio = Various Numerators / Census Total



# Proportions of Segments Exceeding Threshold: Ratio = Locatable ABS / Various Denominators





# Pairwise Comparison: With and Without NS File

<b>Effect of Augmenting Frame with Active NoStat Addresses: Percentage of Estimates Exceeding Threshold</b>				
<b>Threshold=0.85</b>				
		<b>ABS+NS/CensusOcc</b>		
		<b>0</b>	<b>1</b>	<b>Total</b>
<b>ABS/CensusOcc</b>	<b>0</b>	4.5%	3.0%	7.5%
	<b>1</b>	0%	92.5%	92.5%
	<b>Total</b>	4.5%	95.5%	100%

# Pairwise Comparison: Occupied vs. Total HUs

<b>Effect of Using Census Occupied or Total HUs as “Truth”: Percentage of Estimates Exceeding Threshold</b>				
<b>Threshold=0.85</b>				
		<b>ABS/CensusTot</b>		
		<b>0</b>	<b>1</b>	<b>Total</b>
<b>ABS/CensusOcc</b>	<b>0</b>	7.5%	0%	7.5%
	<b>1</b>	7.5%	85.0%	92.5%
	<b>Total</b>	15.0%	85.0%	100%

# Pairwise Comparison: Census vs. ACS

<b>Effect of Using Census or ACS as “Truth”: Percentage of Estimates Exceeding Threshold</b>				
<b>Threshold=0.85</b>				
		<b>ABS/ACS_Occ</b>		
		<b>0</b>	<b>1</b>	<b>Total</b>
<b>ABS/CensusOcc</b>	<b>0</b>	6.1%	1.4%	7.5%
	<b>1</b>	4.6%	87.9%	92.5%
	<b>Total</b>	10.8%	89.3%	100%

# Considerations for Predicting Net Coverage of Segments

- **Be aware of frame inclusions and exclusions**
  - How many addresses dropped as unlocatable or undeliverable?
- **Know the types of segments you have**
  - Which segments are expected to have problems?
- **Consider choice of “truth”**
  - Decennial census may be outdated, but only option for blocks
  - ACS or Claritas may be good options for block groups or larger
  - Which estimate of “truth” (Occ or Total HUs) best represents your purpose?
- **Clarify purpose of net coverage estimates**
  - ABS as frame for *occupied* HUs or as replacement for FE *total* HUs?

# References

- American Association for Public Opinion Research (2016). *Address-based Sampling*. Report prepared for AAPOR Council by the Task Force on Address-based Sampling.
- Cunningham, D., Hunter, S., Justin, L., Morton, K., & Stolzenberg, S. (2006). *2005 National Survey on Drug Use and Health Count and List Validity Study: Overall summary report* (prepared for the Substance Abuse and Mental Health Services Administration, Office of Applied Studies, under Contract No. 283-2004-00022, RTI/0209009). Research Triangle Park, NC: RTI International.
- Dohrmann, S., Han, D., & Mohadjer, L. (2006). Residential address lists vs. traditional listing: Enumerating households and group quarters. In *Proceedings of the 2006 Joint Statistical Meetings, American Statistical Association, Survey Research Methods Section, Seattle, WA* (pp. 2959-2964). Alexandria, VA: American Statistical Association.
- Dohrmann, S., Han, D., & Mohadjer, L. (2007). Improving coverage of residential address lists in multistage area samples. In *Proceedings of the 2007 Joint Statistical Meetings, American Statistical Association, Section on Survey Research Methods, Salt Lake City, UT* (pp. 3219-3126). Alexandria, VA: American Statistical Association.
- Dohrmann, S., & Sigman, R. (2013). Using an area linkage method to improve the coverage of abs frames for in-person household surveys. In *Proceedings of Federal Committee on Statistical Methodology (FCSM) research conference*. Washington, DC: Federal Committee on Statistical Methodology.

# References (cont.)

- Eckman, S., & English, N. (2012a). Geocoding to create survey frames. *Survey Practice*, 5(4), 1-8. Retrieved from <http://www.surveypractice.org/index.php/SurveyPractice/article/viewFile/29/pdf>
- Eckman, S., & Kreuter, F. (2013). Undercoverage rates and undercoverage bias in traditional housing unit listing. *Sociological Methods and Research*, 42, 264-293. <https://doi.org/10.1177/0049124113500477>
- Hsu, V., Montaquila, J. M., & Brick, J. M. (2010). Using a "match rate" model to predict areas where USPS-based address lists may be used in place of traditional listing. In *Proceedings of the 2010 Joint Statistical Meetings, American Statistical Association, Section on Survey Research Methods, Vancouver, British Columbia, Canada* (pp. 1549-1560). Alexandria, VA: American Statistical Association.
- Iannacchione, V. G. (2011). *The changing role of address-based sampling in survey research*. *Public Opinion Quarterly*, 75(3), 556–575.
- Iannacchione, V., McMichael, J., Shook-Sa, B., & Morton, K. (2012). *A proposed hybrid sampling frame for the National Survey on Drug Use and Health* (prepared for the Substance Abuse and Mental Health Services Administration under Contract No. 283-2004-00022). Research Triangle Park, NC: RTI International.

# References (cont.)

- Kennel, T. L., & Li, M. (2009). Content and coverage quality of a commercial address list as a national sampling frame for household surveys. In *Proceeding of the 2009 Joint Statistical Meetings, American Statistical Association, Section on Survey Research Methods, Washington, DC* (pp. 2364-2378). Alexandria, VA: American Statistical Association.
- Kish, Leslie. 1965. *Survey Sampling*. New York: John Wiley & Sons.
- McMichael, J. P. (2015, August). ABS coverage evaluation: Recommendations for evaluating the household coverage of address-based sampling frames. In *Proceedings of the 2015 Joint Statistical Meetings, American Statistical Association, Section on Survey Research Methods, Seattle, WA* (pp. 2279-2280). Alexandria, VA: American Statistical Association.
- Montaquila, Jill M., Valerie Hsu, and J. Michael Brick. 2011. "Using a "match rate" model to predict areas where USPS-based address lists may be used in place of traditional listing." *Public Opinion Quarterly* 75:317-335.
- O'Muircheartaigh, C., English, N., Eckman, S., Upchurch, H., Garcia, E., & Lepkowski, J. (2006). Validating a sampling revolution: Benchmarking address lists against traditional listing. In *Proceedings of the 2006 Joint Statistical Meetings, American Statistical Association, Section on Survey Research Methods, Seattle, WA* (pp. 4182-4196). Alexandria, VA: American Statistical Association.

# References (cont.)

- O'Muircheartaigh, C., English, N., & Eckman, S. (2007). Predicting the relative quality of alternative sampling frames. In *Proceedings of the 2007 Joint Statistical Meetings, American Statistical Association, Section on Survey Research Methods, Salt Lake City, UT* (pp. 3239-3248). Alexandria, VA: American Statistical Association.
- O'Muircheartaigh, C., English, N., Latterner, M., Eckman, S., & Dekker, K. (2009). *Modeling the need for traditional vs. commercially available address listings for in-person surveys: Results from a national validation of addresses*. American Association for Public Opinion Research Conference, Hollywood, FL. In *Proceedings of the Section on Survey Research Methods, American Statistical Association*, pp. 6193-6202.
- Shook-Sa, Bonnie E., Douglas B. Currivan, Joseph P. McMichael, and Vincent G. Iannacchione (2013). "Extending the coverage of address-based sampling frames beyond the USPS computerized delivery sequence file." *Public Opinion Quarterly* 77:994-1005.
- Shook-Sa, B. E., McMichael, J. P., Ridenhour, J. L., & Iannacchione, V. G. (2010). The implications of geocoding error on address-based sampling. In *Proceedings of the 2010 Joint Statistical Meetings, American Statistical Association, Section on Survey Research Methods* (pp. 3303-3312). Alexandria, VA: American Statistical Association.



# References (cont.)

- Unangst, J. J., & McMichael, J. (2015). Tracking and evaluating changes to address-based sampling frames over time. In *Proceedings of the 2015 Joint Statistical Meetings, American Statistical Association, Section on Survey Research Methods, Seattle, WA* (pp. 4310-4316). Alexandria: American Statistical Association.
- Zandbergen, P. A. (2011). Influence of street reference data on geocoding quality. *Geocarto International*, 26(1), 35-47.

# Contact Information

**Rachel Harter**

**Senior Research Statistician**

**919-541-6472**

**[rharter@rti.org](mailto:rharter@rti.org)**