

# Comparing Spatial Patterns: How much statistical power do I have?

Li Zhu

Surveillance Research Program

National Cancer Institute

Linda Pickle

StatNet Consulting

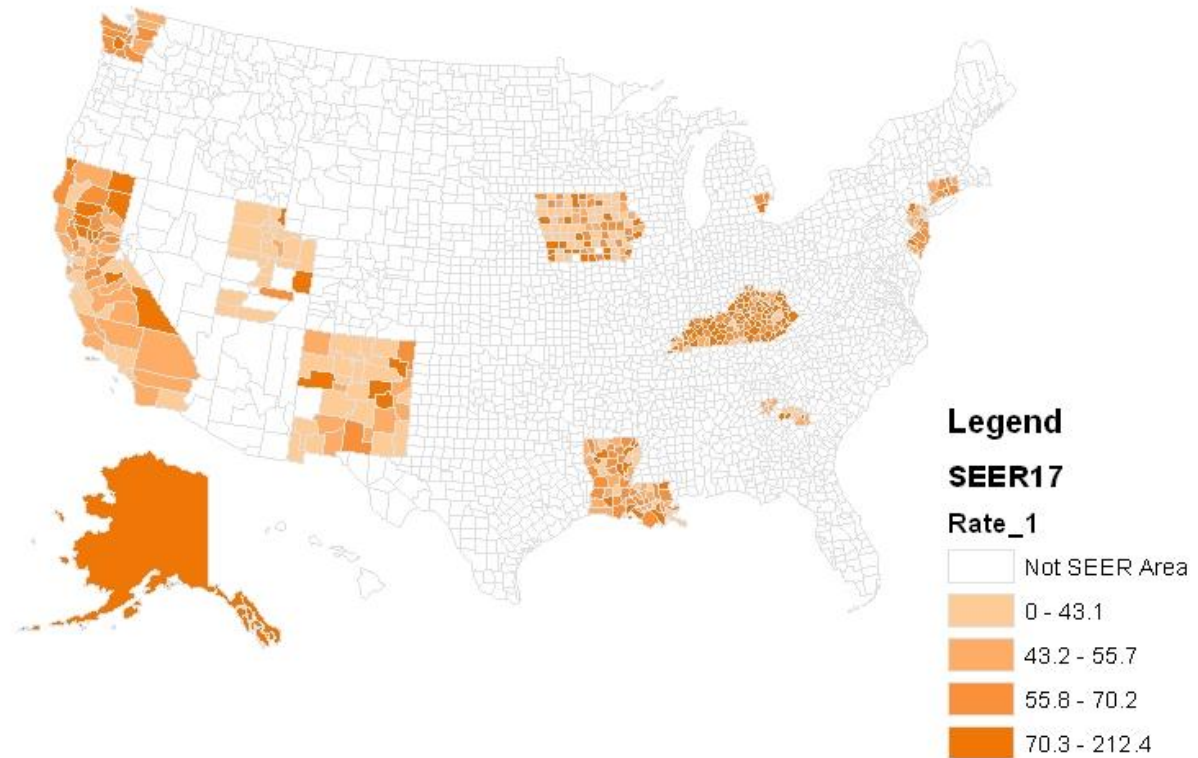
# Motivation

- New incidence data become available every year.
- 1973 – 2009 Data (Nov 2011 Submission)
- 2009 new cases: about 400K malignant cases

# Motivation

- Are there any differences
  - in the (temporal) trends of incidence and mortality?
  - in the (spatial) patterns of incidence and mortality?
- Are there any new hotspots in late-stage diagnosis?
- How to compare spatial patterns?

# County-level Female Lung Cancer Incidence Age-Adjusted Rates (per 100,000) in SEER17 Area, 2009



# Questions to Ask

- How different are the age-adjusted rates at the county level?
  - Compared to the average rates in the state
  - Compared to the average rates in the past (2000-2008)

# Absolute Percent Difference

$$Abs\_pct\_diff = \frac{|2009Rate - AverageRate|}{AverageRate} \times 100\%$$

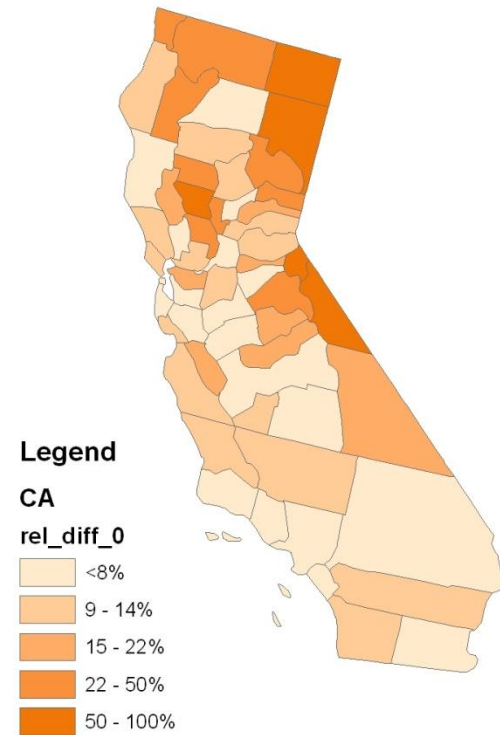
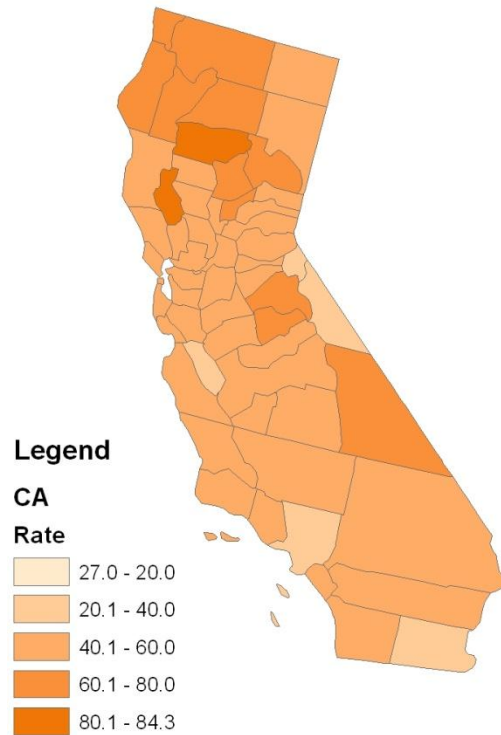
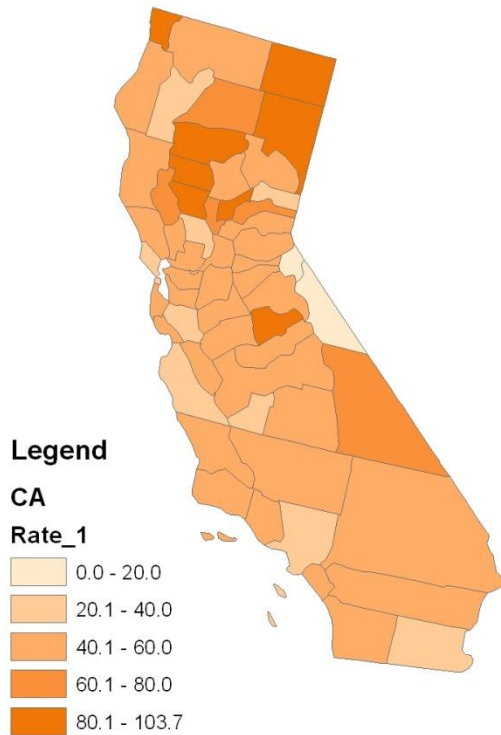
Value  $\geq 0$ , without upper limit  
Related to Rate Ratio

# Comparing County Rates to the Average Rates in the Past

2009

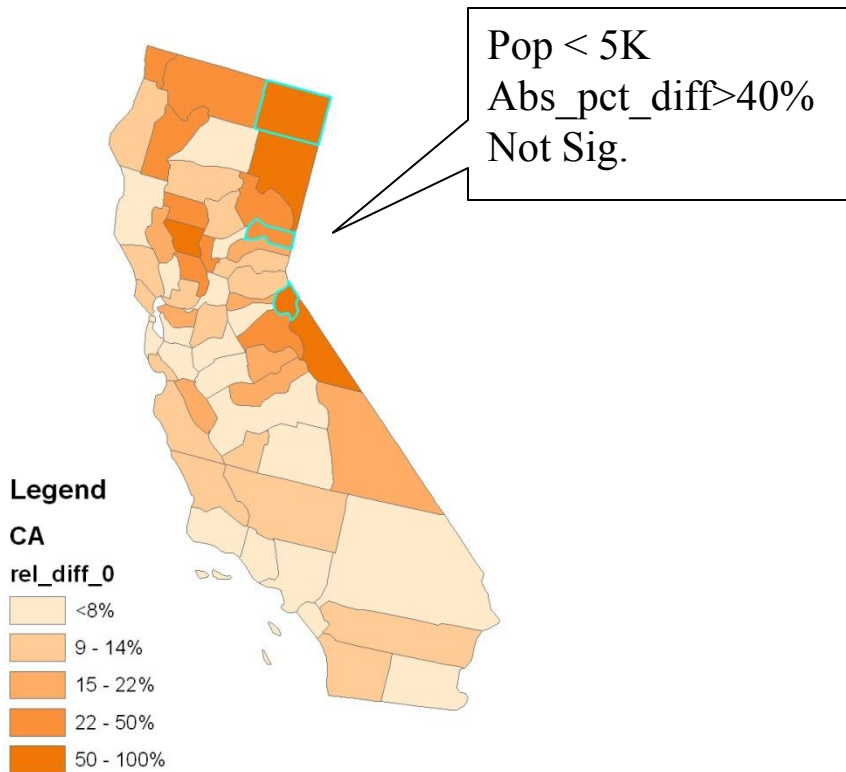
2000-2008

Absolute Percent  
Difference

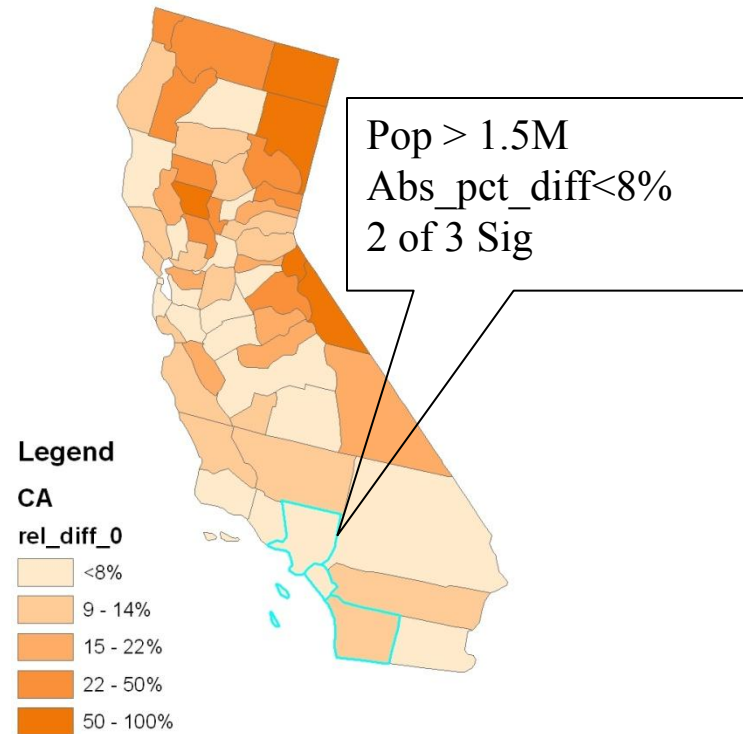


# Comparing County Rates to the Average Rates in the Past

Absolute Percent  
Difference



Absolute Percent  
Difference

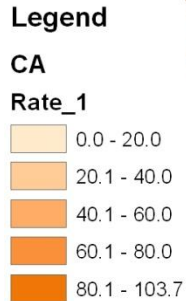
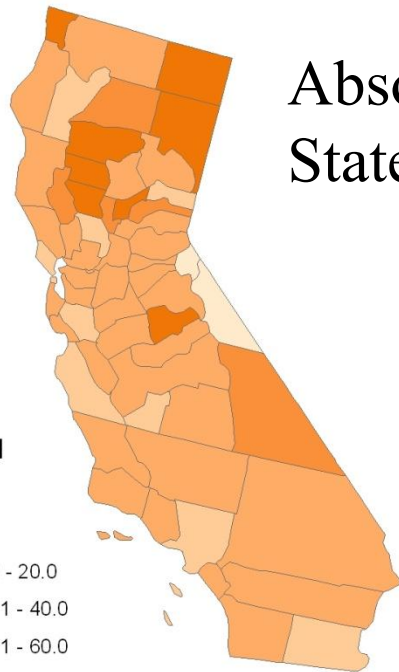




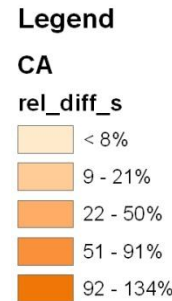
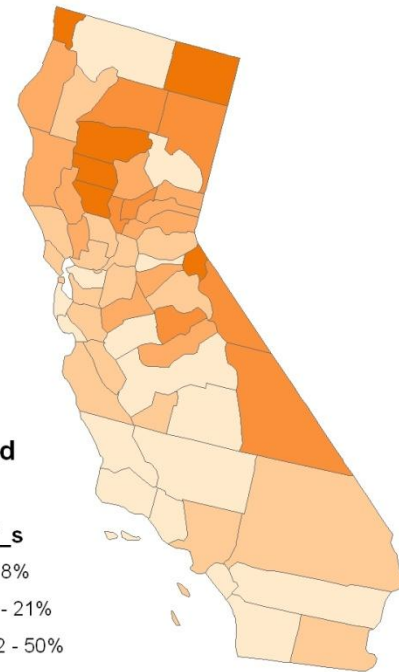
# Comparing County Rates to the State Average

2009

Abs. Pct. Diff



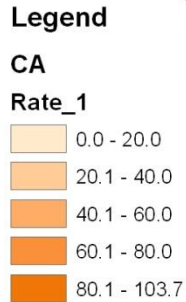
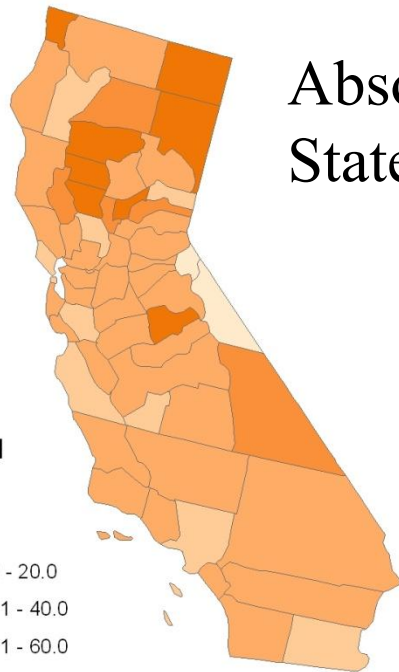
Absolute Percent Difference  
State Average Rate of 44.3



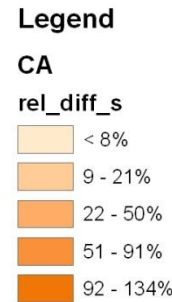
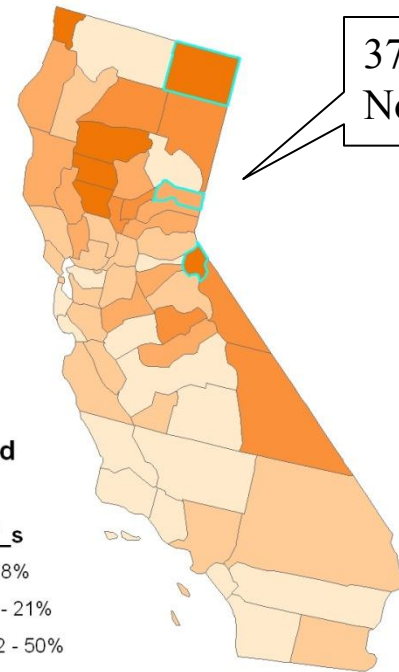
# Comparing County Rates to the State Average

2009

Abs. Pct. Diff



Absolute Percent Difference  
State Average Rate of 44.3

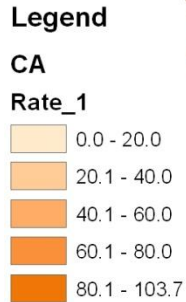
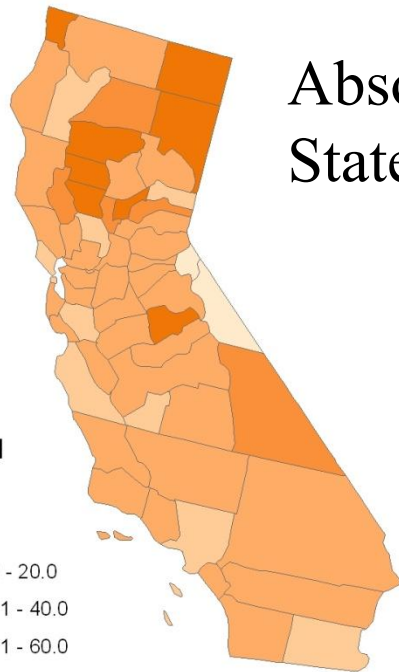


37-133%  
Not Sig.

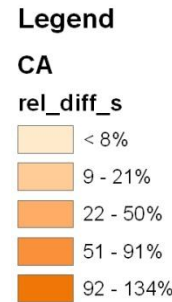
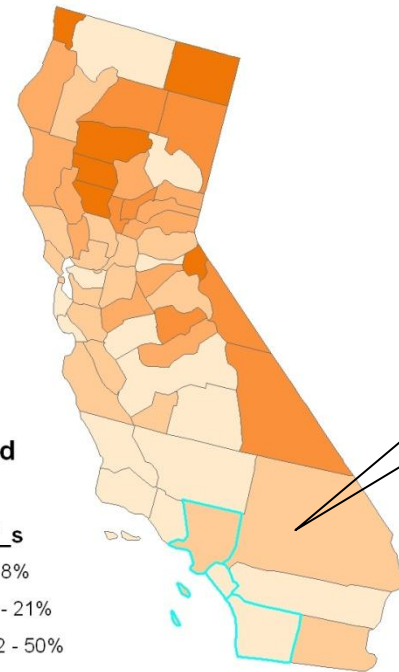
# Comparing County Rates to the State Average

2009

Abs. Pct. Diff



Absolute Percent Difference  
State Average Rate of 44.3

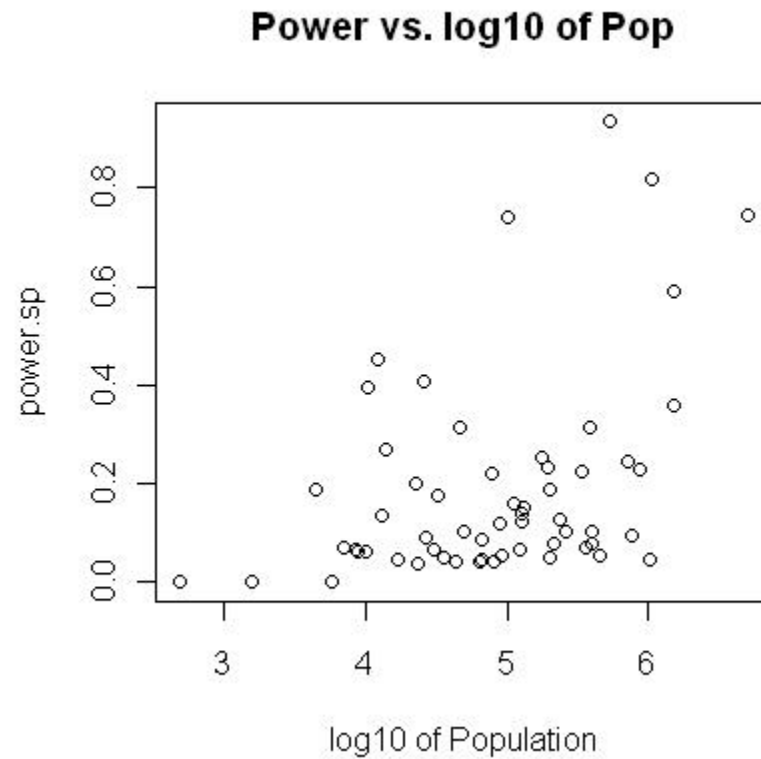


1 - 16%  
1 of 3 Sig.

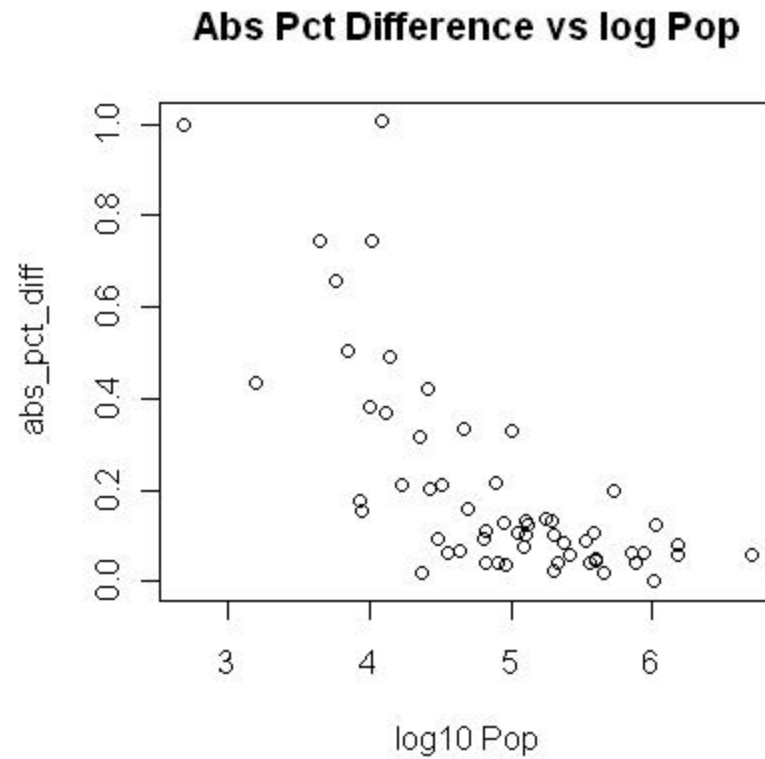
# Differences in Rates

- Big differences are not always significant.
- Significant differences are not always big.
- How much statistical power do I have?

# Power Increases with Pop

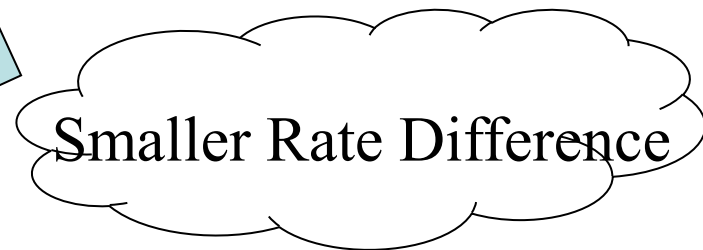
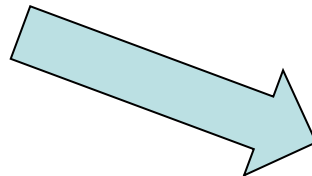
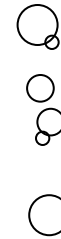
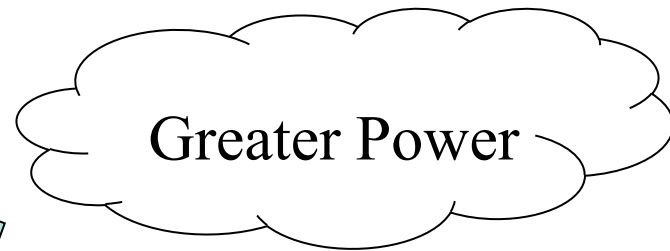
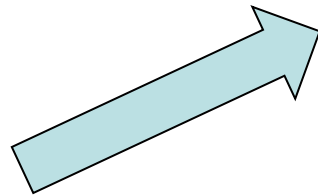


# Abs. Pct. Diff Decreases with Pop



# Statistical Power

Larger Population Size

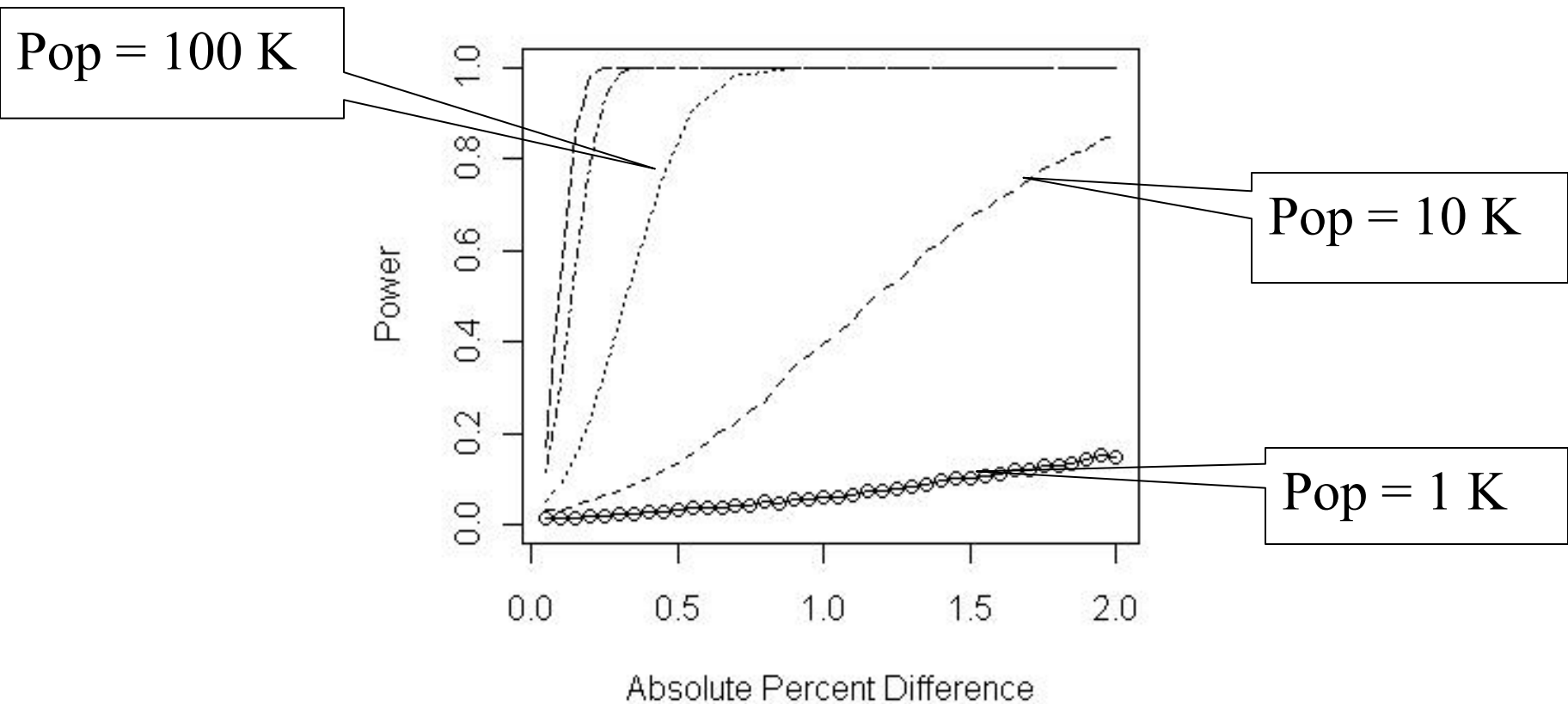


# Clean Up the Mess

- Simulation study
- Population size: 1K, 10K, 100K, 500K, 1M
- Absolute Percent Difference: 0 – 200%
- Pure Poisson Distribution



# Power for Simulated Data: Pop=1K, 10K, 100K, 500K, 1M



# Various Geographic Units in SEER

	# counties	%	# HSAs	%	states	
1 - 1K	35	1.11%	1	0.11%	0	0.00%
1K - 10K	665	21.18%	41	4.34%	0	0.00%
10,001-100K	1918	61.08%	420	44.44%	0	0.00%
100,001-500K	410	13.06%	355	37.57%	1	1.96%
500,001-1M	77	2.45%	69	7.30%	7	13.73%
1,000,001-10M	35	1.11%	59	6.24%	36	70.59%
10,000,001-100M	0	0.00%	0	0.00%	7	13.73%
total # geographic units	3140		945		51	

# Various Geographic Units in SEER

	# counties	%	# HSAs	%	states	
1 - 1K	35	1.11%	1	0.11%	0	0.00%
1K - 10K	665	21.18%	41	4.34%	0	0.00%
10,001-100K	1918	61.08%	420	44.44%	0	0.00%
100,001-500K	410	13.06%	355	37.57%	1	1.96%
500,001-1M	77	2.45%	69	7.30%	7	13.73%
1,000,001-10M	35	1.11%	59	6.24%	36	70.59%
10,000,001-100M	0	0.00%	0	0.00%	7	13.73%
total # geographic units	3140		945		51	

# Various Geographic Units in SEER

	# counties	%	# HSAs	%	states	
1 - 1K	35	1.11%	1	0.11%	0	0.00%
1K - 10K	665	21.18%	41	4.34%	0	0.00%
10,001-100K	1918	61.08%	420	44.44%	0	0.00%
100,001-500K	410	13.06%	355	37.57%	1	1.96%
500,001-1M	77	2.45%	69	7.30%	7	13.73%
1,000,001-10M	35	1.11%	59	6.24%	36	70.59%
10,000,001-100M	0	0.00%	0	0.00%	7	13.73%
total # geographic units	3140		945		51	

# Conclusions

- As population size increases, in general power increases.
- On the other hand, larger population often is associated with smaller rate difference.
- For counties with  $< 100$  K pop, rate difference has to be very high to reach enough power.
- Current geographic units in SEER have varying pop (and power).

# Future Work

- Exploring other geographic units (e.g. PUMA)?
- Considering covariates
- Adding spatial autocorrelation