

Issues for Discussion

Ecological fallacy: the analysis is at the county level, but is interested in socio-demographics, disability and COVID-19 at the individual level.

Other conceptualizations by Chakraborty use urban - urban-adjacent - rural classifications or the covid vulnerability index to create clusters

GEE models are run at the county level, but counties also vary tremendously by population.

Should clusters be weighted, and are there consequences for not doing so?

SpatialEpi's Kulldorff doesn't include any overlap in the clusters (see line 199 of <https://github.com/rudeboybert/SpatialEpi/blob/master/R/kulldorff.R>) The algorithm creates zones starting from every location (county) and expanding the radius until the maximum population has been included (typically 50% of the total population), ultimately creating the `nearest.neighbors` list of lists. See <https://github.com/rudeboybert/SpatialEpi/blob/9c3c9e3f1583338134f14c62e1a18ade4eca9a7d/R/zones.R> For every `nearest.neighbors` item then, the log-likelihood is calculated, and then the whole list is sorted by descending likelihood. Every single zone is checked for potential inclusion as secondary cluster in descending order, excluding any zones with overlap, until zones are no longer above the specified significance level (usually 0.05). The Monte Carlo simulation is used to determine maximum likelihoods under random distribution of cases, from which p-values can be derived.

SatScan reports secondary clusters as follows: "The most likely cluster will always be reported. Secondary clusters can be reported based in either or both of two criteria. The first is a purely hierarchical criterion where secondary clusters with $p < 1.0$ are reported depending on their degree of geographical overlap with more likely clusters already reported. The second is a criteria based on the Gini index, which is a measure of statistical dispersion. With this criterion, SaTScan selected the group of non-overlapping clusters that maximizes the Gini index, so that there is a big difference in rates between the cluster and non-cluster areas. These latter are called 'gini clusters'." (pg 68 of the user guide)

SMR = Standardized Mortality Ratio = ODE (ordinary differential equation)

note: while spatialEpi uses 0.5 significance threshold, satscan appears to use 0.1.

Should we have used county-level local relative risk or cluster-level cluster relative risk? This probably indicates that we should use county-level local relative risk: "estimate relative risk (RR) for COVID-19 incidence rates at the county level"

There are definitely differences in Kulldorf spatial scan statistic between SatScan and SpatialEpi. What are they?

Unplanned Deviations

- add cluster relative risk score

Results

- add relative risk Results

Discussion

- discuss differences between local and cluster relative risk results