*A poisson distribution was selected because it’s a flexible model that is ideal for attempting to predict the probability of a certain number of times an event occurs (i.e. the number of individuals getting lung cancer). In addition, there are no negative values in such a distribution.*

|  |  |
| --- | --- |
| **Symbol** | **Meaning** |
| i (subscript) | Corresponding to county i |
| j (subscript) | Corresponding to year j |
| Y | Observed cases |
| E | Expected cases |
| θ | Relative risk |
| Γ | Intercept |
| u | Random effect from that specific area (CAR) |
| v | General random effects (normal i.i.d.) |
| β | Global trend (i.e. for all of Texas) |
| δ | Trend for that specific area (i.e. a specific county in Texas) |
| t | Time |

Model 1: **Spatiotemporal General Model for All Lung Cancers**

Yij ~ Poisson (Eijθij)

log(θij) = Γ + ui + vi + (β + δi) x tj

Model 2: **Spatiotemporal Model with Air Quality, Poverty levels & rurality as Covariates**

Model 3:

**Mixture Variable Between Spatial & Spatiotemporal**

**COVID-19 Mortality Rate as a Covariate**

**Rurality as a Covariate**

**Poverty/ Socioeconomic Status as a Covariate**

**Trends for Specific Age/ Race/ Gender Groups**

**Some other potential covariates to investigate based on literature review with SIR as response variable:**

* Air quality/ pollutants
* Poverty rates
* Rurality
* Radon levels
* Asbestos
* Farmland data (pesticides)
* Coalminer data/ other occupation data