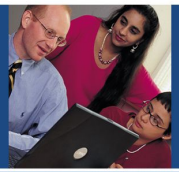
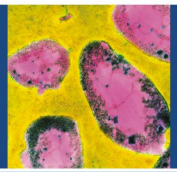
JOHNS HOPKINS
UNIVERSITY

JOHNS HOPKINS BLOOMBERG SCHOOL *of* PUBLIC HEALTH





Protecting Health, Saving Lives—*Millions at a Time*

JOHNS HOPKINS BLOOMBERG SCHOOL of PUBLIC HEALTH

County Lung Cancer Mortality: A Decision Tree Model for Control and Prevention

Norma Kanarek, PhD, Johns Hopkins School of Public Health, Department of Environmental Health Sciences
Shu-Chih Su, MS, JHSPH, Department of Biostatistics
Melissa Brower, MPH, Bassett Healthcare
Haomiao Jia, PhD, Mercer University School of Medicine
Brian Fitzek, BA, JHSPH Department of EHS

Specific Aims

- Develop or aggregate county-level data pertinent to lung cancer prevention and control
- Use a classification and regression tree (CART) approach to illuminate county groupings
- Extend the use of CART for policy making, program planning and research

County-Level Indicators

- **Lung cancer** mortality (1999-2001, NCHS)
- **Smoking** prevalence (2000, BRFSS; Jia)
- Risk of **excess cancer death** estimates (National Air Toxics Assessment, 1996)
- Population **size** and **density, urban/rural status**, and **poverty** (2000, US Census)
- EPA/DHHS **region**
- Average **indoor radon** level > 4 pCi/L (EPA)
- **Coronary Heart Disease mortality** (1988-90, NCHS)

Classification and Regression Tree

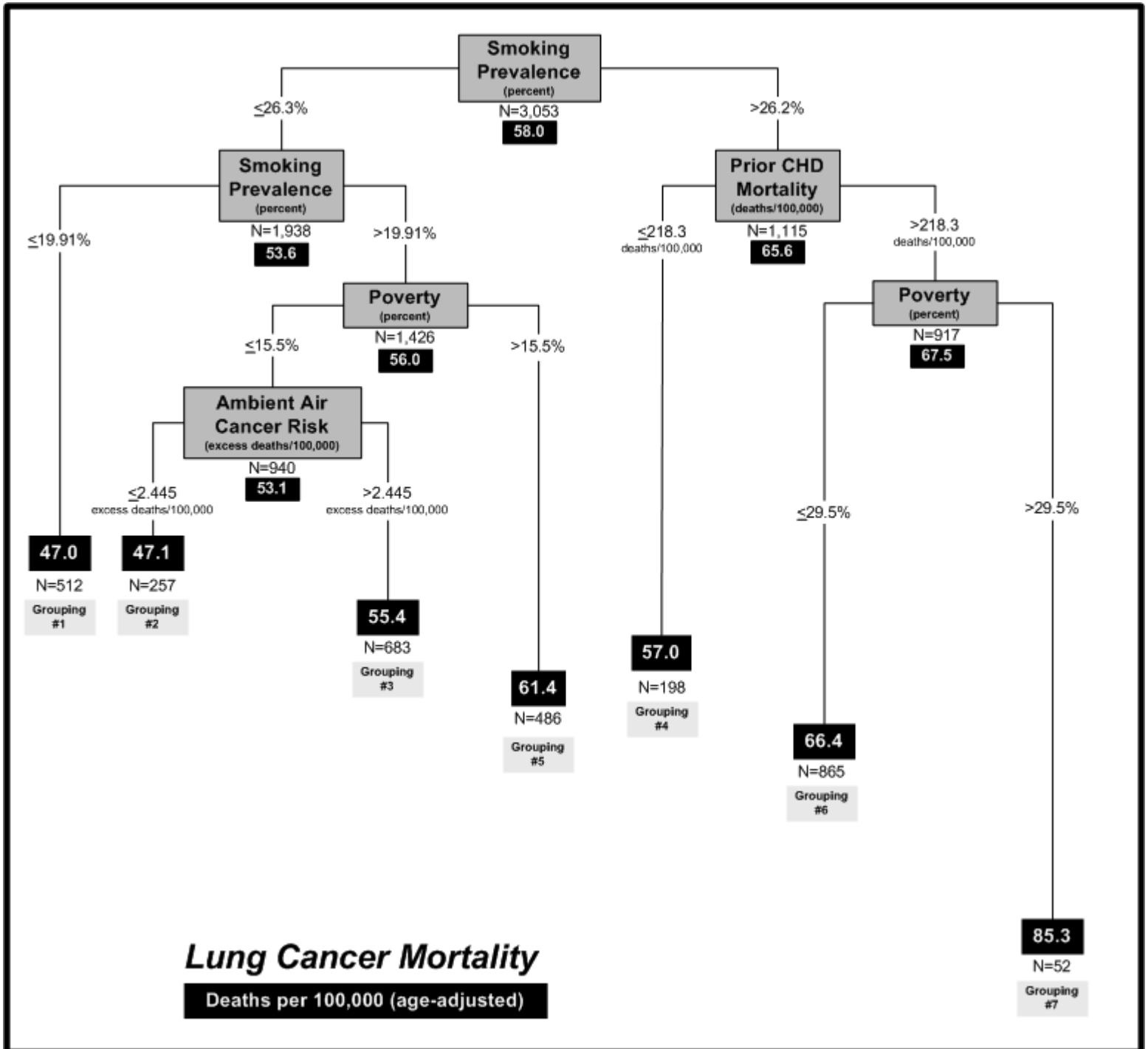
- Software, R Stat
- Two choices of tree building:
  - regression (maximizes the prediction) or
  - classification (maximizes sensitivity and specificity)
- Process:
  - grow a full tree (all significant leaves)
  - prune the tree (optimal number of leaves, balancing accuracy with complexity)

Questions that can be addressed using CART

- At this time, what is the local baseline lung cancer mortality in the US?
- Which counties need tailored messages to a low education/income population?
- How many communities or people are affected in each grouping?

Questions that can be answered using CART (continued)

- What groupings predominate in my state?
- Is there a geographic cluster of counties with high lung cancer mortality?
- What is the attributable risk percent for poor air quality?


Smoking Prevalence (percent)
N=3,053
58.0
≤26.3%
>26.2%
Smoking Prevalence (percent)
N=1,938
53.6
≤19.91%
>19.91%
Prior CHD Mortality (deaths/100,000)
N=1,115
65.6
≤218.3 deaths/100,000
>218.3 deaths/100,000
Poverty (percent)
N=1,426
56.0
≤15.5%
>15.5%
Poverty (percent)
N=917
67.5
Ambient Air Cancer Risk (excess deaths/100,000)
N=940
53.1
≤2.445 excess deaths/100,000
>2.445 excess deaths/100,000
≤29.5%
>29.5%
47.0
N=512
Grouping #1
47.1
N=257
Grouping #2
55.4
N=683
Grouping #3
61.4
N=486
Grouping #5
57.0
N=198
Grouping #4
66.4
N=865
Grouping #6
85.3
N=52
Grouping #7
Lung Cancer Mortality
Deaths per 100,000 (age-adjusted)

Data and CART Issues

- Contiguous US
- Correlations affect the tree:
  - education level and low income
  - radon and smoking prevalence
- Broad range of lung cancer mortality within grouping

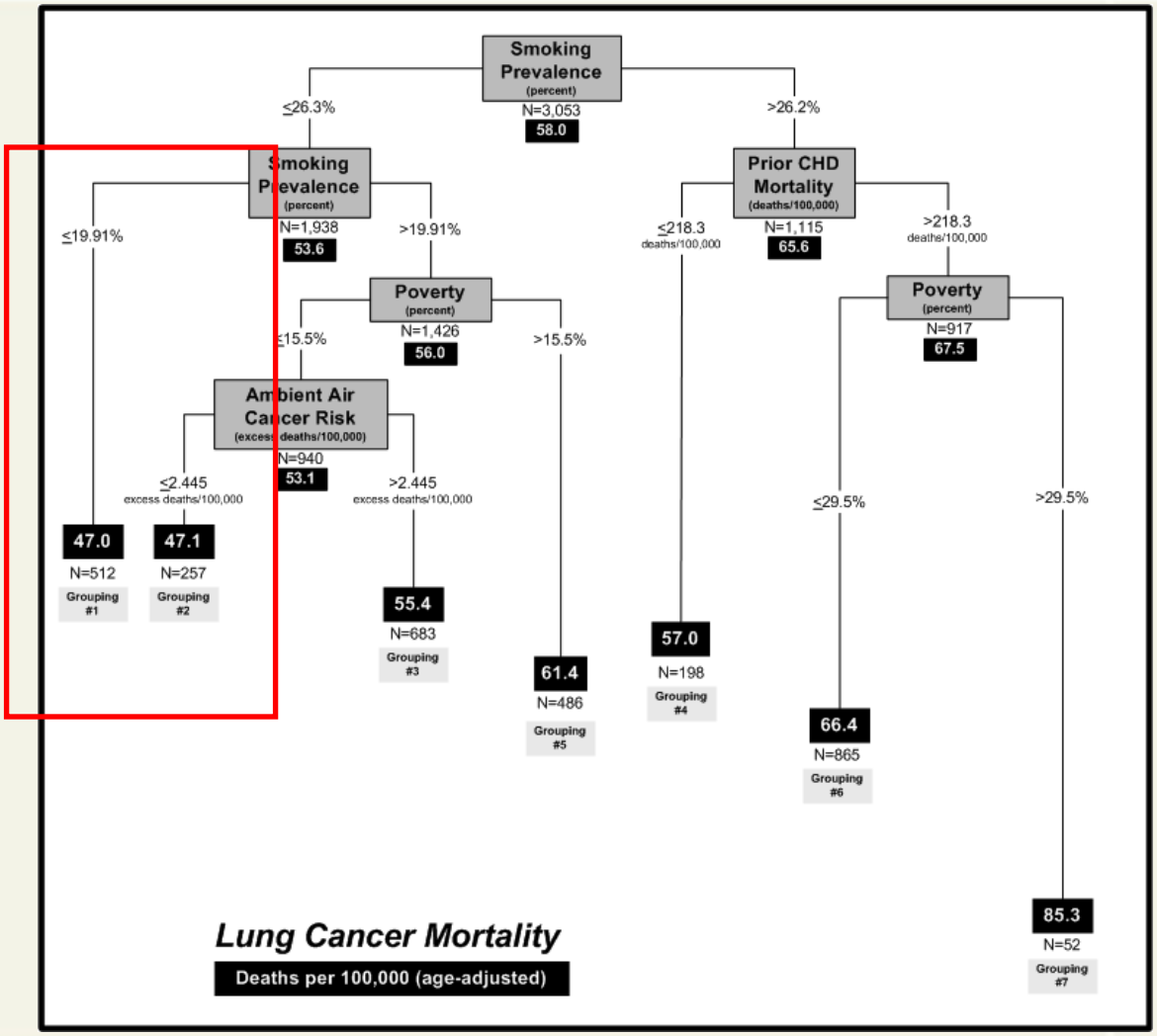
CART Advantages

- Graphic – “Pathways to Health”
- Exploratory and data-based
- Simplifying
- Continuous or categorical outcomes

Utility of CART

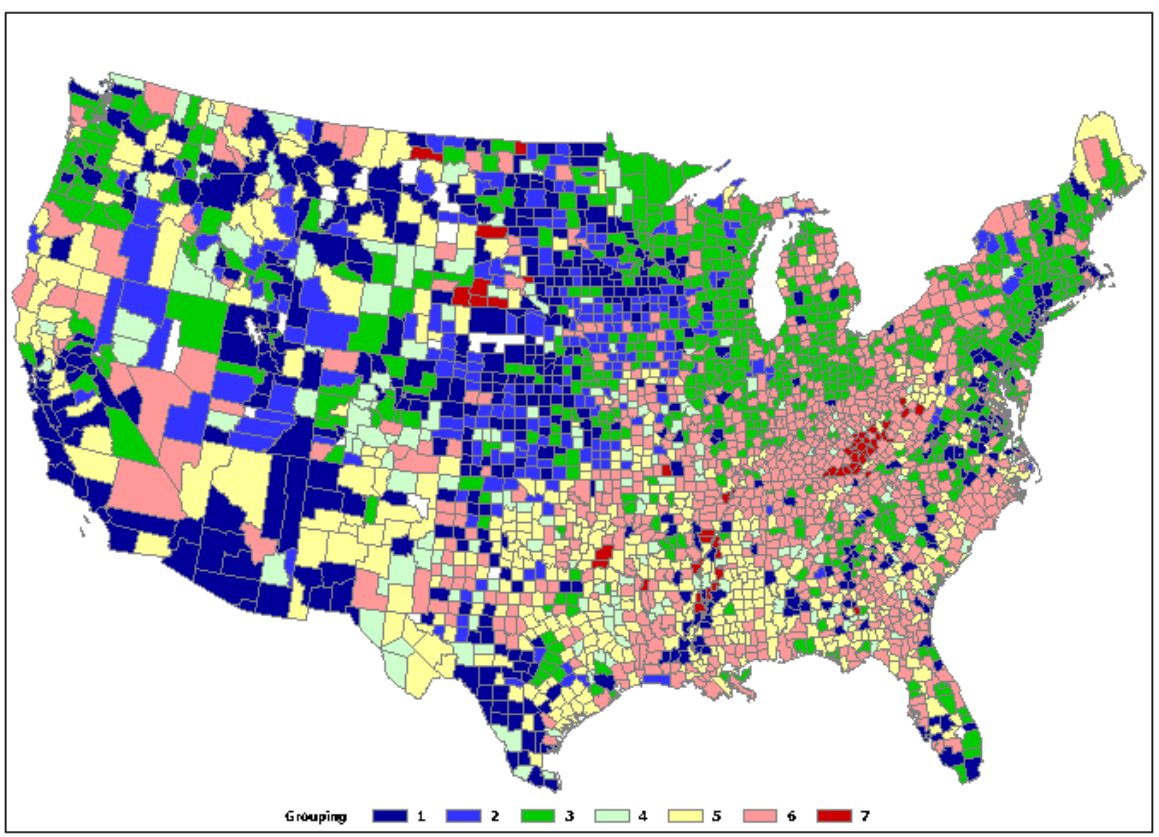
- Program management
- Policy making
- Research

At this time, what is the local baseline lung cancer mortality in the US?



Which counties need tailored messages to a low income/education population?

Groupings #5 and #7



What groupings predominate in my state?

Grouping #1

Carroll
Harford
Howard
Montgomery
Prince Georges

Grouping #2

Garrett

Grouping #3

Anne Arundel
Baltimore
Caroline
Charles
Frederick
Kent
Queen Anne's
St. Mary's
Talbot
Washington
Wicomico
Worcester

Grouping #4

Calvert

Grouping #5

Somerset

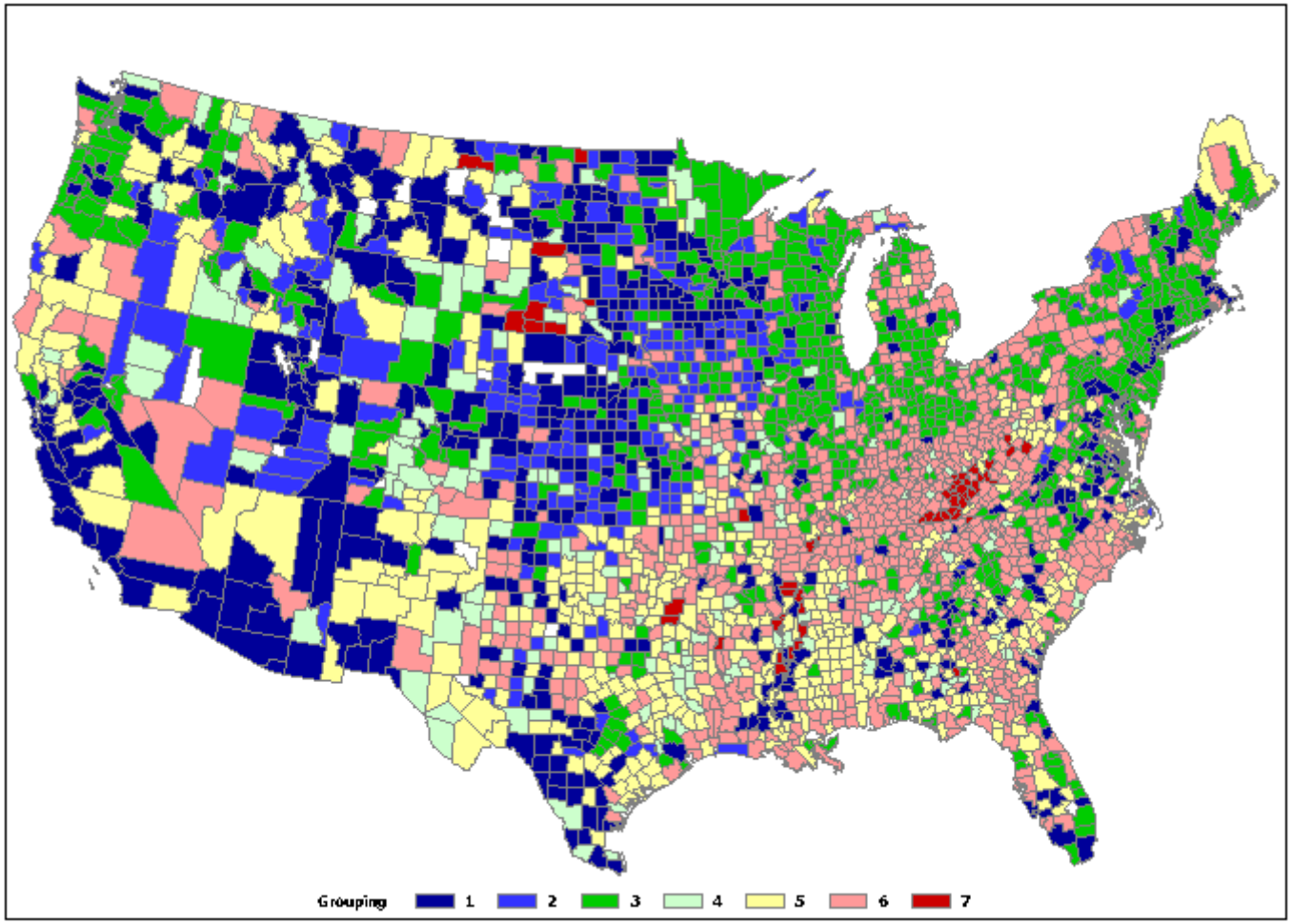
Grouping #6

Allegany
Baltimore City
Cecil
Dorchester

Grouping #7

The sample state, Maryland, has 23 counties and one city.
MD is statistically different from the nation ($p<0.05$).

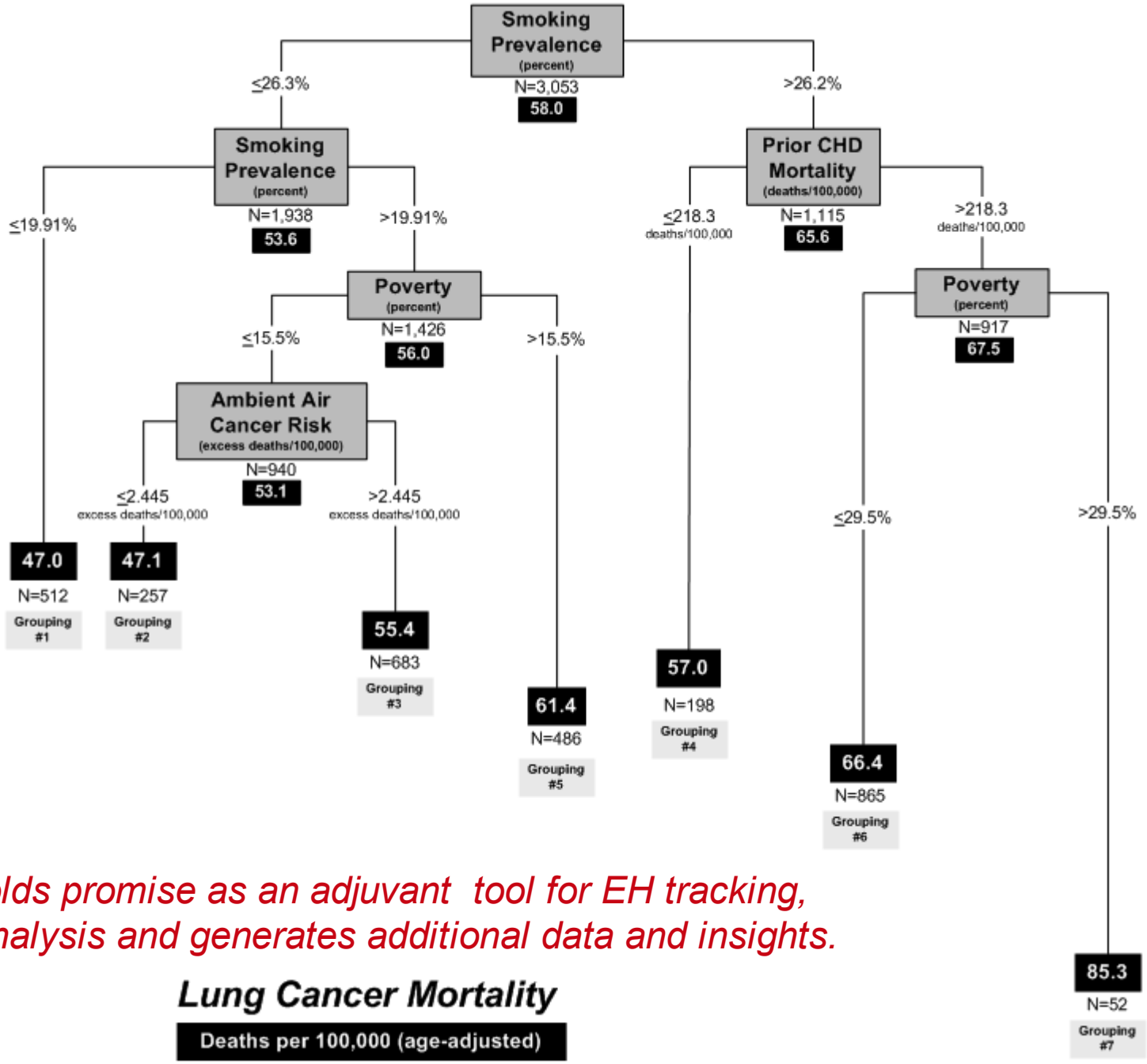
Is there geographic clustering of counties with high lung cancer mortality?



How many communities or people are affected in each grouping?

Grouping	#1	#2	#3	#4	#5	#6	#7
Lung Cancer Mortality (per 100,000)	47.0	47.1	55.4	57.0	61.4	66.4	85.3
Population Size (%)	90.1 m (33%)	2.9m (1%)	90.5m (32%)	6.6m (2%)	31.6m (12%)	56.6m (20%)	1.1m (<1%)
Number of Counties (%)	512 (17%)	257 (8%)	683 (22%)	198 (6%)	486 (16%)	865 (28%)	52 (2%)

What is the attributable risk percent for poor air quality?



CART holds promise as an adjuvant tool for EH tracking, guides analysis and generates additional data and insights.

Acknowledgements

Funding was provided in part by the Maryland Cigarette Restitution Fund and the Center for Excellence in Environmental Public Health Tracking at Johns Hopkins Bloomberg School of Public Health (Fellow).