Correcting Coverage Deficiencies in Address-Based Frames: The Use of Enhanced Listing

Ned English, Katie Dekker, and Colm O'Muircheartaigh
NORC at the University of Chicago

NORC
at the UNIVERSITY *of* CHICAGO

Outline

- The delivery-sequence file (DSF or CDSF) in survey research
- Coverage evaluations
- "Enhanced" (dependent) listing where imperfect
- Results of national evaluation of listing methods

Introduction

- "Traditional listing" former gold-standard
  - Record all addresses in selected areas
  - Time consuming, costly
- US Postal Service computerized delivery sequence file
  - All addresses receiving mail in USA
    – CDSF nearly complete coverage of USA
- Survey organizations examined alternatives since 2002
  - CDSF equivalent /better than traditional listing in urban, suburban
  - Still limited in rural areas with non city-style delivery
    – Requires listing or other augmentation

➢ Undercoverage is generally clustered, predictable

➢ How do we handle areas where DSF known insufficient?

Introduction Contd.

- In today's presentation we consider:
  - How should we determine when CDSF-alone insufficient?
    - When? Where?
  - What methods could we use to augment the list where deficient?
    - "Enhanced" or "Dependent" listing
  - How do methods compare?
    - Present evaluation results
  - What may be suitable in the future?

Background: The CDSF

- US Postal Service (USPS) delivery sequence file
  - All addresses receiving mail in USA
  - CDSF* had 98% coverage of USA in 2008 (Link et al.)
    - Much closer to 100% now
- Organizational tool for USPS
  - All mailable addresses in urban and suburban areas
  - All non-vacant mailable addresses in rural areas
  - Updated by individual carriers via "edit books"
    - Operational incentives for updating
- USPS provides licensing arrangements
  - Direct-mail, market research
  - *Valassis, MSG, SSI, CIS*

Background Contd: CDSF Evaluations

- Potential Application to Survey Research
  - Addresses in standard format
  - Can be geocoded and mapped, except in rural areas
    - Geocoding error affects coverage (Eckman and English 2011)
  - Basis for US Census Master Address File (MAF)

- Evaluated since early 2000s
  - Cover mid to upper 90's% of US households for face-to-face
    - Missing: simplified addresses, PO BOXes, long-term vacants
      - **Impact depends on mode, sampling approach**
  - Can be all non-city style in rural areas: no info re dwellings

Determining when CDSF Sufficient

- Multiple potential approaches:
    1. Direct: compare count of geocoded addresses to control
        - Requires: geocoding; reliable source of counts; threshold
            - **Limited by data (ACS? Market Research sources?)**
            - **Can be difficult in areas of rapid growth**
            - **Threshold definition situational, budget driven**

    2. Modeled: predict areas requiring augmentation
        - Census, NORC have pursued (O'Muircheartaigh et al. 2007, 2009)
        - Urbanicity, growth, demographics

    3. Follow Census: Type of Enumeration Area (TEA) code

- Recommend starting with *a priori* idea, and then examine
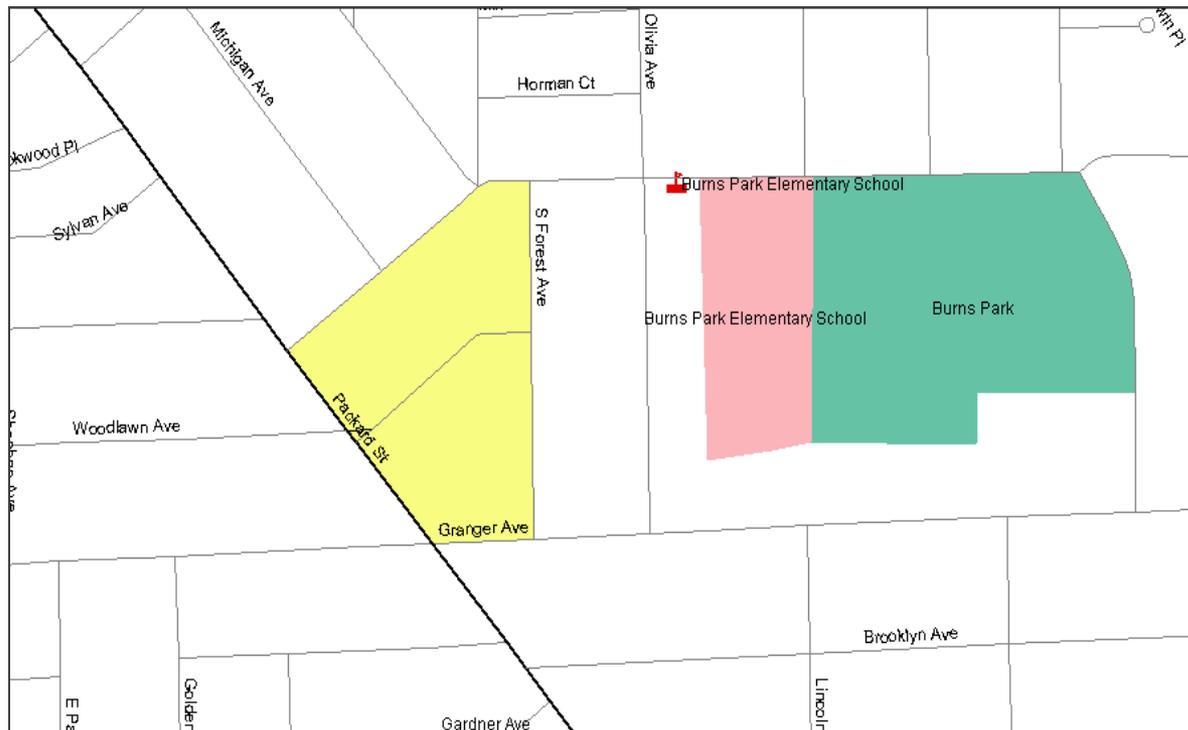
Where Listing Likely Necessary

2000
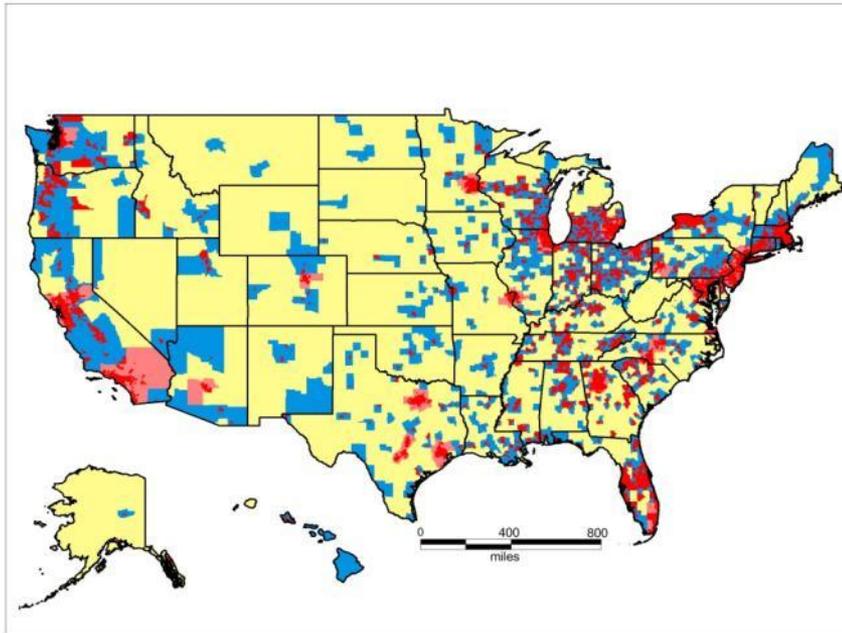
2010
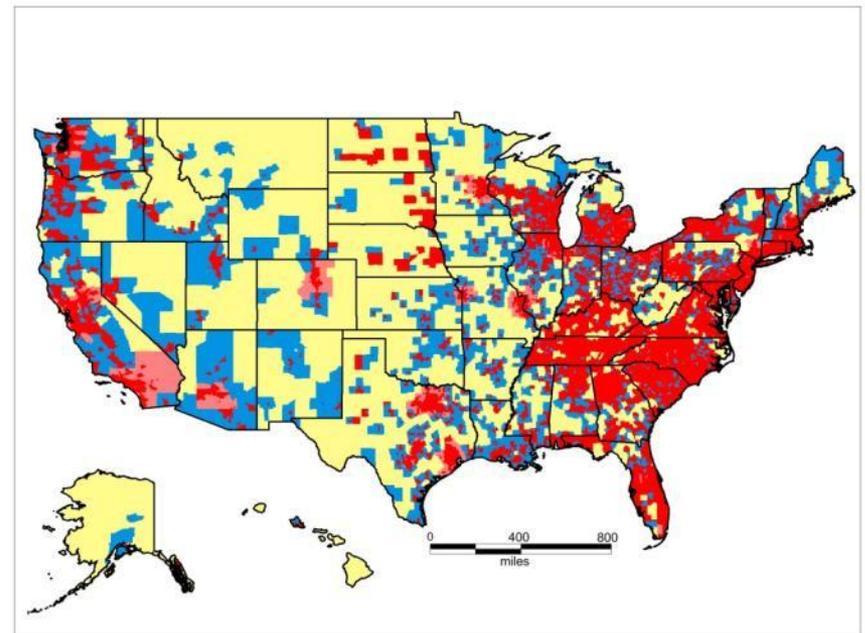
- Red areas "largest urban", Blue "mid-urban"
- Yellow "rural", likely requiring listing
- 71% pop 2000, 86% 2010 likely not requiring listing

➤ What if not sufficient?

"Traditional" Listing

- Lister given only map of selected area

- Record addresses by block via PAPI, hand-held

"Dependent" or "Enhanced" Listing

- Lister has map and initial frame (*input list*)
  - From previous listing
  - Or postal database
- Update frame in the field
  - *Add*   missing addresses
  - *Delete*   inappropriate addresses
  - *Confirm/Edit*   existing addresses

- Executed by Census for MAF update, NSFG (CDC)
- Can be done paper-and-pencil, electronically

Enhanced vs. Traditional

- Assume enhanced more efficient than traditional
  - Could it perpetuate known drawbacks of DSF in particular areas?
    - "Confirmation bias" (Eckman 2010)
  - Can we predict where one preferable?

- Conducted evaluation in 2011 in two counties
  - Mix of rural, urban, and suburban environments
  - 13 pairs of segments
  - One member of each pair listed traditionally
  - Second member listed using enhanced (dependent) method

- Listings were then independently checked
  - "Frame checking" for gold standard
  - Can see which method captured more "reality"

Evaluation of Methods Contd.

- Both E or T captured nearly all of "reality"
- Segments where one did somewhat better
  - Haphazard over-coverage on both lists
    – Question of cost
- Cost expressed as "minutes per unit"
  - Rural county: 3.9 for enhanced and 5.1 for traditional
  - Urban county: 2.9 for enhanced and 3.1 for traditional
- Enhanced did best in urban areas with substantial initial list

- Caveats: Small sample size, not-nationally representative, paper and pencil
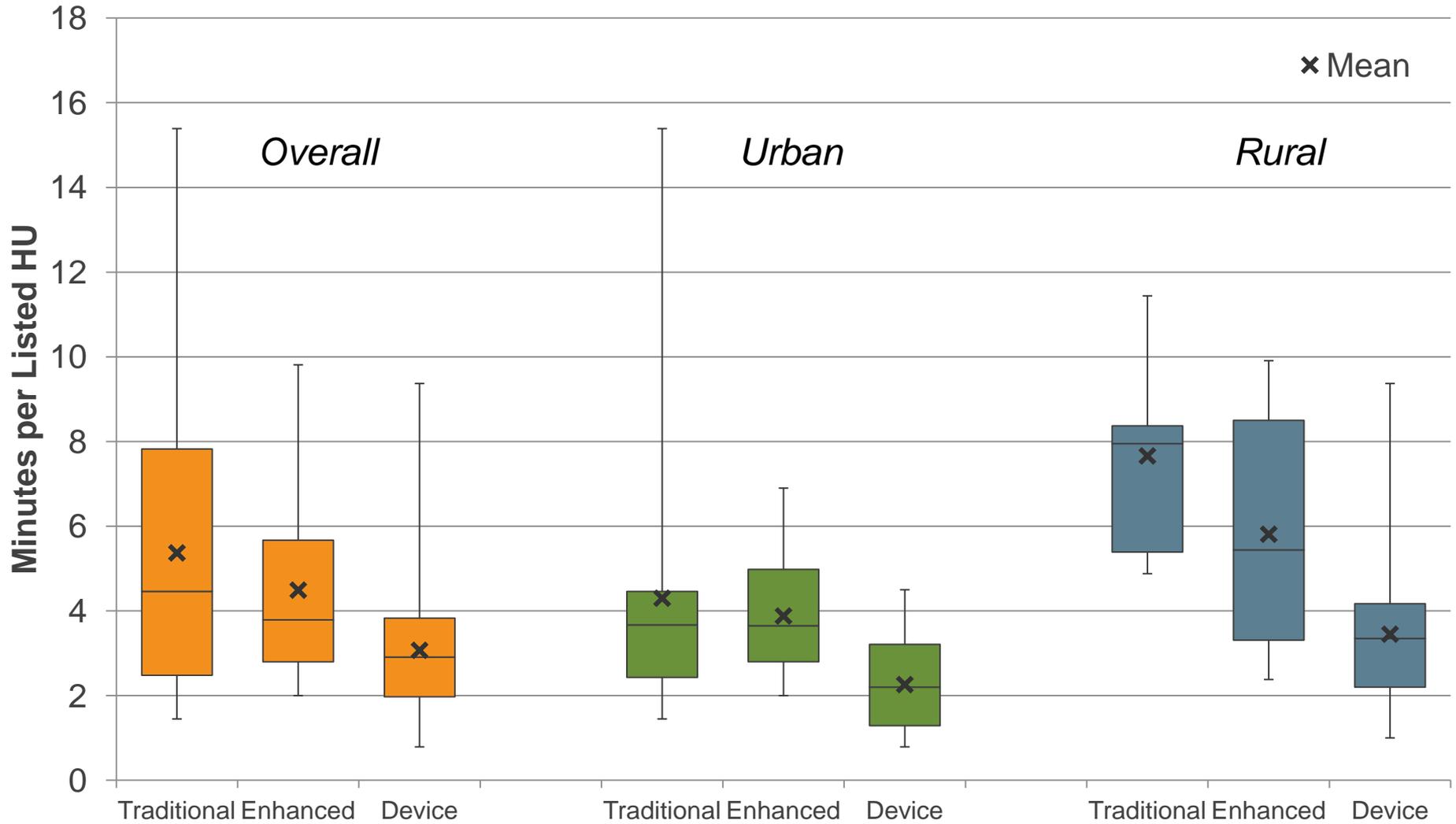➢ Post-hoc national comparison using hand-helds

Enhanced Listing: Example on Hand-Held

- Pre-loaded addresses are displayed, searchable
- Confirm, edit, delete
- Photos, description, longitude/latitude
- Synchronization with NORC server

Listing Minutes per Housing Unit (HU)

Potential Future Approaches

- ## Specialized lists
  - *InfoUSA, MSG, Valassis, Etc.*
  - Expect lots of over-coverage

- ## Air photo interpretation
  - Human, machine
  - Conducted evaluation (Curtis and English 2012)
    - Matched housing units from air-photos to listings
      - **Aerial listing missed 13% of listings**
      - **Listing missed 11% of aerial review**
      - **No validation of who right**
    - Real cost savings: aerial listing 25% time and 10% costs

Potential Future Approaches Contd.

- Issues:
  - Cloud, tree cover
  - Determining Hus
  - Multi unit buildings?

Discussion and Conclusions

- CDSF suitable in urban areas, many non-urban
  - Need some augmentation in rural areas
- Rural often have some CDSF addresses as starting point
  - Argue enhanced listing in instances CDSF not suitable alone
    - Improves coverage in all environments
    - Lower or equivalent relative cost
- Technology adds considerable efficiency, utility
  - Least-costly listing
  - Collect photo, GPS coordinates
- CDSF quality predictable
  - Segment
  - Sub-segment (block)
- Should implement "surgical" approach to frame construction

Ned English, english-ned@norc.org
Katie Dekker, dekker-katie@norc.org

Thank You!

Evaluation of Methods Contd. (consolidate onto one)

Question 1: Where is the DSF alone most suited (B ∩ U)

Variable	Sign
Ratio DSF/Census***	+
% HU Urban ***	+
% HU TEA 1 ***	+
HU density ***	+
% HU occupancy ***	+
Segment area ***	+
% White non-Latino ***	-
Median Household Income ***	+
HU increase since 2000*	+
Multi unit buildings+	+

- Measures of urbanicity, geocoding, DSF consistency

Evaluation of Methods Contd.

Question 2: Where does Enhanced Listing have better coverage properties than Traditional Listing

Variable	Sign
Ratio DSF/Census***	+
Segment Area***	+
Median Household Income+	+

- Measures of urbanicity, geocoding

Evaluation of Methods Contd.

Question 3- Where E listing adds more missed units than T

Variable	Sign
% HU Urban**	+
% HU TEA 1 ***	+
HU density***	+
% HU occupied***	+
Segment area**	+
Block Count***	+
% White non-Latino***	+
Ratio DSF/Census***	+

- E adds more in urban areas

Segment Count/Density

	# Segments	Mean Density
Overall Traditional	19	1,479 HU/square mile
Overall Enhanced	19	1,303 HU/square mile
Overall Device	144	309 HU/square mile
Urban Traditional	13	2,134 HU/square mile
Urban Enhanced	13	1,860 HU/square mile
Urban Device	46	835 HU/square mile
Rural Traditional	6	59 HU/square mile
Rural Enhanced	6	95 HU/square mile
Rural Device	98	69 HU/square mile

Background Contd: CDSF Evaluations

- DSF not a sampling frame by design
    - Requires processing
    - Is organized by Postal geographies
        - ZIP code
        - ZIP + 4
        - Carrier Route
        - Walk Sequence
    - Geocoding to associate with Census areas
        - Tract
        - Block Group
        - Block
    - One decides which addresses to include
        - Vacant, seasonal, college
        - PO BOX, RR BOX

Evaluation of Methods Contd.

- Listings were then independently checked
  - "Frame checking" for gold standard
- Have address frames:
  - U- USPS DSF list
  - T- Traditionally-listed frame
  - E- Enhanced frame
  - B- "Best" frame, representing reality
- Can quantify performance by:
  - B ∩ U
  - B ∩ T
  - B ∩ E
- Logistic regression for prediction