Strategies for Increasing Geocoding Accuracy

David Rust, MS
Kentucky Cancer Registry
Overview

• Importance of Geocoding
• Geocoding at the Kentucky Cancer Registry
• Geocoding Difficulties
• KCR’s Initial Geocoding Attempt
• Strategies to Improve Geocoding Practices
• Results
Importance of Geocoding

• KCR geocodes patient addresses at diagnosis
  – Determine patterns and trends important to cancer community
  – Exploring Environmental Exposure

– KCR Examples:
  • Appalachian residents for D.O.D. Lung Study
  • City vs. Well Water Lines
Geocoding Difficulties

• Kentucky is a rural state
  – Rural Routes and PO Boxes instead of residential addresses

• Typographical errors and misspellings

• Informal place names
  – Informal: Las Vegas Strip or “The Strip”
  – Formal: South Las Vegas Boulevard
Complication of a Rural State

• Street-Level Geocoding is desired

• Kentucky’s rural nature makes geocoding patient addresses difficult
  – Rural Routes and PO Boxes do not provide enough information for street-level geocoding
Rural vs Urban Kentucky Counties

• Only 5 of Kentucky’s 120 counties are considered urban counties

• Urban county
  – 75% population is urban in 2000 census
  – Otherwise, the county is considered rural
Envinsa Geocoding Outcomes

• **~30,000** Patient Addresses
  – 30% Urban
  – 70% Rural

• **8449** Street Level Urban Addresses
  – 94% of Urban Total
  – Urban County Goal >= 90%

• **14760** Street Level Rural Addresses
  – 70.2% of Rural Total
  – Rural County Goal >= 80%
Kentucky Voter Data

• Similar to the Atlanta SEER Registry, the KCR obtained voter registration files that include both mailing and residential addresses
• Using Link Plus, a CDC record linkage software, the KCR matched the mailing addresses of KY voters and KY cancer patients
• This mapped PO Box mailing addresses to a residential street
Kentucky Voter Data & Link Plus

- **6265** Total Rural Addresses not geocoded to Street-Level

- Using Link Plus to map these rural address to Voter data
  - **1072** or ~17% rural mailing addresses were linked to KY Voter mailing addresses
Linked Results + Envinsa Geocoding

- **553** addresses geocoded to street-level
- Now there are **15,313** Rural Address geocoded to street-level
- + **2.6%** increase in Rural County geocoding
Melissa Data & Address Scrubbing

• Recommended by Atlanta SEER Registry
• Used an address scrubbing technique
  – Corrects street addresses that may have typographical errors and misspellings
• Transformed each address to the USPS standard
• Melissa Data was able to geolocate 300 additional Rural addresses to street level
• +1.5% increase in Rural County geocoding
Informal Place Name Database

- Replaced informal place names
- Geocoded addresses according to formal locale name
- 15 more street-level rural addresses
- \(<1\%\) increase in Rural County Addresses
Results

• Kentucky Voter Data
  +2.6% in Rural Counties

• Melissa Data & Address Scrubbing
  + 300 records = +1.5%

• Informal Place Name Database
  < 1%

• New Rural Street-Level Geocoded Addresses
  
  – 70.2% +2.6% + 1.5% + <1% = ~74%
Questions?