Poisson regression models.

Before introducing spatial models we first consider the Poisson regression model that represents the starting point of statistical methods in ecological analysis. Let $\{Y_i, i = 1, \ldots, n\}$ be the set of observed number of events of a certain disease and $\{E_i, i = 1, \ldots, n\}$ the set of expected number under a reference set of age-specific rates for n areas of the region of interest. Then Y_i follows a Poisson distribution with expectation:

$$\mu_i = \theta_i E_i,$$

where θ_i is the relative risk for the site i. The maximum likelihood estimates of θ_i, under a saturated model, are given by the standardized mortality ratios:

$$SMR_i = \frac{Y_i}{E_i}.$$

This model can be extended to a set of explanatory variables $X_1, X_2, \ldots, X_H$ in a log-linear formulation:

$$\log \mu_i = \log E_i + \sum_{h=1}^{H} \beta_h x_{ih}.$$

Maximum likelihood estimates of the coefficients β_h can be obtained in a generalized linear models framework.

A Poisson regression model can include the influences on disease of many ecologic factors (the covariates X_h) but it does not control for the autocorrelation and for the extra-Poisson variability, which may arise due to, for example, unobserved confounder variables. Some authors argued that non linear ecological models give biased estimates of the individual level coefficients. This bias is negligible for moderately large risk ratios (see e.g. Richardson et al., 1987 and Greenland, 1992).

Bayesian mixed models.

Unstructured and structured extra-Poisson sources of variability can be further considered by the following generalized linear mixed model:

$$\log \theta_i = t_i + u_i + v_i,$$

where

$$t_i = \sum_{h=1}^{H} \beta_h x_{ih}$$

denote the fixed regression component, u_i is the random unstructured terms (named *heterogeneity*) and v_i the random spatial structured terms (named *clustering*).

Introducing the heterogeneity and clustering terms represents a way of taking account of unmeasured covariates. Defining appropriate prior distribution on the hyperparameter involved in the model, a Bayesian inference using the posterior distribution of θ_i can be done. In particular estimates of the relative risks can be computed by running MCMC algorithms. This model was introduced by Clayton and Kaldor (1987) in disease mapping framework and then it was developed by Besag et al. (1991) and Clayton et al. (1993) in ecological analysis.