Geocoding and selection bias in epidemiologic research using GIS

M. Norman Oliver, M.D., M.A.
Associate Professor,
Departments of Family Medicine,
Public Health Sciences,
and Anthropology
University of Virginia Health System

Acknowledgements

- Kevin A. Matthews, M.S.
- Mir Siadaty, M.D., M.S.
- Fern R. Hauck, M.D., M.S.
- Linda W. Pickle, Ph.D.

- NCI K07 CA099983; HRSA CFDA No. 93.984, Academic Units in Primary Care-Family Medicine

Place matters

- Family medicine, public health, and epidemiological researchers using GIS to assess association between population health and area characteristics
  -- surveillance, cluster analysis, exposure, measured & unmeasured factors affecting disease.

Locating the place

- Initial task: assign geographic location to study subjects – geocoding
- Completeness varies
  -- positional accuracy
  -- differential match rates by region
- Incomplete geocoding can lead to biased results

Selection bias

- Differential match rates by geographic region can lead to biased results owing to unrepresentative data and a consequent selection bias
- Non-random missingness: social, economic, political, other reasons
- Place matters, and social determinants may be confounded with place

Prostate cancer in Virginia, 1990-99

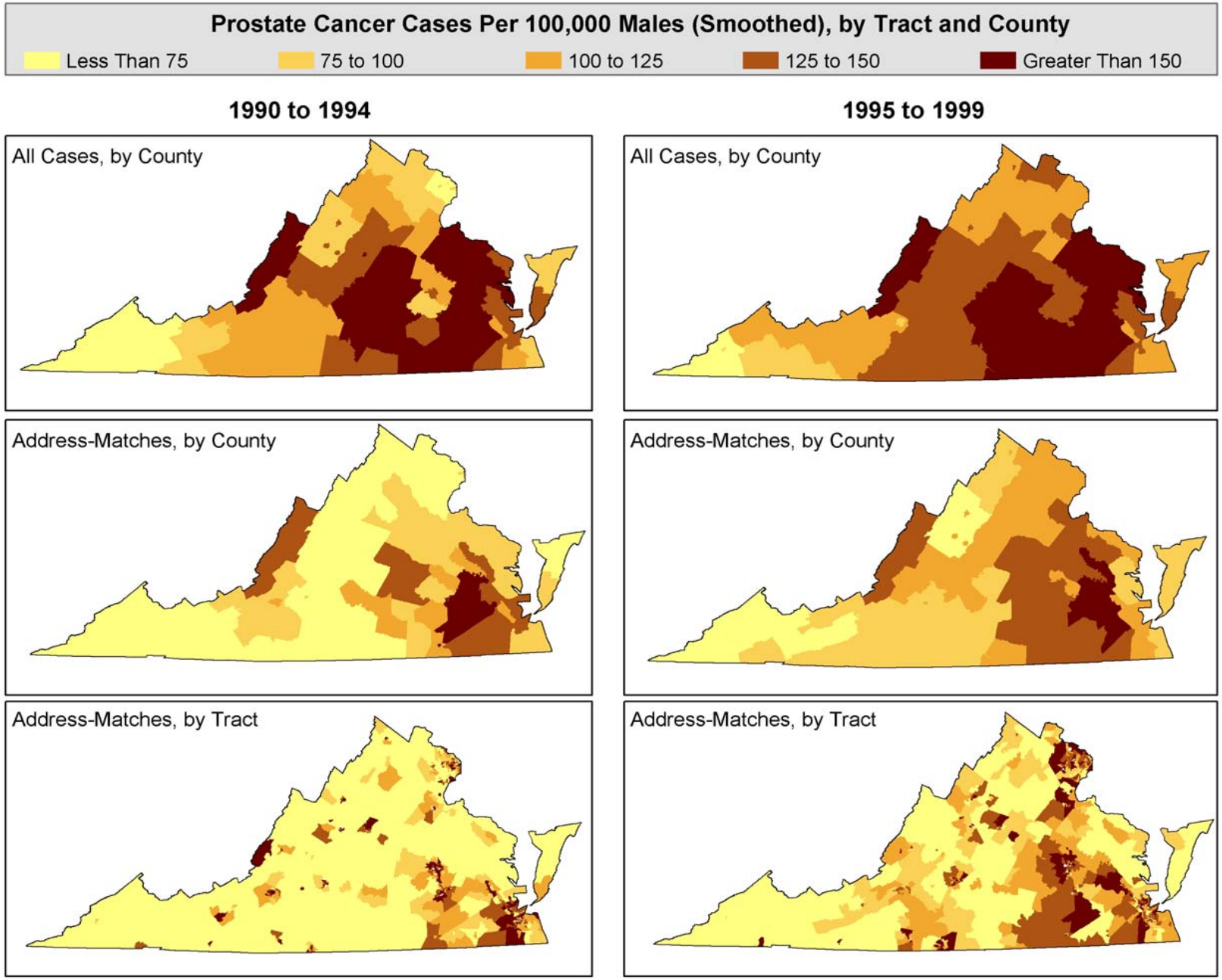
- Study of CaP incidence, assessing association of age, racial category, and area-level measures of SES with this outcome
- Positive assoc btw CaP incidence and income, urban status (all)
- Negative assoc btw CaP incidence and poverty, low educ (whites only)
- These effects seen only at the census-tract level
- MAUP?

Table 1. Census Tract geocoding results broken down by address type.

			% Of address types			
	No.	**%**	**Street Addresses**	**Rural Routes[a]**	**P.O. Boxes[a]**	**Other[a,b]**
Matched[c]						
African American	6,060	74.0	92.7	0.0	0.0	0.0
White	20,278	73.4	92.6	0.0	0.0	0.0
Unmatched[c]						
African American	2,192	26.0	7.3	100.0	100.0	100.0
White	7,136	26.6	7.4	100.0	100.0	100.0

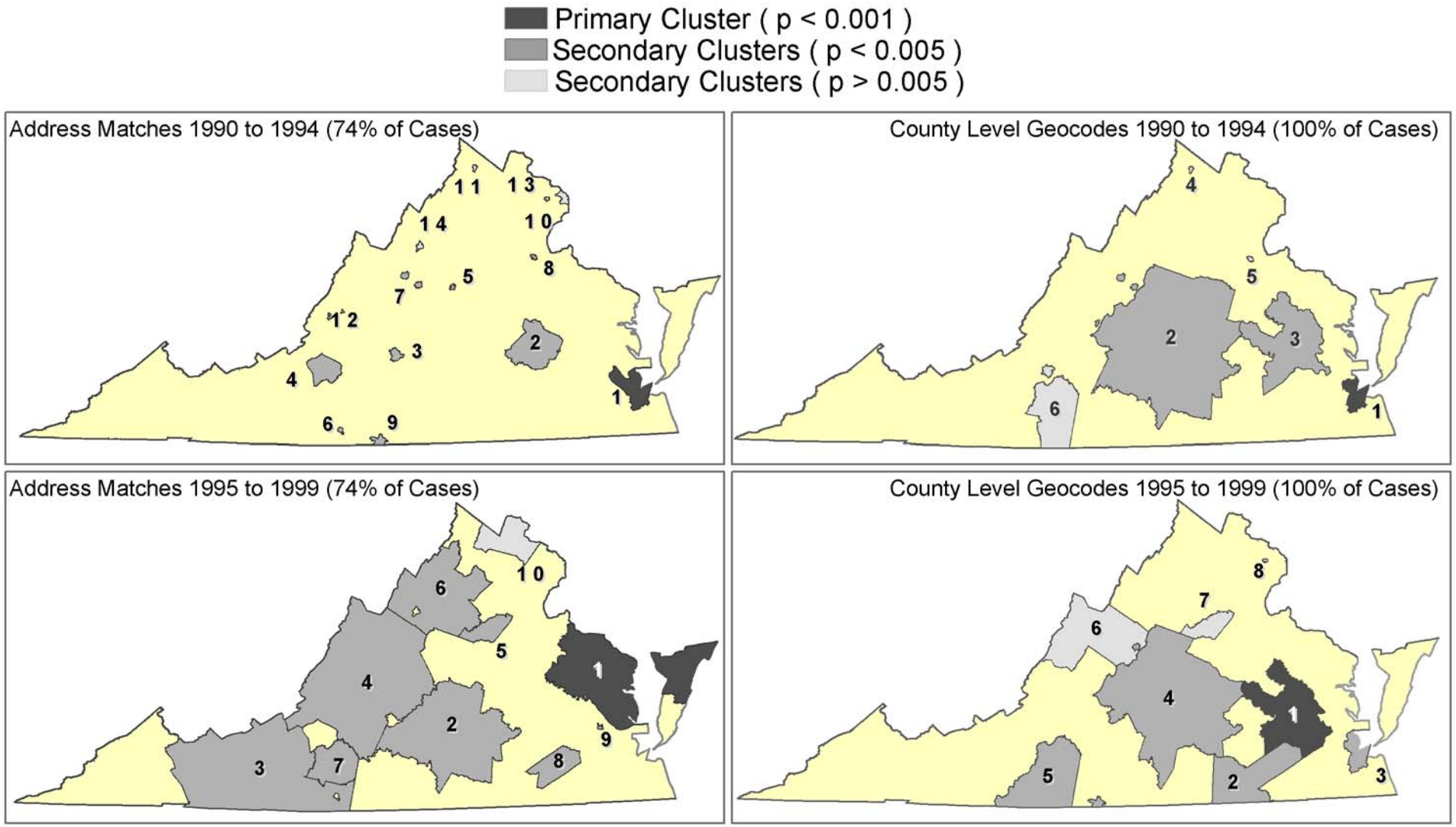
[a]Accurate geocoding to the Census tract cannot be performed on this address type.
[b]Includes garbled and incomplete addresses.
[c]To the Census Tract.

Figure 1. Annualized age-adjusted prostate cancer incidence, Virginia 1990 - 99



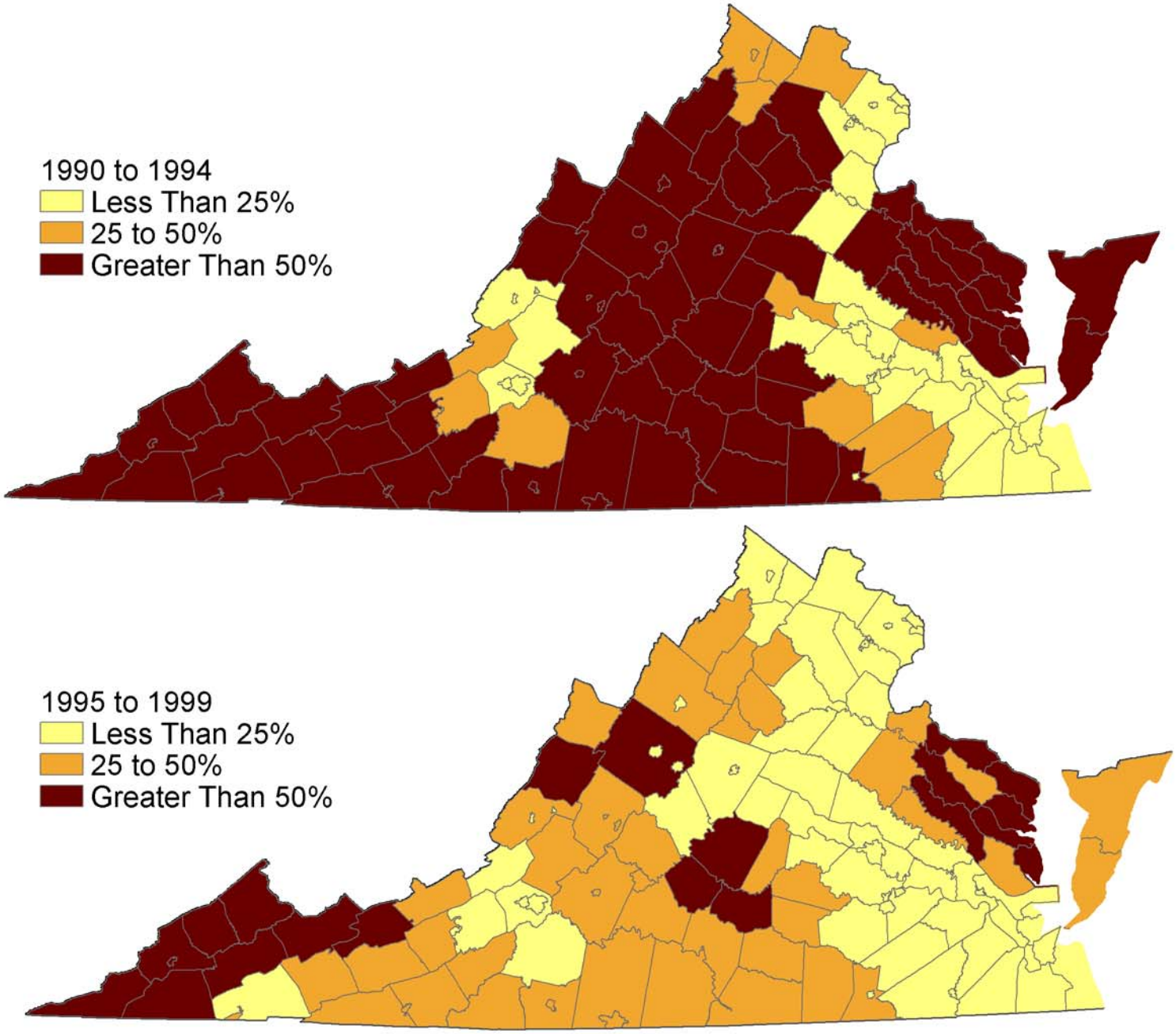
Source: 1990 to 1999 Virginia Cancer Registry and 1990 Population Census

Figure 2. Prostate cancer incidence clusters, Virginia 1990 - 99



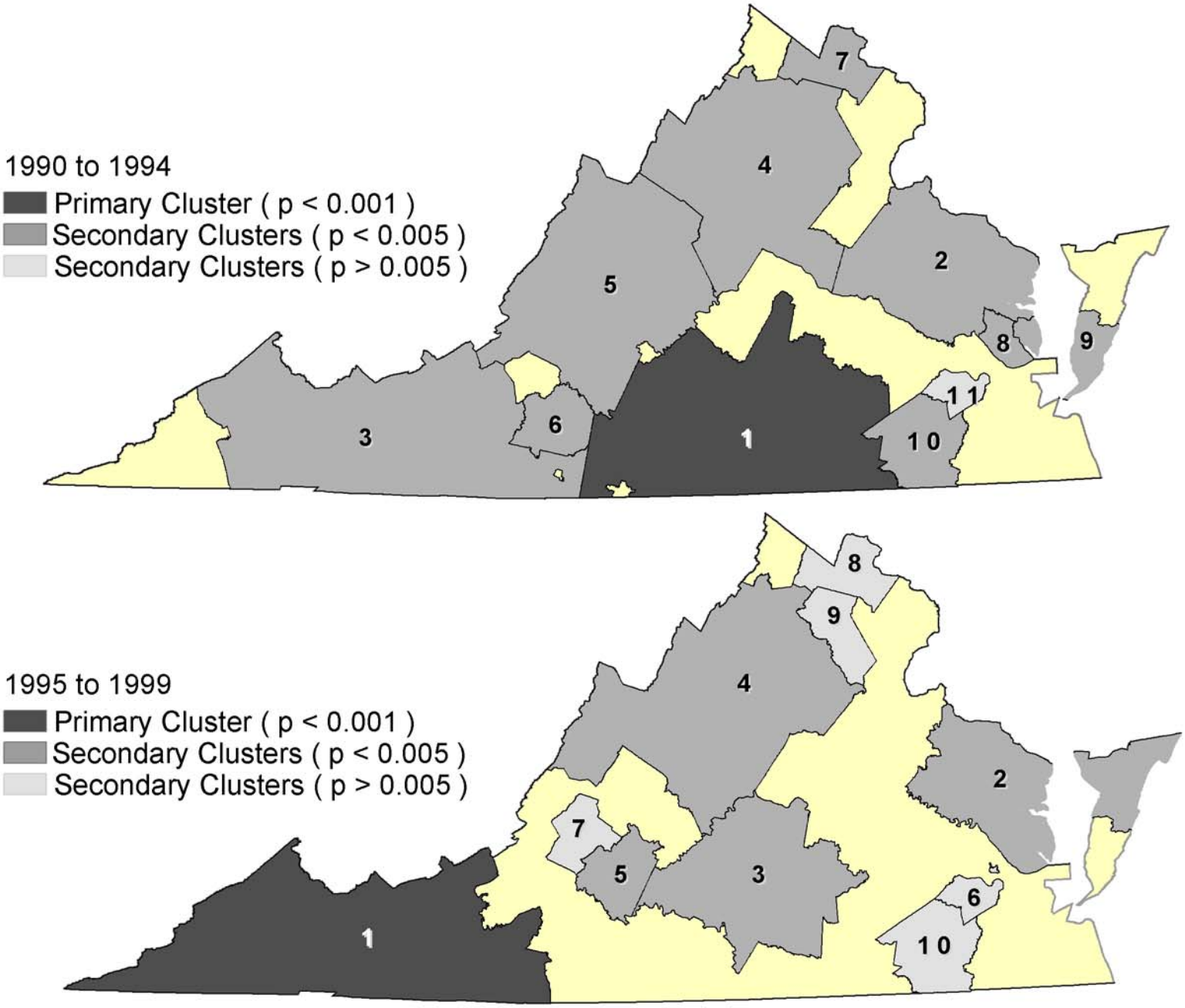
Source: 1990 to 1999 Virginia Cancer Registry and 1990 Population Census

Figure 3. Proportion of unmatched prostate cancer cases



Source: 1990 to 1999 Virginia Cancer Registry and 1990 Population Census

Figure 4. Clusters by proportion of unmatched prostate cancer cases



Source: 1990 to 1999 Virginia Cancer Registry and 1990 Population Census

Geographic patterns: Are they real?

- GIS used to identify geographic patterning
- In our VCR study, spatial patterns may be reflection of data distribution rather than underlying disease patterns
- Cluster analysis of proportion of missing cases shows significantly different patterns resulting from non-random differences in geocoding completeness

Cartographic confounding

- Classic epi: measure of the effect of one factor on disease risk biased because of its assoc with another factor (confounder) *and* the disease
- *Similarly, when the factor of interest is* ***geographic,*** *a factor related to the disease that is not distributed randomly across the study area can confound the appearance of maps of that disease.*

Location, location, location

- Systematically missing data resulting from location
- However, location's sociodemographic characteristics associated with likelihood of missing data from that location
- As well as location being associated with likelihood of disease in that area.
- Spatial disease patterns – the look of the map – may confound location with social determinants of disease
- Standard methods of dealing with this challenge are not enough – ignore geography.
  -- case ascertainment (90%), multivariate analysis.
- Must assess geographically – e.g., cluster analysis
- Iterative process of statistical and spatial analyses

Geocoding and selection bias in epidemiologic research using GIS

M. Norman Oliver, M.D., M.A.
Associate Professor,
Departments of Family Medicine,
Public Health Sciences,
and Anthropology
University of Virginia Health System