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

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 constraints). The established Kulldorff spatial scan statistic was then applied to this data using 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 method generates a large number of distinct circular windows, each including a different set of neighboring counties. At each such point location, 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.

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. 2009). 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.

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 based on their consequentiality to the map theme (Kahl 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 "spatial" 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/jr19/cervicalreferences.pdf

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 & Wingo 2005) despite evidence that the urban-rural divide does not have a significant effect on Pap smear screening (e.g., Swin et al. 2003).