

turning knowledge into practice

Linking Mailing Addresses to a Household Sampling Frame Based on Census Geography

Katherine Morton, Joseph McMichael, James Cajka, Ross Curry,
Vincent Iannacchione, and David Cunningham
AAPOR, Anaheim, CA, May 17-20, 2007



RTI International is a trade name of Research Triangle Institute

3040 Cornwallis Road
Phone 919-541-7181

■ P.O. Box 12194

■ Research Triangle Park, North Carolina, USA 27709

e-mail bowman@rti.org

1

Introduction

- Census geography often serves as the primary sampling unit in household surveys.
- Counting and listing methods can be used to create household frames within sampled areas. However, these methods are often time and cost prohibitive.
- Residential mailing lists are inexpensive and can be purchased closer to the start of data collection.

Introduction (continued)

- Prior research has shown undercoverage of rural areas and group quarters in mailing lists.
- This presentation:
 - assumes the use of residential mailing addresses,
 - provides a method for linking mailing addresses to Census geography (geocoding),
 - evaluates geocoding, and
 - suggests sources for improving the coverage of mailing lists.

Census Geography vs. Postal Geography

	Pros	Cons
Census Geography	<ul style="list-style-type: none">■ Complete geographic coverage of target population■ Matching to external data	<ul style="list-style-type: none">■ Geocoding error
Postal Geography	<ul style="list-style-type: none">■ Complete list of mailing addresses within sampled area	<ul style="list-style-type: none">■ Incomplete geographic coverage

Mailing Address versus Count and List (MAVCAL) Research Project

- Selected 50 area segments (one or more Census blocks) across the state of North Carolina with probability proportional to size.
- Mailing lists were purchased for the sampled areas.
- Field staff enumerated the households in sampled areas using traditional count and list methods.
- After count and list was complete, field staff were provided the residential mailing addresses for the sampled area. They then returned to the segment to verify the accuracy of the mailing list.

MAVCAL Research Project (continued)

- For RTI's research project, a Geographic Information System (GIS) was used to overlay the postal carrier routes with the Census geography.
- In order to ensure that all mailing addresses within the sampled areas were purchased, a buffer zone was created.

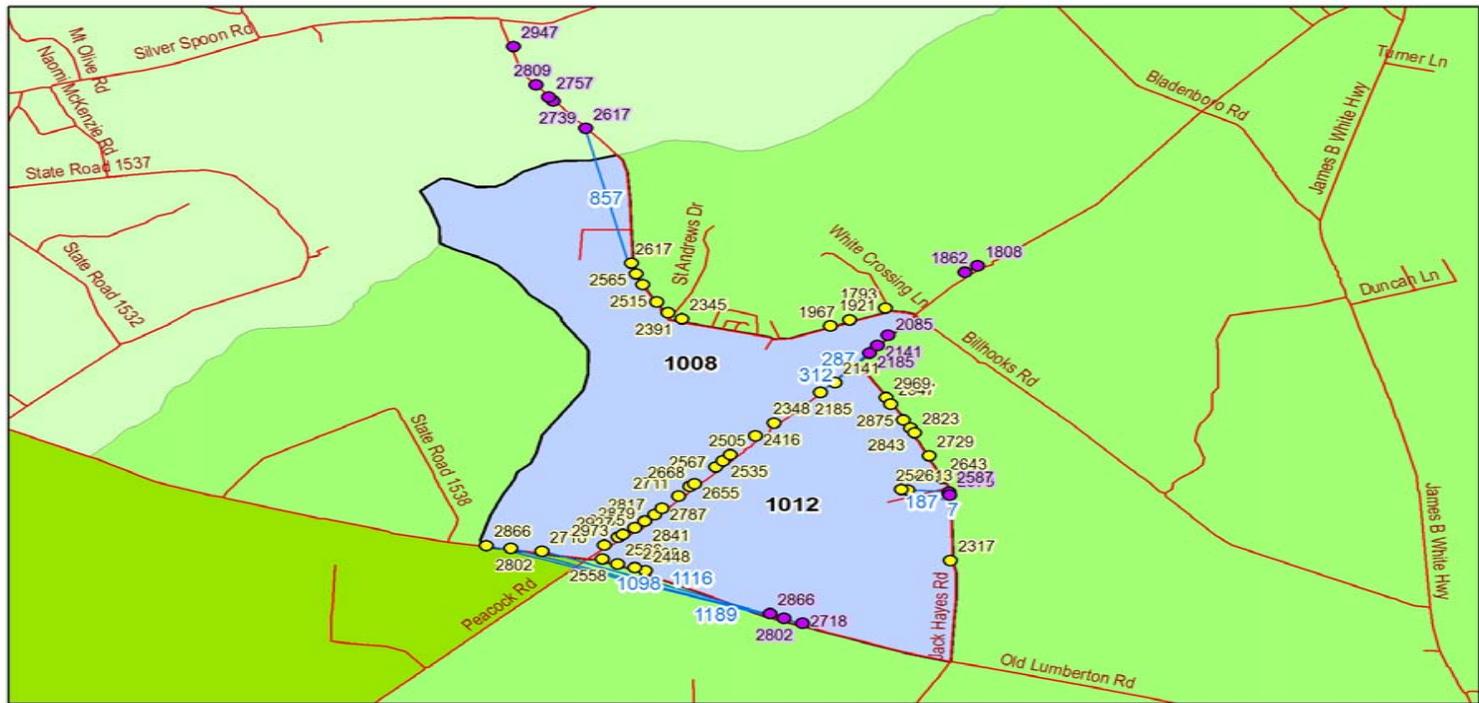
Linking Mailing Addresses to Census Geography (continued)

- Geocoding: the process by which geographic coordinates are assigned to mailing addresses by interpolating the address location along a section of street based on address ranges assigned to that section.
- Using geocoding, a mailing list vendor can determine the Census geography (e.g., blocks, block groups, Census tracts) for any given address.
- The vendor can then provide all mailing addresses within the sampled Census geography.
- How accurate is geocoding?

Evaluation of Geocoding for Linking Mailing Addresses to Census Geography

Accuracy of Geocoding	Rural (%)	Urban (%)	Total (%)
Correct Block	37.6	73.4	65.6
Correct Block Group	89.9	91.7	91.3
Correct Census Tract	91.7	92.6	92.4
Correct County	99.9	100.0	100.0
<i>Sample Size</i>	<i>1,200</i>	<i>4,318</i>	<i>5,518</i>

Evaluation of Geocoding for Linking Mailing Addresses to Census Geography (continued)



- Geocoded Coordinates
- GPS Coordinates
- TIGER road
- MAVCAL segment
- TIGER blockgroups

0 250 500 750 1,000 Meters

Sources for Supplementing Mailing Lists

- Households without city-style mailing addresses (e.g. telephone books, driver's license files, credit card files)
- Group quarters (e.g. an education survey frame for college dormitories)
- Counting and listing (e.g. in rural areas with poor coverage of mailing addresses)
- Half-open interval frame linking procedure

Conclusions

- Census geography is preferred to postal geography because it ensures complete geographic coverage of the target population.
- Geocoding can be used to link mailing addresses to Census geography.

Conclusions (continued)

- Due to geocoding error, it is recommended that larger geographic areas (e.g. block groups or Census tracts) serve as the primary sampling units when sampling from a residential mailing list.
- Sampling households across larger geographic areas will also decrease the variance of estimates since clusters are less homogeneous (i.e. there is less intra-cluster correlation).
- Supplemental sources can be used to improve the coverage of mailing lists.

Additional Information

- Presentation: www.rti.org/aapor
- Email: bowman@rti.org