

# Using the NAACCR Geocoder Metadata to Assess Quality of Geocoded Data and Impact on Results of Spatial Studies

Recinda Sherman  
NAACCR Annual Meeting,  
Wed June 25, 2014

# Overview

---

- NAACCR geocoder
- Geocoding quality
- K-nearest tool
- Using the NAACCR Geocoder Metadata to Assess Quality of Geocoded Data ~~and Impact on Results of Spatial Studies~~

# NAACCR geocoder

---

- Free
- Transparent
- NAACCR quality codes
- Quality metadata
- Accessed through MyNAACCR
- High quality underlying data
  - Parcels, NAVTEQ
- Uniform geocoding

# Geocoded data quality

- Census data
  - Scale, sample, time, availability
- Geocoding
  - Missing, incorrect, incorrect
- Misclassification bias
- Reduce power
- Cartographic confounding

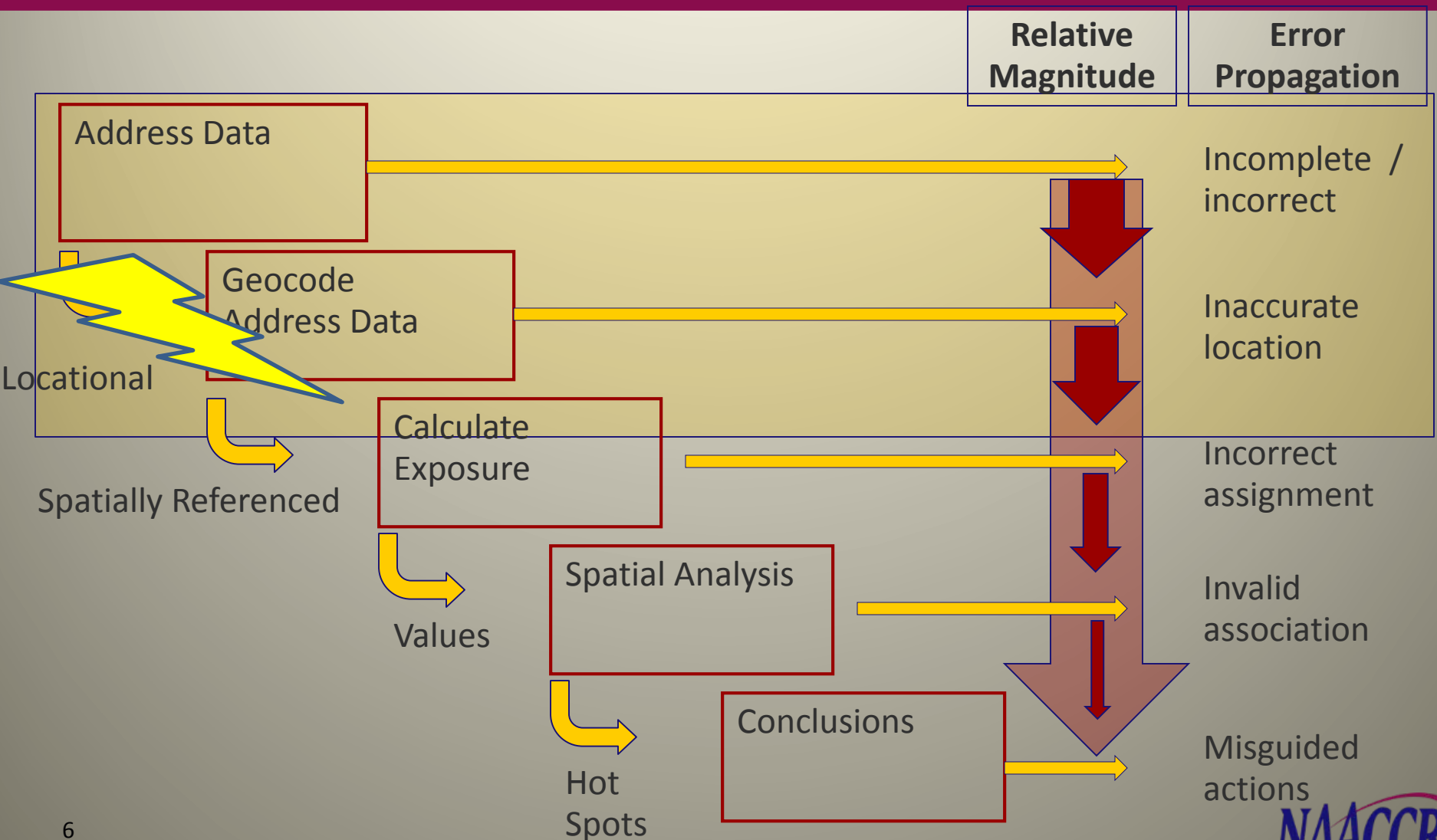


# Geocoding quality components

- Match
  - Match rate
    - Coverage, % with spatial location
  - Quality of match
    - Parcel match versus interpolation along a street segment
    - GIS Coordinate quality NAACCR Item #366
- Precision
  - Scale
    - County center versus census block
    - Census Tract Certainty
    - NAACCR Items #364,#365,#367
- Accuracy
  - Correct location

# Geocoding error

- Error introduction/propagation



# Imprecise geocodes

Census tract certainty >

Zipcode level matches

Often removed from analysis

PO Box versus residential

Cartographic confounding

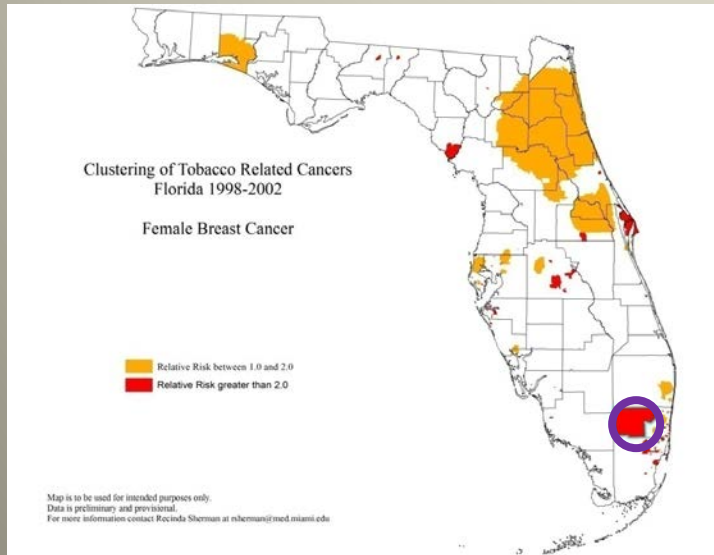
Rural, older, institutionalized

1	Census tract based on complete and valid street address of residence
2	Census tract based on residence ZIP + 4
3	Census tract based on residence ZIP + 2
4	Census tract based on residence ZIP code only
5	Census tract based on ZIP code of P.O. Box
6	Census tract/BNA based on residence city where city has only one census tract, or based on residence ZIP code where ZIP code has only one census tract
9	<del>Not assigned, geocoding attempted</del>
Blank	<del>Not assigned, geocoding not attempted</del>

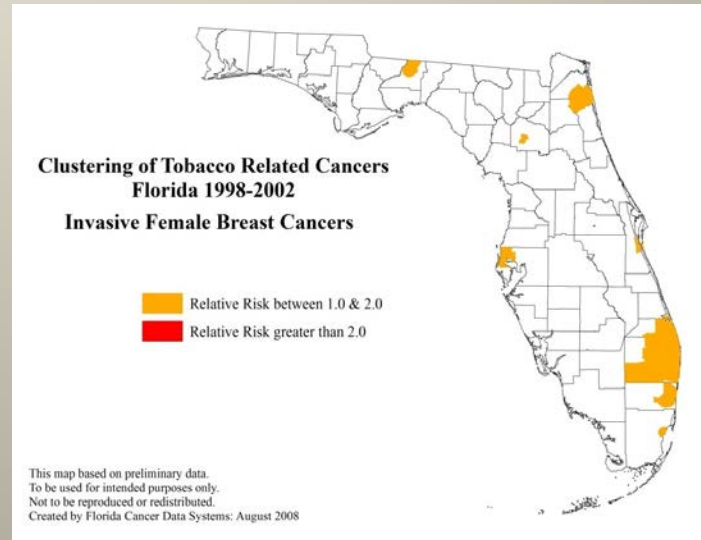


# Specificity ?

## Old Data:



## Improved Data:





# Validity ?

## Old Data :

### Oral Cancer by SES

- Wealthy
  - 34.0 ref
- Mid High
  - 36.6 RR 1.08
- Mid Low
  - 39.1 RR 1.15
- Poorest
  - 46.3 RR 1.36

## New Data :

### Oral Cancer by SES

- Wealthy
  - 37.3 ref
- Mid High
  - 40.1 RR 1.08
- Mid Low
  - 45.4 RR 1.22
- Poorest
  - 49.2 RR 1.32

# Methods

---

- Florida data
  - Geocoding
    - 95% street level, 3% PO Box, 2% Zip code level residence
    - Proprietary vendor, no manual follow-up
- Breast cancer cases, 2007-2011
  - NAACCR geocoder
  - 6% Census Tract Certainty 4
  - Zip code centroid based on residential address
  - n=5,732
- K-nearest Tool
  - 5 nearest census tracts

# Preliminary results

---

- “nearest” 100% match
- 1-4 nearest
- Long/lat
- Distance (meters)
- Census tract
- **Missing county**
  - Census tracts are only unique when combined with county (state)
  - Results may be conservative

# Preliminary results

---

- Census tract-level rates
  - Cluster detection, cancer “cluster” investigation
- Population size
  - Ratio,  $< 1$  selected census tract larger
  - Zero pops excluded
- Nearest
  - Mean change ratio 1.18
- Average of 4 nearest
  - Mean change ratio 4.68

# Results tract-level poverty

		1st Nearest				
		nomatch	lowest	2	3	highest
Selected	lowest		25%	47%	20%	6%
	2	1%	25%	29%	34%	10%
	3	3%	10%	27%	38%	23%
	highest	4%	3%	10%	34%	49%

# Results tract-level poverty

		4th Nearest				
		nomatch	lowest	2	3	highest
Selected	lowest	2%	25%	32%	34%	4%
	2	5%	17%	29%	32%	15%
	3	12%	9%	23%	35%	16%
	highest	5%	7%	18%	29%	40%

# Results tract-level poverty

		1st Nearest--Blacks				
		nomatch	lowest	2	3	highest
Selected	lowest		22%	52%	19%	7%
	2		23%	18%	52%	7%
	3	3%	3%	24%	29%	39%
	highest	1%	3%	7%	26%	63%



# Issues

---

- Beta mode
- Speed issues
- County
- “nearest” vs contiguous

# Future Directions

---

- Fix county/Reassess
- Age-specific population
- Race/ethnicity specific population
- Aggregate census tracts for analysis of imprecise cases
  - Incorporate matched zipcode area
  - Add variance for distance studies
- Guidelines, thresholds for case inclusion

# Questions?

---

Recinda Sherman

[rsherman@naaccr.org](mailto:rsherman@naaccr.org)

217-698-0800 x 6