

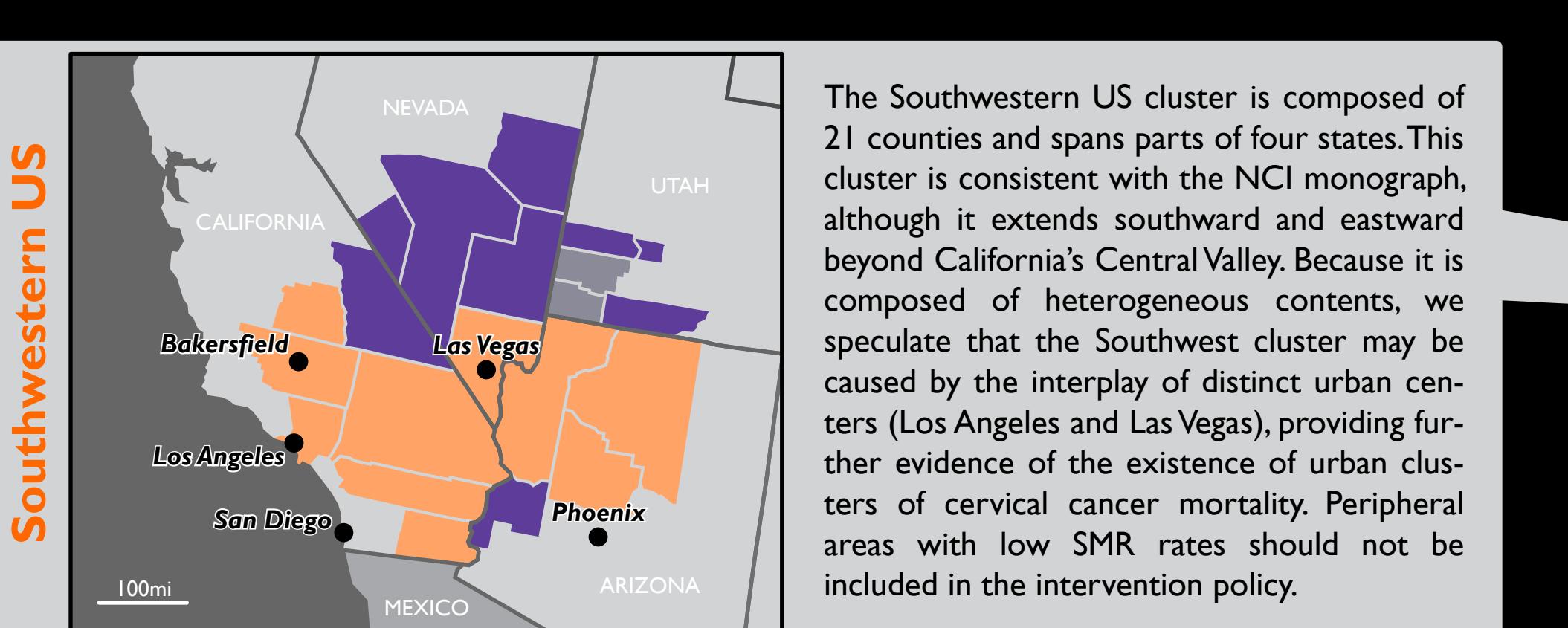
US Cervical Cancer Mortality 2000-2004: New Methods and New Findings

combining Geocomputational, Geovisual Analytic, and Cartographic techniques to identify geographic clusters of elevated cervical cancer

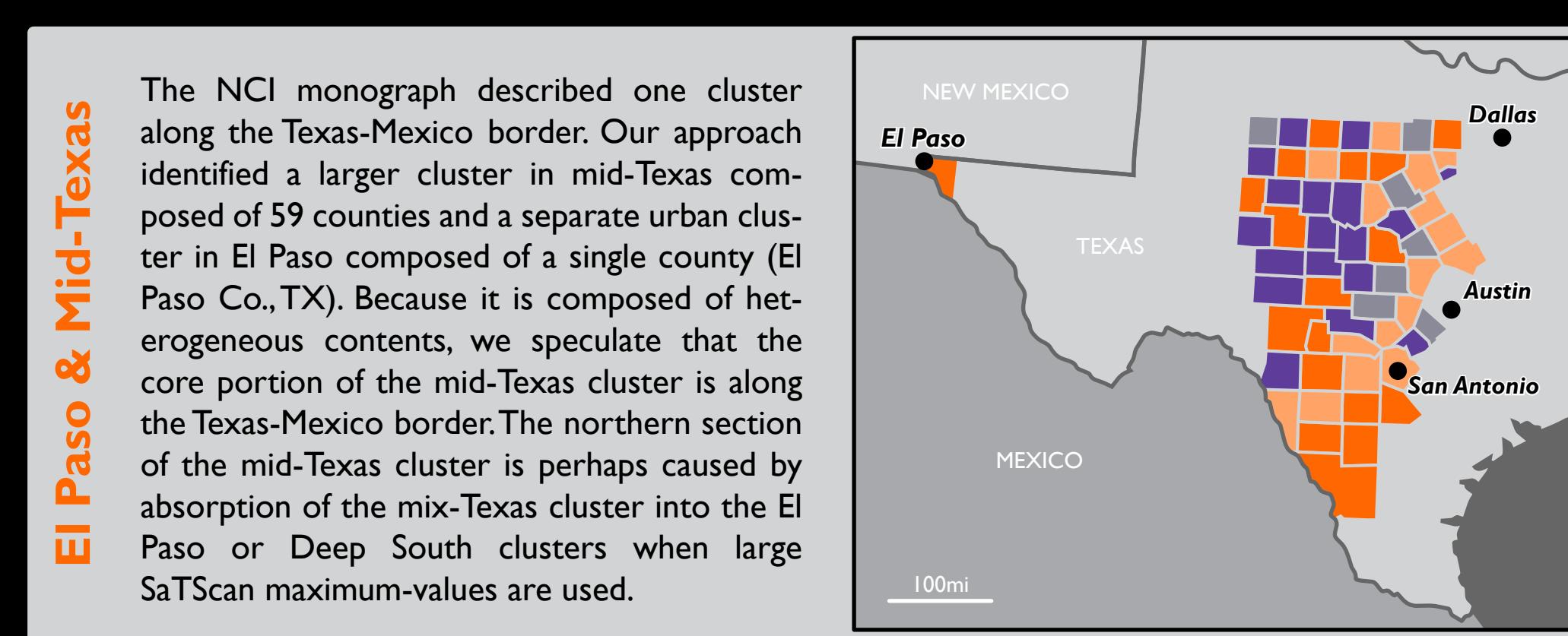
GEOGRAPHIC DISPARITIES IN CERVICAL CANCER MORTALITY

The American Cancer Society estimate that 11,070 American women were diagnosed with cervical cancer in 2008 and that the disease contributed to 3,870 deaths. These figures are troubling because cervical cancer is considered to be **largely preventable**. Since Pap smear screening was introduced in the 1950s, annual US cervical cancer mortality rates have been cut in half (Yabroff 2003). Recent introduction of a vaccine for HPV – which is directly responsible for 90-95% of all cervical cancer cases – offers further promise for the prevention of cervical cancer (Schultz 2003). However, while cervical cancer mortality at the scale of the nation is in decline, the **geographic disparities remain**.

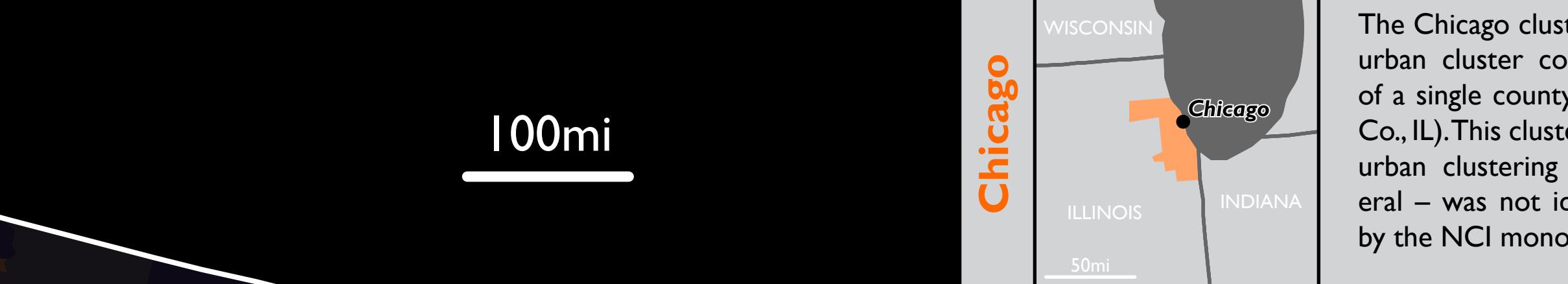
A 2005 monograph published by the National Cancer Institute identified **five geographic clusters** in primarily rural regions of the continental United States exhibiting an elevated risk of cervical cancer mortality: (1) Appalachia, (2) California's Central Valley, (3) the Deep South, (4) the Northern Plains, and (5) the Texas-Mexico border (Freeman & Wingrove 2005). Of particular concern when enacting region-specific health care policies is the method of demarcation (i.e., how were the regions identified and marked?), the reliability of the demarcations (i.e., do the borders of the marked regions change greatly when different clustering parameters are used?), and the homogeneity within the clusters (i.e., do all places in this marked region exhibit similar health risks?). Unfortunately, **robust statistical clustering methods that produce reliable and homogeneous results were not used in the NCI monograph**, making it difficult to determine which regions require immediate health care intervention.



The Southwestern US cluster is composed of 21 counties and spans parts of four states. This cluster is consistent with the NCI monograph, although it extends southward and eastward beyond California's Central Valley. Because it is composed of heterogeneous contents, we speculate that the Southwest cluster may be caused by the interplay of distinct urban centers (Los Angeles and Las Vegas), providing further evidence of the existence of urban clusters of cervical cancer mortality. Peripheral areas with low SMR rates should not be included in the intervention policy.



The NCI monograph described one cluster along the Texas-Mexico border. Our approach identified a larger cluster in mid-Texas composed of 59 counties and a separate urban cluster in El Paso composed of a single county (El Paso Co., TX). Because it is composed of heterogeneous contents, we speculate that the core portion of the mid-Texas cluster is along the Texas-Mexico border. The northern section of the mid-Texas cluster is perhaps caused by absorption of the mid-Texas cluster into the El Paso or Deep South clusters when large SaTScan maximum-values are used.



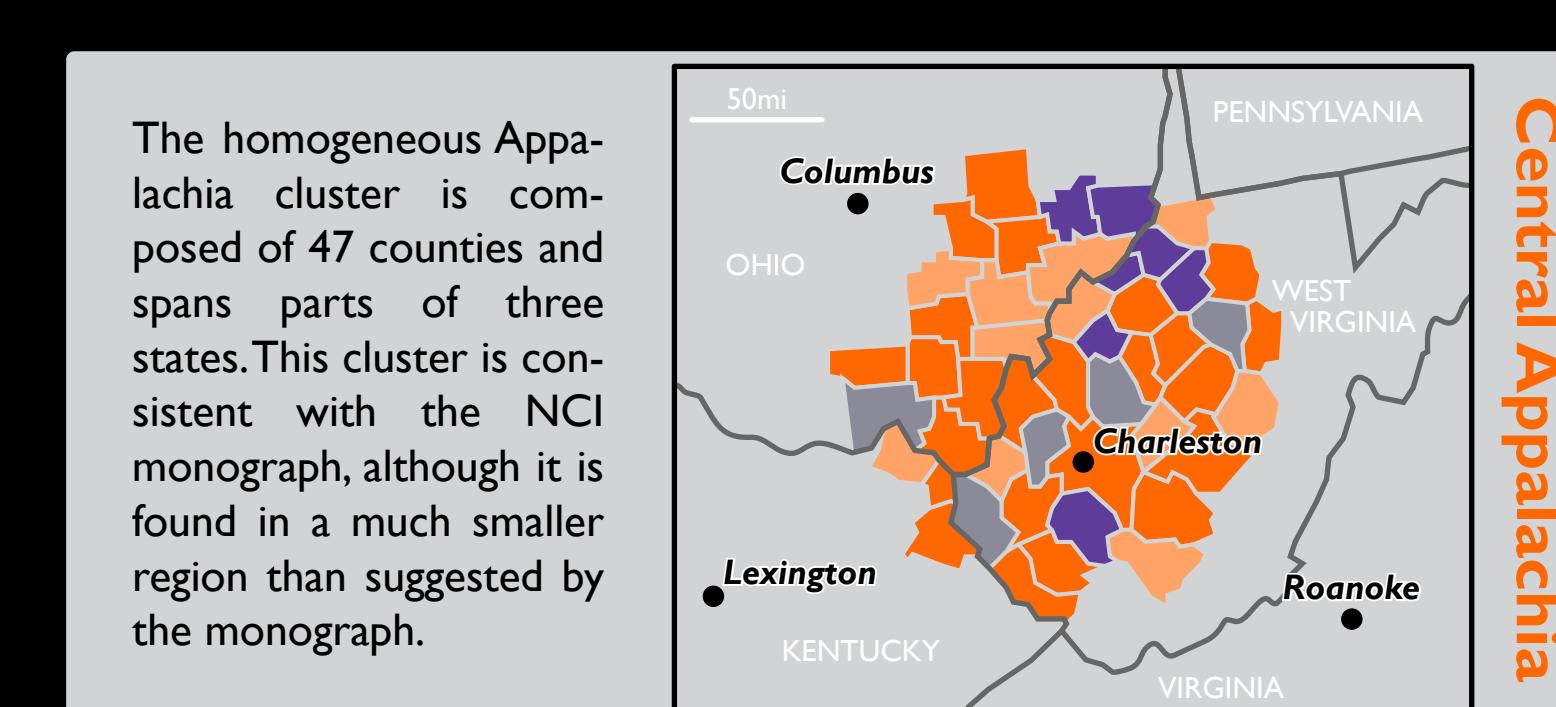
The Chicago cluster is an urban cluster composed of a single county (Cook Co., IL). This cluster – and urban clustering in general – was not identified by the NCI monograph.

NEW METHODS: HOW DID WE MAKE THIS MAP?

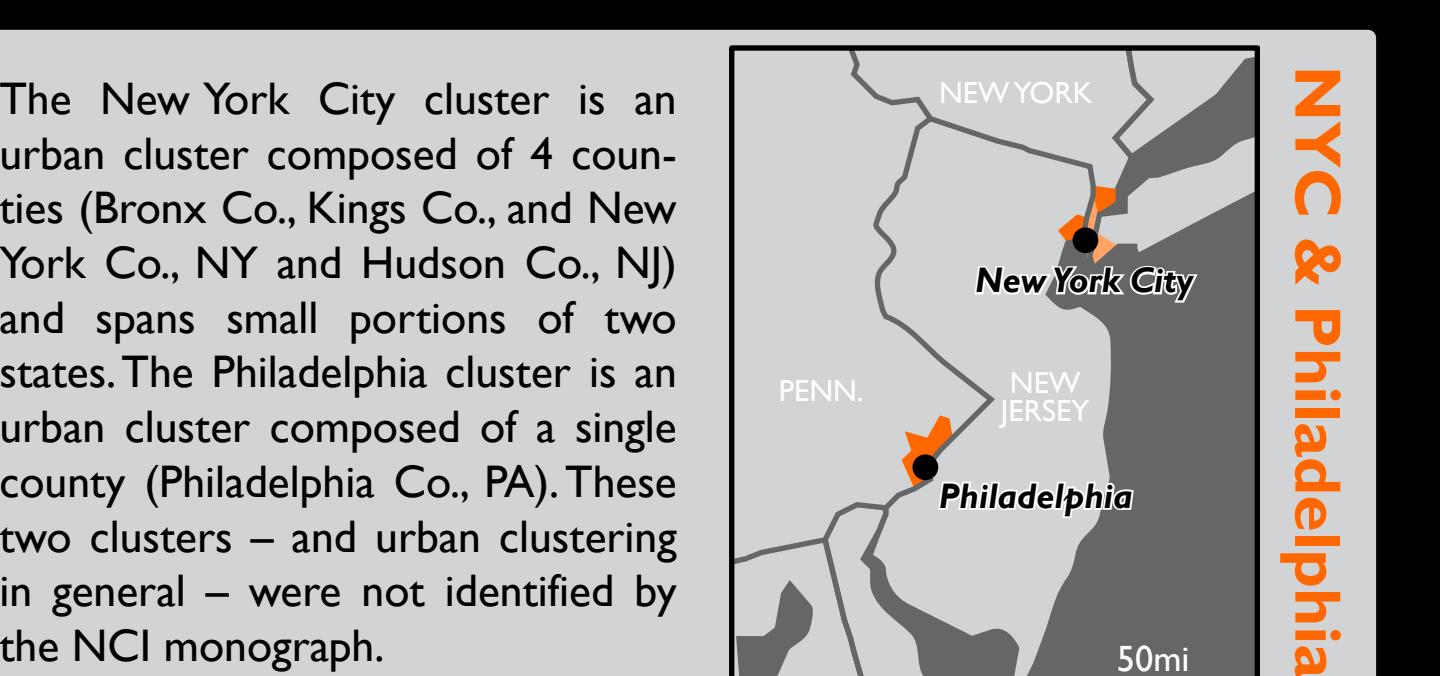
Geocomputation: the Kulldorff spatial scan statistic

The raw population and cervical cancer mortality data for 2000-2004 were obtained from the National Cancer Institute through the SEER program, allowing for the calculation of standardized mortality ratios (SMRs) for each county (excluding Alaskan and Hawaiian counties for geographic continuity). The established **Kulldorff spatial scan statistic** was then applied to this data using the SaTScan software (Kulldorff 1997). The Kulldorff spatial scan statistic imposes a circular window on the map and moves the circle center over each county centroid so that the window includes different arrangements of neighboring centroids at different positions. By adjusting the center location and radius,

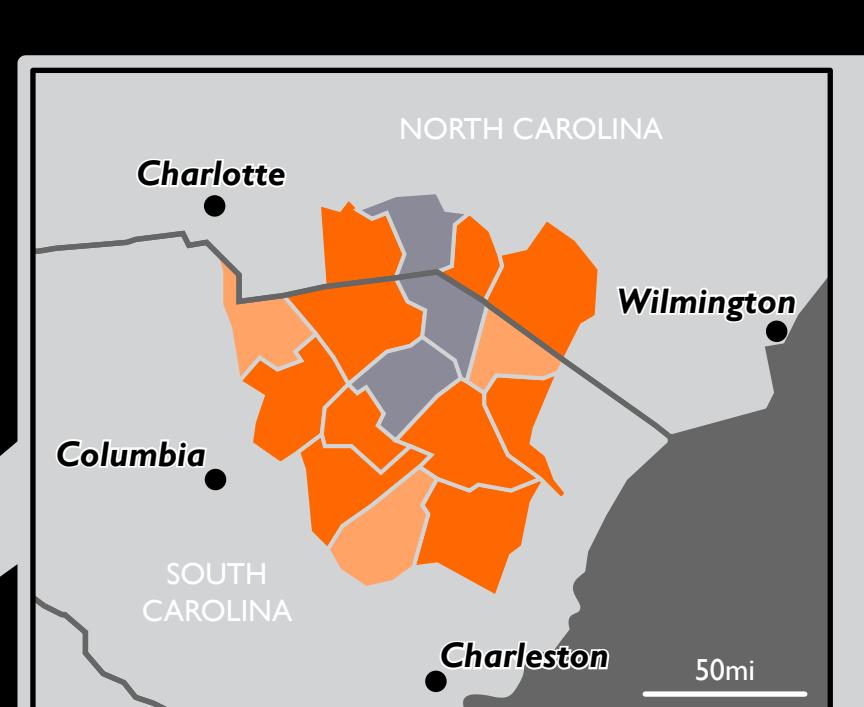
the radius of the circle is increased continuously from 0% to a user-defined **maximum-size** based upon a percentage of the total population-at-risk within the circle. Any circle with a statistically significant level of clustering is reported by SaTScan.



The homogeneous Appalachia cluster is composed of 47 counties and spans parts of three states. This cluster is consistent with the NCI monograph, although it is found in a much smaller region than suggested by the monograph.



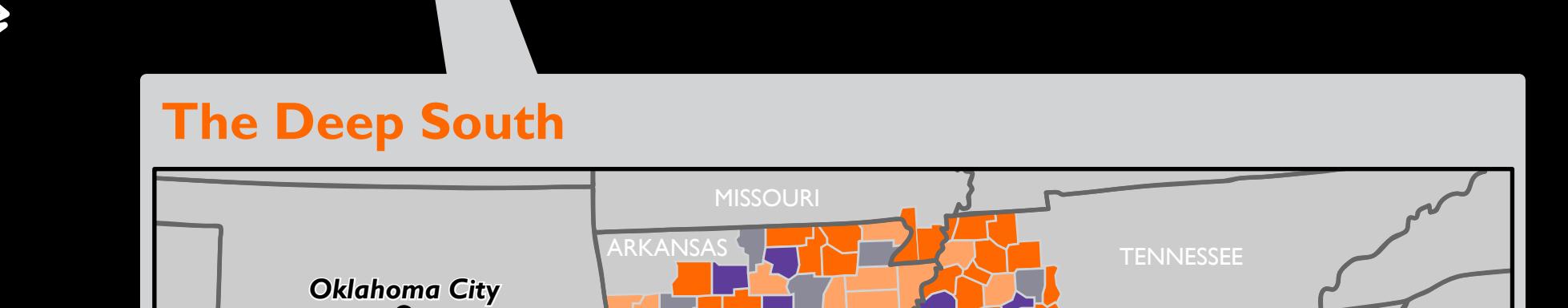
The New York City cluster is an urban cluster composed of 4 counties (Bronx Co., Kings Co., and New York Co., NY and Hudson Co., NJ) and spans small portions of two states. The Philadelphia cluster is an urban cluster composed of a single county (Philadelphia Co., PA). These two clusters – and urban clustering in general – were not identified by the NCI monograph.



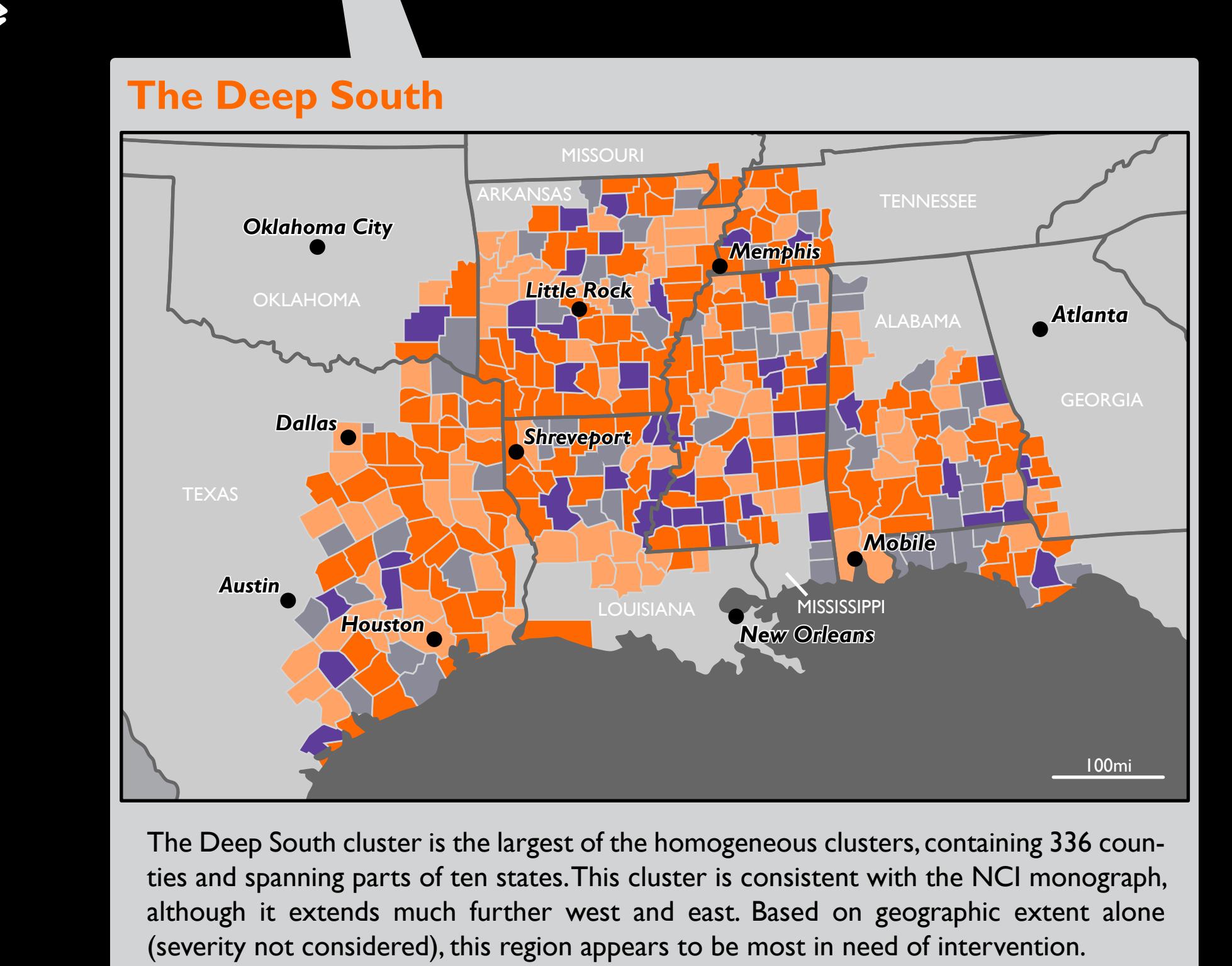
The homogeneous Eastern Carolinas cluster is composed of 16 counties and spans parts of two states. This rural cluster was not identified explicitly by the NCI monograph, although it may have been implicitly included in the Deep South cluster.

Geovisual Analytics: the Reliability Visualization

One primary limitation to the Kulldorff spatial scan statistic is that size and position of the reported clusters are sensitive to the selection of the maximum-size parameter. Too large of a maximum-size value can hide small, homogeneous clusters within larger, heterogeneous ones, and too small of a maximum-size can miss significant, regional-level clusters. We developed a novel **reliability visualization** technique to identify and discriminate reliable clusters of homogeneous contents (see Chen et al. 2008). The method visualizes the reliability that a county is reported within a cluster when SaTScan is run multiple times with a systematically varying maximum-size parameter. Eight SaTScan runs were incorporated into our reliability visualization using maximum-size values of 4%, 6%, 8%, 10%, 20%, 30%, 40%, and 50% of the total population-at-risk. Continuous regions receiving the highest two reliability scores are considered reliable clusters. The inset maps show the homogeneity of these clusters.



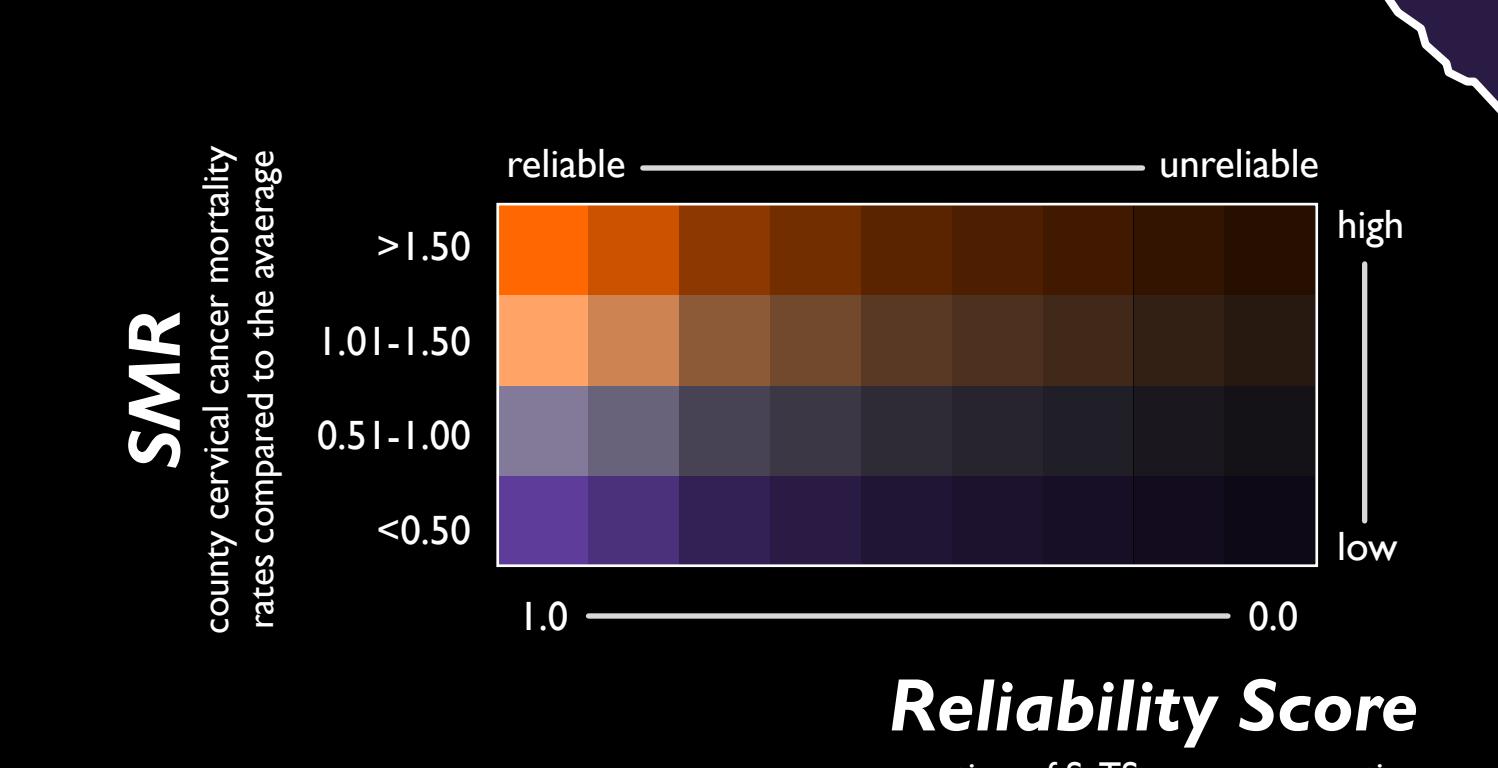
The Deep South cluster is the largest of the homogeneous clusters, containing 336 counties and spanning parts of ten states. This cluster is consistent with the NCI monograph, although it extends much further west and east. Based on geographic extent alone (severity not considered), this region appears to be most in need of intervention.



The Deep South cluster is the largest of the homogeneous clusters, containing 336 counties and spanning parts of ten states. This cluster is consistent with the NCI monograph, although it extends much further west and east. Based on geographic extent alone (severity not considered), this region appears to be most in need of intervention.

NEW RESULTS: WHAT DOES THE MAP SHOW?

Here we report on our research to systematically identify clusters of elevated cervical cancer using established and novel geocomputational, geovisual analytic, and cartographic techniques. The central map shows the **nine geographic clusters** identified by our approach. Like the NCI monograph, we identified rural clusters (1) in central Appalachia, (2) in the Deep South, (3) in the Southwest (including the southern portion of the Californian Central Valley), and (4) in mid-Texas along the Texas-Mexico border. We also found a rural cluster in (5) the Eastern Carolinas that was not explicitly identified in the NCI monograph. We did not find a cluster in the Northern Plains, although this is possibly because of the low population-at-risk in this region yielding unstable mortality rates over the time period. Using our approach, we identified four urban clusters: (6) Chicago, (7) El Paso, (8) New York City and (9) Philadelphia. Because it is composed of heterogeneous contents, we speculate that the Southwest cluster may be caused by the interplay of distinct urban centers (Los Angeles and Las Vegas) combined with edge effects in the analysis method. **Urban cervical cancer clustering is an important finding, as elevated cervical cancer mortality is thought to be primarily a rural phenomenon** (e.g., Freeman & Wingrove 2005) despite evidence that the urban-rural divide does not have a significant effect on Pap smear screening (e.g., Swan et al. 2003).



Reliability Score
proportion of SaTScan runs reporting the county in a significant cluster

Cartographic Representation: Value-by-Alpha Mapping
To generate the central map, we applied a novel value-by-alpha mapping technique to visually equalize the county-level SMR values by their reliability scores. **Value-by-alpha mapping** relates the alpha channel of each displayed enumeration unit to its value in a relevant statistical variable, visually

ranking the enumerations units based on their consequentiality to the map theme (Roth et al. submitted). In the central map, counties with low reliability scores appear to dissolve into the background, causing them to effectively disappear from the map. This disappearance produces a 'spotlight' effect over the counties with high reliability scores due to the color contrast, causing the map reader to focus on the clusters of elevated cervical cancer only.

We would like to thank **Alan MacEachren** and **Gene Lengerich** for their assistance with developing and guiding this research project and **Adam Naito** for his contribution to the cervical cancer data preparation. A full listing of the cited references can be found at: <http://www.personal.psu.edu/rer198/cervicalReferences.pdf>