Project 1.1

Non-spatial and spatial analyses was performed for lung cancer data in 88 counties of Ohio over the years 1968-1988. Race, sex and eight different age groups were also used in the analysis. The data was stratified into groups based on these age, sex, race and year cohorts (i.e. strata 1 represents age group = 1, sex = 1, race =1 and year = 1968). The risks, denoted as $\hat{q}_k = \frac{Y_k}{N_k}$, where Y_k and N_k represent the number of deaths and population for strata $k = 1, ..., 672$ were used as the reference risks. The expected number for each county $i = 1, ..., 88$ $\hat{E}_i = \sum_k N_{ik}\hat{q}_k$, was determined. The standardized mortality rates and SMR standard errors, with these expected numbers adjusted for gender, race, sex and year are provided in a map in Figure 2. The indexes of each county were mapped in Figure 1 for easy reference to particular counties.

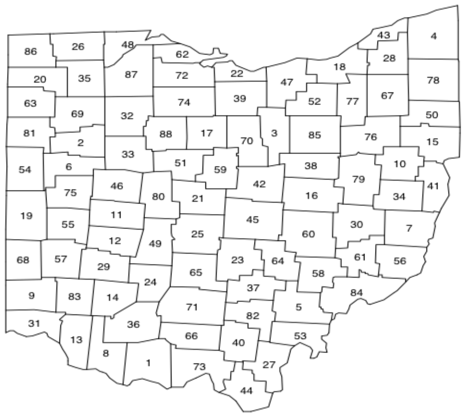

Figure 1: Ohio Map Index

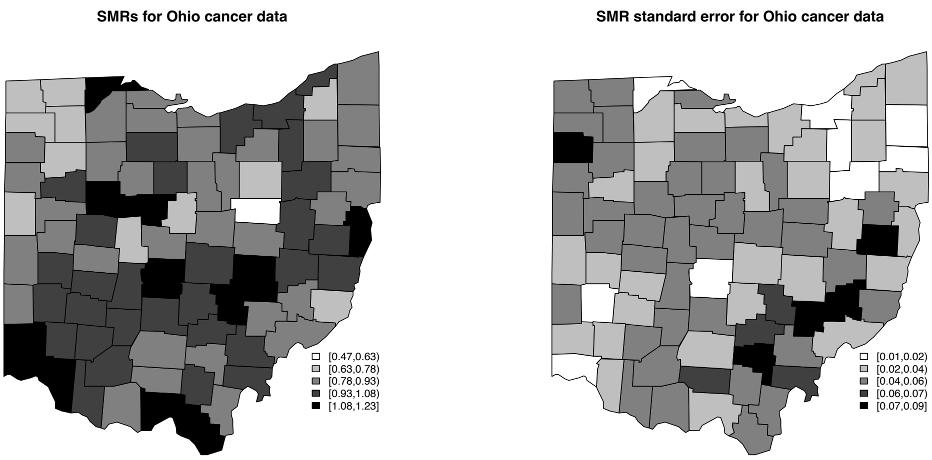

Figure 2: SMRs for Ohio cancer data with expected numbers adjusted for age, sex, race and year

The following Poisson log linear model was fit which provides a basic model to see affect over time:

$$Y_{ij} \sim Poisson(E_{ij} \exp(\alpha_0 + \alpha_1 year))$$
$$\text{where } i = 1, ..., 88 \; j = 68, ..., 88 \text{ and}$$

E_{ij} represents the expected numbers for each county and year conditioned on race, sex and age. The SMRs and their standard errors with expected numbers conditioned on age, sex and race for the years 1968 and 1988 are shown in Figure 3 and 4. The mapped $\hat{\alpha_1}$ estimates (MLE of negative binomial dispersion parameter) and the histogram in Figure 5 show that for most of the counties year does not have a large association with the number of deaths we see. However, from the map of the slopes we can see there are about 5 counties, 8, 58, 66, 69 and 88 where the time has a larger association. A plot of SMR vs year for one of these counties is shown in Figure 6. From this plot of the SMR for county 8 over time, we see there appears to be a general upward trend in the number of lung cancer deaths. For many of the other counties this is not necessarily the case.

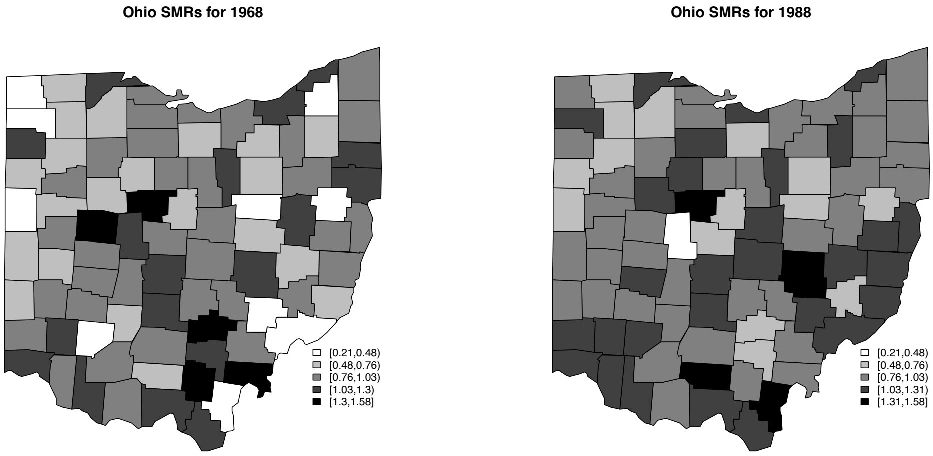

Figure 3: SMRs in Ohio for years 1968 and 1988

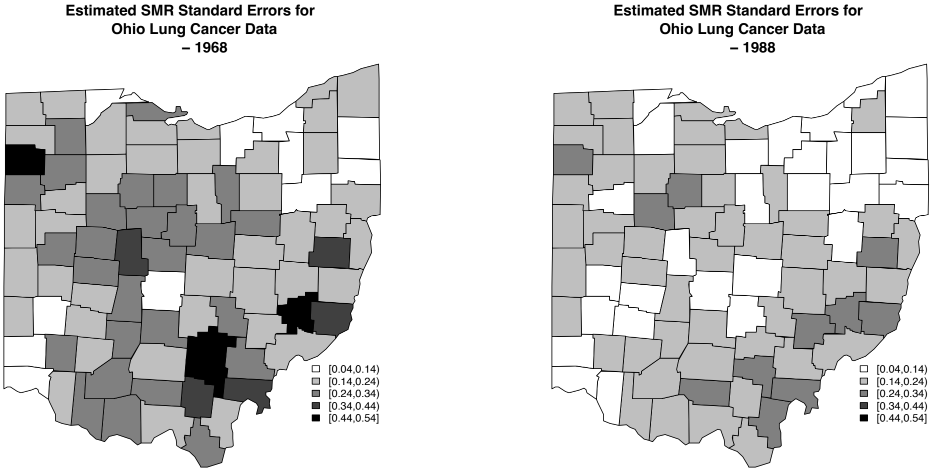

Figure 4: SMR standard errors in Ohio for years 1968 and 1988

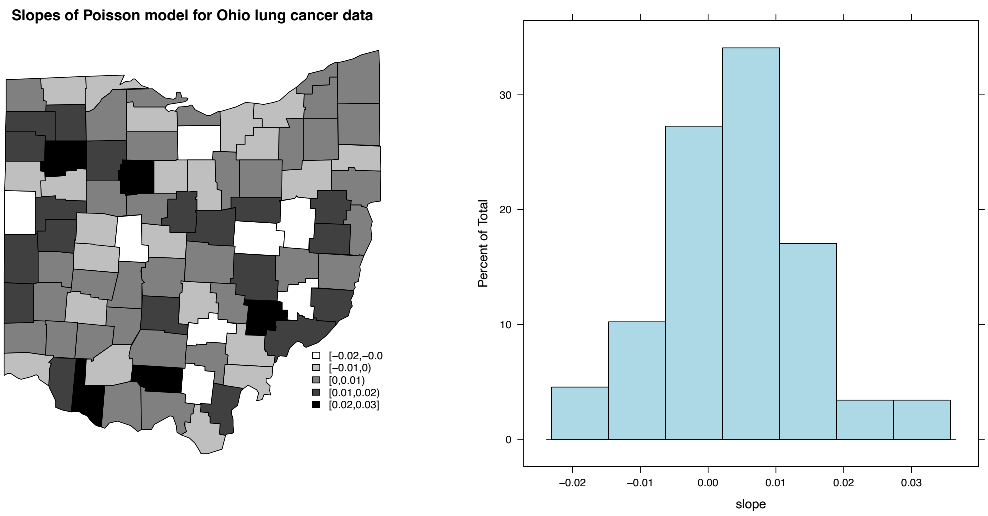

Figure 5: Map and histogram of slopes from the log linear Poisson model

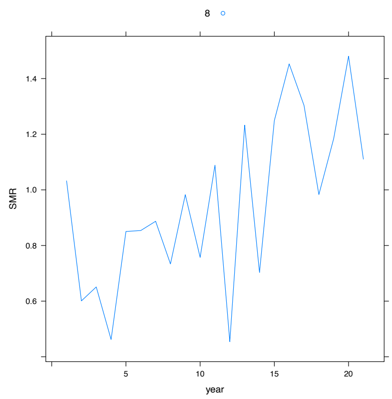

Figure 6: SMR over time for county 8

A map of the posterior mean smoothed empirical Bayes (with gamma prior) estimates $\hat{RR}_i$ are shown in Figure 7. The expected numbers were stratified by age, sex and race. From this we can see that the smoothed estimates have a smaller range than the SMR estimates. In this case, $\hat{\alpha} = 48.3$ and the weights, $w_i = \frac{E_i}{(\hat{\alpha}+E_i)}$ are quite large thus the random effects have a tight spread and there is more shrinkage. This means that the SMRs that are far from unity are inconsistent with the majority of the estimates and have a smaller weight assigned to it. From the plot of the posterior standard deviations the empirical Bayes estimates are less variable than the SMRs because of global smoothing.

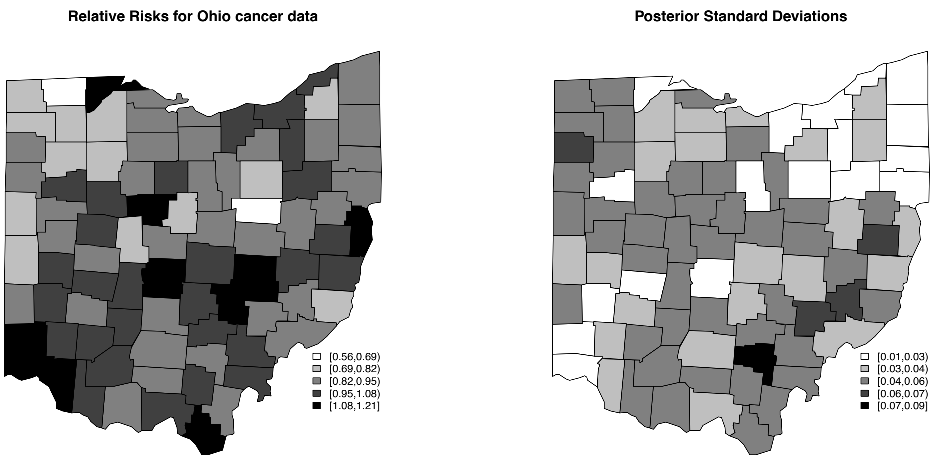

Figure 7: Empirical Bayes relative risks and plot of posterior standard deviations

Next another class of prior models that incorporates the map's geographical structure was considered. In this case the relative risks are conditionally independent of one another except for a small set of neighbours sharing a border with the county in question. For instance, for county 1, the neighbours are counties 8, 36, 66 and 73.

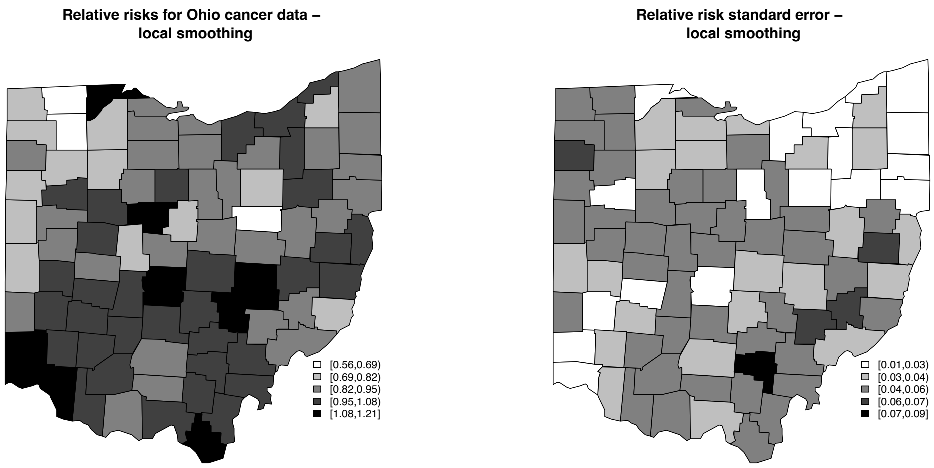

Figure 8: Empirical Bayes relative risks with local smoothing and standard deviations

The relative risk estimates for each area are directly influenced by the neighbouring area estimates and indirectly influenced by the area estimates in the rest of the map. Therefore each county's estimate is shifted toward a local mean rather than a global mean. This means that the prior takes into account the heterogeneity of the geographic structure rather than assuming a homogeneous geographic structure. However, from the maps in Figure 7 and 8, it appears that the result of local smoothing is very similar to global smoothing for both the estimates and their standard errors. In this case $\hat{\alpha}^* = 73.7$, which is ever larger than the global smoothing case thus there is more shrinkage than before. From the above analyses we can see the deficiency of a non-spatial analysis due to areas with smaller populations and potentially greater instability. With spatial analysis, one does not need to assume that the geographical structure is homogeneous, which is ideal since estimates vary quite greatly spatially.