Note: This is a preliminary extended abstract detailing the model as originally developed, for cancer by site. It will be extended at PAA to encompass all causes of death, and beyond a regression framework.

Model for assessing effects of an exposure on multiple cancers simultaneously

Alan Cohen, Groupe de Recherche PRIMUS

Can be used for:

- Geographical region incidence studies (linear model, geographic covariates)
- Individual-level epidemiological studies (logistic or survival model, age/sex covars)
- Laboratory studies (mice/rats)
- Others?

$$RR(\text{Cancer type j}) \sim \alpha_j + (\beta_1 + \beta_2 \times Inf_j + \gamma_j) \times Exp + \boldsymbol{\beta}_{\mathbf{cov(j)}} \times \mathbf{Covs} + \varepsilon_j$$

$$\gamma_j \sim Normal(0, \sigma^2)$$
$$\varepsilon_j \sim Normal(0, \tau_j)$$
$$\tau_j \sim Normal(0, \theta)$$

For j = 1,2,…,n, where n= number of cancer types
RR= Relative Risk
α_j = Intercept for cancer type j
β_1 = Shared (generalized, average) effect of exposure on all cancers
β_2 = Shared (generalized, average) effect of exposure on inflammatory cancers
γ_j = Specific effect of exposure on cancer type j
Inf_j = Whether or not cancer type j is inflammatory (1 or 0 for yes/no)
Exp = Whatever exposure variable we want (could be several)
$\boldsymbol{\beta}_{\mathbf{cov(j)}}$ = a vector of coefficients for whatever other covariates we want, specific to cancer type j
Covs = a vector of all the covariates as measured
ε_j = error term (measurement error, etc)
σ^2 = variance of the effects of the exposure on different types of cancer
τ_j = error variance for cancer type j (unexplained variance)
θ = Variance in the error variances across cancer types

Example:

$$RR(\text{Liver cancer}) \sim \alpha_{LC} + (\beta_1 + \beta_2 \times Inf_{LC} + \gamma_{LC}) \times Exp + \boldsymbol{\beta}_{\mathbf{cov(LC)}} \times \mathbf{Covs} + \varepsilon_{LC}$$
$$RR(\text{Breast cancer}) \sim \alpha_{BC} + (\beta_1 + \beta_2 \times Inf_{BC} + \gamma_{BC}) \times Exp + \boldsymbol{\beta}_{\mathbf{cov(BC)}} \times \mathbf{Covs} + \varepsilon_{BC}$$
$$RR(\text{Colon cancer}) \sim \alpha_{CC} + (\beta_1 + \beta_2 \times Inf_{CC} + \gamma_{CC}) \times Exp + \boldsymbol{\beta}_{\mathbf{cov(CC)}} \times \mathbf{Covs} + \varepsilon_{CC}$$
$$RR(\text{Prostate Cancer}) \sim \alpha_{PC} + (\beta_1 + \beta_2 \times Inf_{PC} + \gamma_{PC}) \times Exp + \boldsymbol{\beta}_{\mathbf{cov(PC)}} \times \mathbf{Covs} + \varepsilon_{PC}$$
$$RR(\text{Leukemia}) \sim \alpha_{Leuk} + (\beta_1 + \beta_2 \times Inf_{Leuk} + \gamma_{Leuk}) \times Exp + \boldsymbol{\beta}_{\mathbf{cov(Leuk)}} \times \mathbf{Covs} + \varepsilon_{Leuk}$$

Etc.

Variations, as necessary:

- Lump rare cancers together to improve parameter estimation
- Use same error, covariate, or intercept terms across cancer types so that rarer cancer type effects can be estimated

- Combinations that allow variation of ancillary parameters across major cancers, but fixed for minor cancers

Questions the model can answer:

- How much does the exposure affect risk/rates of cancer in general? (β_1)
- How much does the exposure affect risk/rates of each cancer type? (γ_j)
- How much does the effect depend on whether or not the cancer is inflammatory? (β_2)
- How much does the effect vary across types of cancers? (σ^2)

Outcome measures should be standardized so that effect sizes are easily interpretable. I have run preliminary analyses using this method on incidence data for six types of cancer in 48 US states in 2006 relative to state smoking levels in 1995. The model was able to detect an overall effect of smoking on cancer, and beyond that to find additional effects on lung and colon cancer, but not on leukemia or breast, liver, or prostate cancer, confirming what Pearson correlations suggest.