## Generalized Linear Models

- GLM from a Bayesian prospective
- Hierarchical models, focusing on the normal model for the GLM-coefficients
- Hierarchical logistic regression
- Example: rat tumor data



- The purpose of GLM is to extend the idea of linear modeling to cases for which the linear relationship between $X$ and $E(y \mid X)$ or the normal distribution is not appropriate
- In some cases, we can apply a linear model to a suitably transformed outcome variable using suitably transformed (or untransformed) explanatory variables
- However, the relation between $X$ and $E(y \mid X)$ cannot always be usefully modeled as normal and linear, even after transformation



A GLM is specified in 3 stages

1. the linear predictor $\eta = X\beta$
2. the link function $g(.)$ that relates the linear predictor to the mean of the outcome variable

$$\mu = g^{-1}(\eta) = g^{-1}(X\beta)$$

3. the random component specifying the distribution of the outcome variable $y$ with mean $E(y \mid X) = \mu$;



## Notation

- $E(y \mid X) = g^{-1}(X\beta)$
- $X$ is the $n \times p$ matrix of explanatory variables
- $\eta = X\beta$ is the vector of $n$ linear predictor values
- $y = (y_1, \ldots, y_n)$ vector of observations
- $p(y \mid X\beta, \phi) = \prod_{i=1}^{n} p(y_i \mid (X\beta)_i, \phi)$

## Special cases

- Poisson and Binomial model $\phi = 1$
- Normal model is a special case of GLM when $y \sim N(\mu, \phi)$ and $g(\mu) = \mu$

**Poisson**

- $y_i \sim$ Poisson$(\mu_i)$ where $\mu_i = (X\beta)_i$
- $\log \mu_i = \eta_i = (X\beta)_i$
- $p(y \mid \beta) = \prod_{i=1}^{n} \frac{1}{y_i!} \exp(-\exp(\eta_i))(\exp(\eta_i))^{y_i}$

**Binomial**

- $y_i \sim$ Binomial$(n_i, \mu_i)$, $n_i$ known
- $g(\mu_i) =$ logit$(\mu_i)$
- $p(y \mid \beta) = \prod_{i=1}^{n} \binom{n_i}{y_i} \left(\frac{\exp(\eta_i)}{1+\exp(\eta_i)}\right)^{y_i} \left(\frac{1}{1+\exp(\eta_i)}\right)^{n_i - y_i}$





- Overdispersion: possibility of variation beyond that of the assumed sampling distribution
- If we consider a logistic regression in which the sampling unit is the litter of mice and the proportion of the litter born alive is considered binomial with some explanatory variables (such as dose). The data might indicate more variation than expected due to systematic differences among mothers.
- Such variation could be incorporated in a hierarchical model using an indicator for each mother, with these indicators themselves following a distribution such normal or beta.



- Units are clustered (extra-binomial variation)



**Canonical link functions**

- the canonical link is the function of the mean parameter that appears in the exponent of the exponential family form of the probability density

**Offset**

- It is sometimes convenient to express a generalized linear model so that one of the explanatory variables has a known coefficient
- an explanatory variable having a kwown coefficient is called offset

**Example: Model for Poisson data**

- $y$ is the $\#$ of incidents in a given exposure time $T$
- if the rate of occurrence is $\mu$ per unit of time
- then the number of incidents $y$ is Poi($\mu T$), where $E(y) = \mu T$, and $\log(\mu) = X\beta$
- However GLM are parameterized through the mean of $y$ which is $\mu T$, where $T$ now represents the vector of exposure times for the units in the regression.
- We can apply Poisson-GLM by augmenting $X$ with a column containing the values $\log T$; this column of the augmented matrix corresponds to a coefficient with known value (equal to -1)



Bayesian non-hierarchical and hierarchical GLM

Bayesian GLM with

1. non informative prior on $\beta$
2. informative prior on $\beta$
3. hierarchical models for which the prior on $\beta$ depends on unknown parameters
4. some generalized linear models are expressed with a dispersion parameter $\phi$ in addition to the regression coefficients $\beta$



Non informative prior on $\beta$

- the classical analysis of $GLM$ is obtained if a non informative or flat prior distribution is assumed for $\beta$
- the posterior mode corresponding to a non informative uniform prior density is the MLE for $\beta$ which can be obtained using Iteratively Weighted Linear Regression (IWLR)
- approximate posterior inference can be obtained from a normal approximation to the likelihood (see 4.1)



Conjugate prior on $\beta$

- Specify a prior for $\beta$ in terms of hypothetical data obtained under the same model
- $y_0$ is a vector of $n_0$ hypothetical data points
- $X_0$ is the corresponding matrix of $k$ explanatory variables
- this is equivalent to consider an augmented vector $\begin{pmatrix} y \\ y_0 \end{pmatrix}$ with matrix $\begin{pmatrix} X \\ X_0 \end{pmatrix}$ and a non informative uniform prior density on $\beta$

## Non conjugate prior on $\beta$

- $\beta \sim N(\beta_0, \Sigma_\beta)$
- a normal prior on $\beta$ is very convenient when we implement a normal approximation to the likelihood

## Hierarchical Models

- As in the normal linear model, hierarchical prior distributions for GLM are a natural way to fit complex data structures and allow us to include more explanatory variables without encountering the problem of over fitting
- a normal distribution for $\beta$ is commonly used



## Computing

- Posterior inference in generalized linear models typically requires simulation techniques (metropolis within Gibbs)
- Given a method for approximating the likelihood by a normal distribution, computation can proceed as follows
  1. Find the posterior mode using an iterative algorithm where iteration in the mode-finding algorithm uses a quadratic approximation to the log posterior density of $\beta$ (and $\phi$) and weighted linear regression.
  2. Use the normal approximation as a starting point for simulations from the exact posterior distribution



## Normal approximation to the joint posterior distribution

- If the posterior distribution $p(\theta \mid y)$ is unimodal and roughly symmetric, it is often convenient to approximate it by a normal distribution centered at the mode
- A Taylor series expansion of $\log p(\theta \mid y)$ centered at the posterior mode $\hat{\theta}$ gives

$$\log p(\theta \mid y) = \log p(\hat{\theta} \mid y) + \frac{1}{2}(\theta - \hat{\theta})^t \left[ \frac{d^2}{d\theta^2} \log p(\theta \mid y) \right]_{\theta=\hat{\theta}} (\theta - \hat{\theta})$$

- the linear term in the expansion is zero because the log posterior has zero derivative at its mode

$$p(\theta \mid y) \simeq N(\hat{\theta}, I(\hat{\theta})^{-1})$$

- where $I(\theta)$ is the observed information

$$I(\theta) = -\frac{d^2}{d\theta^2} \log p(\theta \mid y)$$



## Normal approximation to the likelihood of a GLM

For each data point $y_i$, we construct a

- pseudo datum $z_i$
- pseudo variance $\sigma_i^2$ so that

$$p(y_i \mid (X\beta)_i, \phi) \simeq N(z_i \mid (X\beta)_i, \sigma_i^2)$$
$$y \mid X\beta, \phi \simeq N_n(z \mid X\beta, diag(\sigma_1^2, \ldots, \sigma_n^2))$$

## Center of the normal approximation

- The normal approximation will depend on the value $\beta$ and $\phi$ at which it is centered $(\hat{\beta}, \hat{\phi})$
- In the mode-finding stage of computations, we iteratively alter the center of the normal approximation. Once we have reached the mode, we use the normal approximation at that fixed value of $\hat{\beta}, \hat{\phi}$.

## Determining parameters of the normal approximation

$$\begin{aligned} p(y_1, \ldots, y_n \mid \eta, \phi) &= \prod_{i=1}^n p(y_i \mid \eta_i, \phi) \\ &= \prod_{i=1}^n \exp\left(L(y_i \mid \eta_i, \phi)\right) \\ L(y_i \mid \eta_i, \phi) &\simeq -\frac{1}{2\sigma_i^2}(z_i - \eta_i)^2 + const \end{aligned}$$

where $z_i, \sigma_i^2$ and the constant depend on $y, \hat{\eta}_i = (X\hat{\beta})_i$ and $\hat{\phi}$

$$\begin{aligned} z_i &= \hat{\eta}_i - \frac{L'(y_i|\hat{\eta}_i,\hat{\phi})}{L''(y_i|\hat{\eta}_i,\hat{\phi})} \\ \sigma_i^2 &= -\frac{1}{L''(y_i|\hat{\eta}_i,\hat{\phi})} \end{aligned}$$

- match the 1-th and 2-th order terms of the Taylor series of $L(y_i \mid \eta_i, \phi)$ centered at $\hat{\eta}_i = (X\beta)_i$
- the $i$-th data point is approximately equivalent to an observation $z_i$ with mean $\eta_i$ and variance $\sigma_i^2$



## Binomial logistic model

$$\begin{aligned} L(y_i \mid \eta_i, \phi) &= y_i \log\left(\frac{\exp(\eta_i)}{1+\exp(\eta_i)}\right) \\ &\quad + (n_i - y_i)\log\left(\frac{1}{1+\exp(\eta_i)}\right) \end{aligned}$$

Find $dL_i/\eta_i,\ d^2L_i/\eta_i^2,\ z_i$ and $\sigma_i^2$

## Combining likelihood with an informative or hierarchical prior

- Any normal prior distribution for $\beta$ is conjugate to the normal approximation to the likelihood of $\beta$ (conditional on $\phi$).
- If a non-normal prior model is used, then one can construct a normal approximation for the prior density as well, using the same method of fitting to the linear and quadratic terms of the Taylor series.



## Finding the posterior mode $\hat{\beta}, \hat{\phi}$

For non-informative or conjugate prior distributions we can use IWLR (`Splus function glm` )

- Once the mode $\hat{\beta}, \hat{\phi}$ has been reached, one can approximate the conditional posterior of $\beta$ by the output of the most recent weighted least square computation, that is

$$\begin{aligned} \beta \mid \hat{\phi}, y &\simeq N(\beta \mid \hat{\beta}, V_\beta) \\ V_\beta &= \left(X^t diag(L''(y_i \mid \hat{\eta}_i, \hat{\phi})X\right)^{-1} \\ p_{approx}(\phi \mid y) &\propto p(\hat{\beta}(\phi), \phi \mid y)|V_\beta(\phi)|^{-1/2}, \quad (9.7) \end{aligned}$$

where $\hat{\beta}$ and $V_\beta$ are the mode and the variance matrix of the normal approximation conditional on $\hat{\phi}$



## Sampling from the posterior distribution

- Sampling from the exact posterior distribution for a GLM can be drawn by means of iterative simulations using the above normal distribution as starting distribution.
- Unfortunately methods from sampling from the exact posterior distribution must be developed separately for each class of models

## Bayesian Computations for GLM

### Summary

- $\beta =$ regression parameters
- $\phi =$ overdispersion parameter
- $p(y \mid X, \beta) = \prod_{i=1}^{n} p(y_i \mid (X\beta)_i, \phi)$
- $E[y \mid X\beta] = g^{-1}(X\beta)$
- Three cases:
  1. $\phi$ known $\beta$ unknown, $p(\beta) \propto c$
  2. $\phi$ known $\beta$ unknown, $\beta \sim N(\beta_0, \Sigma_0)$
  3. $\phi$ and $\beta$ unknown $p(\beta, \phi) = p(\phi)p(\beta \mid \phi)$



### 1. $\phi$ known $\beta$ unknown, $p(\beta) \propto c$

- $p(\beta \mid y) \propto p(y \mid \beta)$
- posterior mode for $\beta$ is equal to the MLE
- if $n$ is large, then we can obtain our posterior inference by using a normal approximation to the likelihood function

$$p_{\text{approx}}(\beta \mid y) = N(\hat{\beta}, V)$$

where $\hat{\beta}$ and $V$ are the mle and the corresponding sample covariance matrix (see normal approximation to the posterior distribution)



### 2. $\phi$ known $\beta$ unknown, $\beta \sim N(\beta_0, \Sigma_0)$

- $p(\beta \mid y) \propto p(y \mid \beta) \times N(\beta_0, \Sigma_0)$
- if we use a normal approximation to the likelihood, e.g.

$$p_{\text{approx}}(y \mid \beta) = N(\hat{\beta}, V)$$

where $\hat{\beta}$ and $V$ are the mle and the corresponding sample covariance matrix, then

$$\begin{aligned} p_{\text{approx}}(\beta \mid y) &= N\left(W[V^{-1}\hat{\beta} + \Sigma_0^{-1}\beta_0], W\right) \\ W &= [V^{-1} + \Sigma_0^{-1}]^{-1} \end{aligned}$$



### 3. $\phi$ and $\beta$ unknown $p(\beta, \phi) = p(\phi)p(\beta \mid \phi)$

- $p(\beta, \phi \mid y) \propto p(\beta, \phi)p(y \mid \beta, \phi)$
- find posterior mode $(\hat{\beta}, \hat{\phi})$
- approximate $p(\beta \mid \phi, y)$ by

$$p_{\text{approx}}(\beta \mid \phi, y) = N(\hat{\beta}(\phi), V_\beta(\phi))$$

where $\hat{\beta}(\phi)$ is the conditional mode and $V_\beta(\phi)$ is corresponding sample covariance matrix

- approximate $p(\phi \mid y)$ as

$$p_{\text{approx}}(\phi \mid y) \propto p(\hat{\beta}(\phi), \phi \mid y) \mid V_\beta(\phi) \mid^{\frac{1}{2}}$$

and use inverse CDF method or importance sampling.

**Alternatively...**

Use simulation based methods such as MCMC.

Table 1: *Tumor incidence in a group of rats given three different doses of phenformin, from Tarone(1982). Data from previous experiments are given in table 5.1*

| Dose level, $x_i$ | # rats $n_i$ | # rats with tumor $y_i$ |
|---|---|---|
| 0 | 14 | 4 |
| 1 | 34 | 4 |
| 2 | 34 | 2 |

- the goal of the study is to estimate the dose-response relation - the rate at which the tumor risk increases (or decreases) as function of the dose (see pag 292 text book).
- data from other similar experiments were available (table 5.1) which displays the results of 70 previous experiments in the same strain of rat, all under zero dose.
- the row in table 5.1 for the current experiment is repeated as the first row in Table 9.5 here



## Hierarchical logistic regression model

- current experiment

$$\begin{aligned} y_i &\sim \text{Bin}(n_i, \pi_i) && i = 0, 1, 2 \\ \text{logit}(\pi_i) &= \alpha + \beta x_i && i = 0, 1, 2 \\ \beta &\sim \text{un informative} \end{aligned}$$

- historical data (additional information about $\alpha$)

$$\begin{aligned} y_{j0} &\sim \text{Bin}(n_{j0}, \pi_{j0}) && j = 1, \ldots, 70 \\ \text{logit}(\pi_{j0}) &= \alpha_j && j = 1, \ldots, 71 \\ \alpha_j &\sim N(\mu, \tau^2) \\ \mu, \tau &\sim \text{constant} \end{aligned}$$

- $\pi_{J+1,0} = \pi_0$ and $\alpha_{J+1} = \alpha$



## The joint posterior distribution

$$\begin{aligned} p(\alpha, \beta, \mu, \tau \mid y) &\propto \prod_{j=1}^{J+1} \text{Bin}(y_{j0} \mid n_{j0}, \text{logit}^{-1}(\alpha_j)) \\ &\times \prod_{j=1}^{J+1} N(\alpha_j \mid \mu, \tau^2) \\ &\times \prod_{i=1}^{2} \text{Bin}(y_i \mid n_i, \text{logit}^{-1}(\alpha_{J+1} + \beta x_i)) \\ &\propto \prod_{j=1}^{J+1} e^{\alpha_j y_{j0}} (1 + e^{\alpha_j})^{-n_{j0}} \\ &\times \prod_{j=1}^{J+1} \frac{1}{\tau} \exp\left(-\frac{1}{2\tau^2}(\alpha_j - \mu)^2\right) \\ &\times \prod_{i=1}^{2} e^{(\alpha_{J+1} + \beta x_i) y_i} \\ &\times \prod_{i=1}^{2} (1 + e^{\alpha_{J+1} + \beta x_i})^{-n_i} \end{aligned}$$

- Posterior mode of $p(\alpha, \beta, \mu, \tau \mid y)$ hard to obtain
- Better to work with $p(\alpha, \beta, \mu \mid \tau, y) p(\tau \mid y)$



## Crude estimate of $\beta$ from the current experiment

- Given an uniform prior, the MLE estimate is equivalent to the posterior mode
- we found $\hat{\beta} = -.94, std(\hat{\beta}) = .47$

## Crude estimate of $\beta$ using complete pooling

- we pool the zero-dose data from all the 71 experiments, and we use these values for $y_0$ and $n_0$ in place of the first row of the table
- we found $\hat{\beta} = -.47, std(\hat{\beta}) = .29$
- $\hat{\tau} = \sqrt{var(\text{logit}(\frac{y_{j0}}{n_{j0}}))} = .98$
- sample std of the $\text{logit}(\frac{y_{j0}}{n_{j0}})$

```
# non hierarchical logistic regression - tumor data (pag 292 of Gelman)
make.data_ function() {
  tmp _ cbind(c(4,4,2),c(14,34,34),c(0,1,2))
  tarone _ matrix(scan("tarone.txt"),byrow=T,ncol=2)
  yy0    _ tarone[,1]
  nn0    _ tarone[,2]
  yy _ tmp[,1]       #number of tumors
  nn _ tmp[,2] #total number of births
  xx _ tmp[,3]       #dose
  sf _ cbind(yy,nn-yy)
  sf0 _ cbind(sum(yy0),sum(nn0)-sum(yy0)) #historical data
  sf0_rbind(sf0,sf[2,],sf[3,])
  #crude estimate of tau
for( i in 1 : length(yy0)){
if (yy0[i] == 0) yy0[i] _ .5}
tau.hat_sqrt(var(log( (yy0/nn0)/(1-yy0/nn0))))
return(tmp,sf,xx,sf0,tau.hat)
}
separate_function(){
  DD_make.data()
  glm.out       _ glm(sf ~ xx, family=binomial, data=DD)
  return(glm.out)
}
pooling_function(){
  DD_make.data()
  glm.out       _ glm(sf0 ~ xx, family=binomial, data=DD)
  return(glm.out)
}
##OUTPUT
summary(glm.separate)$coeff
                Value Std. Error   t value
(Intercept) -0.9766773  0.5427958 -1.799346
         xx -0.9451008  0.4777748 -1.978130

summary(glm.pooling)$coeff
                Value Std. Error    t value
(Intercept) -1.7059649  0.0664078 -25.689223
```



```
xx -0.4673848  0.2970020  -1.573675
```



## Computing

- $\gamma = (\alpha_1, \ldots, \alpha_{J+1}, \beta, \mu)$ vector of unknown parameters
- To draw posterior samples of $\gamma, \tau$:
  1. draw $\tau^l$ from $p(\tau \mid y)$ using inverse cdf or importance sampling
  2. run an iterative algorithm to estimate the mode and the curvature of $p(\gamma \mid \tau, y)$ given the just simulated value of $\tau$ $(\hat{\gamma}(\tau^l), V_\gamma(\tau^l))$
  3. draw $\gamma^l$ from a normal approximation based on the computed mode and curvature

$$\begin{aligned} \gamma^l \mid \tau^l, y &\simeq N(\hat{\gamma}(\tau^l), V_\gamma(\tau^l)) \\ V_\gamma(\tau^l)) &= [-L''(\hat{\gamma}(\tau^l))]^{-1} \end{aligned}$$



```
#find conditional mode
logpost_function(gamma,tau=1){
  yy_DD$tmp[,1]
  nn_DD$tmp[,2]
  xx_DD$xx
  tarone _ matrix(scan("tarone.txt"),byrow=T,ncol=2)
  yy0    _ tarone[,1]
  nn0    _ tarone[,2]
  J_length(yy0)-1
  alpha_gamma[(1:(J+1))]
  beta_gamma[J+2]
  mu_gamma[J+3]
 - (alpha[1:J]%*%yy0[1:J]-nn0[1:J]%*%log(1+exp(alpha[1:J]))+(J+1)*log(tau)
          - (1/(2*tau^2))*sum(alpha[1:J]-mu)^2
          + alpha[J+1]*sum(yy)
          + beta*xx%*%yy
          - nn%*%log(1+exp(alpha[J+1]+beta*xx)))
          #min(-logposterior)=max(logposterior)}
mode_ nlminb(start=c(rep(0,71),0,0),obj=logpost)}
```

MCMC methods

- For high dimensional problems to find the posterior mode and the curvature of the log posterior can be very hard
- an alternative here is to implement a Metropolis within Gibbs-Sampler, the full conditionals are:



$$
\begin{aligned}
\alpha \mid \beta, \mu, \tau, y &\sim \prod_{j=1}^{J} e^{\alpha_j y_{j0}} (1 + e^{\alpha_j})^{-n_{j0}} \\
&\times \prod_{j=1}^{J+1} \frac{1}{\tau} \exp\left(\frac{1}{2\tau^2}(\alpha_j - \mu)^2\right) \\
&\times \prod_{i=1}^{2} e^{(\alpha_{J+1} + \beta x_i) y_i} (1 + e^{\alpha_{J+1} + \beta x_i})^{-n_{j0}} \\
\beta \mid \alpha, \mu, \tau, y &\sim \prod_{i=1}^{2} e^{(\alpha_{J+1} + \beta x_i) y_i} \\
&\times \prod_{i=1}^{2} (1 + e^{\alpha_{J+1} + \beta x_i})^{-n_i} \\
\mu \mid \alpha, \beta, \tau, y &\sim N(\tfrac{1}{J+1} \textstyle\sum_{j=1}^{J+1} \alpha_j, \tfrac{\tau^2}{J+1}) \\
\tau^2 \mid \alpha, \beta\mu, &\sim IG(J, \tfrac{1}{2} \textstyle\sum_{j=1}^{J+1} (\alpha_j - \mu)^2)
\end{aligned}
$$



**Hierarchical Poisson Regression**
**Air Pollution Data**

- $y_t^c$ is the daily number of deaths in city $c$
- $x_t^c$ is the air pollution level ($PM_{10}$) in city $c$
- $w_t^c$ confounding variables such as weather, trend and seasonality
- $z^c$ city-specific covariates (such as poverty, SES, traffic)
- Goals of the analysis:
  1. Is there an association between air pollution and mortality within each city?
  2. Can we identify city-specific factors that might modify the association between air pollution and health?



  3. How can we estimate a national relative rate of mortality associated with exposure to air pollution?

**Modeling Approach**

- We assume the following hierarchical Model:

$$
\begin{aligned}
y_t^c &\sim \text{Poisson}(\mu_t^c) \\
\log y_t^c &= \beta_0^c + \beta_1^c x_t^c + \eta^c w_t^c \\
\beta_1^c &= \alpha_0 + \alpha_1 (z^c - \bar{z}^c) + N(0, \tau^2) \\
\boldsymbol{\alpha} &\sim N(\boldsymbol{\alpha}_0, D) \\
\tau^2 &\sim N(0, h) I_{\tau^2 > 0}
\end{aligned}
$$

We need to approximate

$$p(\boldsymbol{\beta}, \boldsymbol{\alpha}, \boldsymbol{\eta}, \tau^2 \mid y) = p(\boldsymbol{\beta}, \boldsymbol{\alpha}, \boldsymbol{\eta} \mid \tau^2, y) p(\tau^2 \mid y)$$

## Bayesian Computation

- Because of the large number of days with air pollution and mortality measurements within each city, we found that the MLE-based normal approximation to the likelihood is adequate, so we replaced the first stage of the hierarchical model with:

$$\begin{aligned}\hat{\beta}^c &\sim N(\beta^c, v^c)\\ \beta_1^c &= \alpha_0 + \alpha_1(z^c - \bar{z}^c) + N(0, \tau^2)\\ \boldsymbol{\alpha} &\sim N(\boldsymbol{\alpha}_0, D)\\ \tau^2 &\sim N(0, h)I_{\tau^2>0}\end{aligned}$$

the implementation of the Gibbs sampling is straightforward



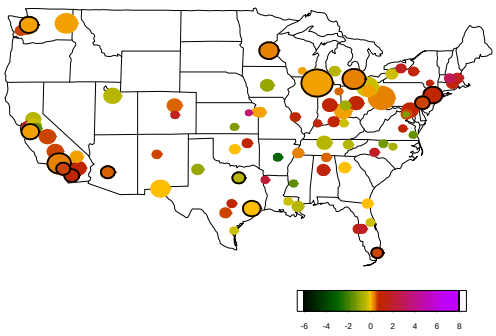


Figure 1: *Map of the 88 largest US cities with the 7 geographical regions of interest. The color scale is proportional to the estimated log relative rates of mortality, which range between $-4\%$ and $4\%$ increase in mortality per 10 $\mu g/m^3$ increase in $PM_{10}$. The circles' areas are proportional to the statistical precisions of the estimates. The larger circles show less statistical uncertainty. The circles with the black outline denote the relative rates that are statistically significant different than zero.*



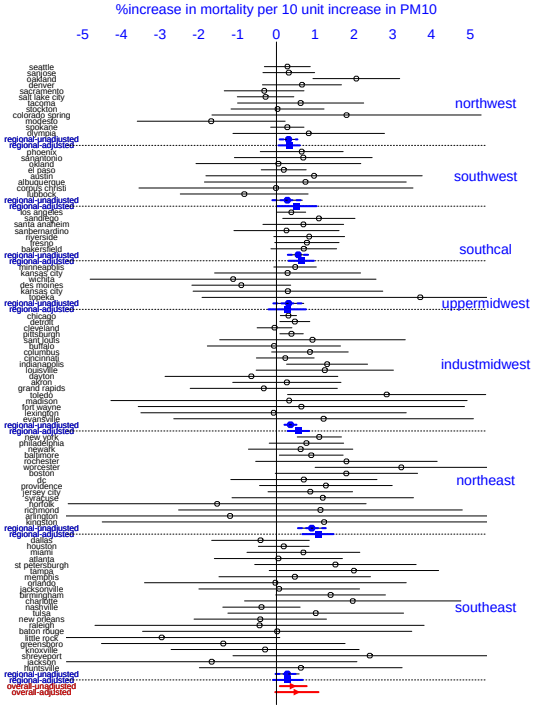


Figure 2: *Maximum likelihood estimates and 95% confidence intervals of the log-relative rates of mortality per $10\mu g/m^3$ increase in $PM_{10}$ for each location. The solid and the square circles with the bold segments denote the posterior means and 95% posterior intervals of the pooled regional effects without and with covariate adjustment, respectively. At the bottom, marked with triangles and bold segments, are the overall effects for $PM_{10}$ for all the cities without and with covariate adjustment.*



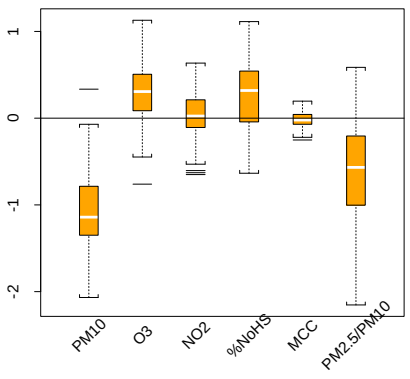


Figure 3: *Boxplots of the posterior distributions of the second-stage regression coefficients.*

Table 2: *Second-stage variables and the rationale for their inclusion in the model. % missing denotes the percentage of cities with a missing data.*

| Predictors | Primary reasons for inclusion | % missing |
|---|---|---|
| $\log \bar{PM}_{10}$ | Possibility of a saturation effect | 0 |
| $\log \bar{O}_3$ | To explore modification of the effect of one pollutant to another | 11 |
| $\log \bar{NO}_2$ | To explore modification of the effect of one pollutant to another | 35 |
| log %NoHS | Potential heterogeneity of the effects associated with socio-demographic factors | 0 |
| logit$MCC$ | Potential heterogeneity of the effects associated with the varying quality of the exposure | 21 |
| $PM_{2.5}/PM_{10}$ | Hypothesized health effects of fine particles | 0 |

$\bar{PM}_{10}$, $\bar{O}_3$ and $\bar{NO}_2$ denote the mean level of $PM_{10}$, mean level of ozone ($O_3$), and the mean level of nitrogen dioxide ($NO_2$) over the period 1987-1994; %NoHS is percentage of persons lacking a high school degree; $MCC$ is median of all pairwise correlations of the $PM_{10}$ measurements among the location specific monitors.

Table 3: *Posterior quantiles of the unadjusted overall $PM_{10}$ effect on total mortality, $\alpha_0$, under four specifications of prior distributions on $\tau^2$.*

| Prior for $\tau^2$ | Scenario | 5% | 25% | 50% | 75% | 95% |
|---|---|---|---|---|---|---|
| $\tau^2 \sim N(0,1)I_{\tau^2>0}$ | Baseline | 0.06 | 0.33 | 0.43 | 0.53 | 0.77 |
| $\tau^2 \sim N(0,1)I_{\tau^2>0}$ | 1 | 0.09 | 0.34 | 0.47 | 0.58 | 0.85 |
| $\tau^2 \sim IG(3,1)$ | 2 | 0.02 | 0.29 | 0.42 | 0.55 | 0.78 |
| $\tau^2 \sim IG(3,1)$ | 3 | 0.00 | 0.30 | 0.45 | 0.60 | 0.90 |

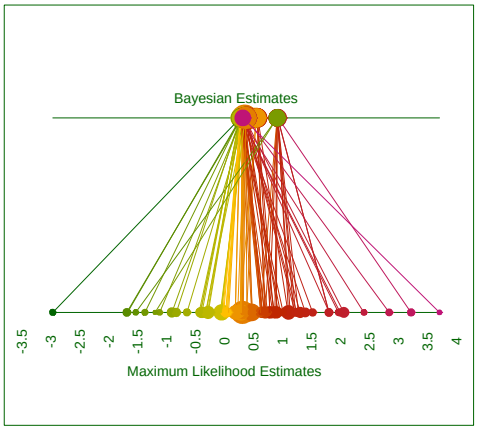


Figure 5: *Maximum Likelihood (bottom) and Bayesian posterior estimates (top) of the relative risks of mortality for the largest 88 cities. Sizes of the circles are proportional to the statistical precisions.*

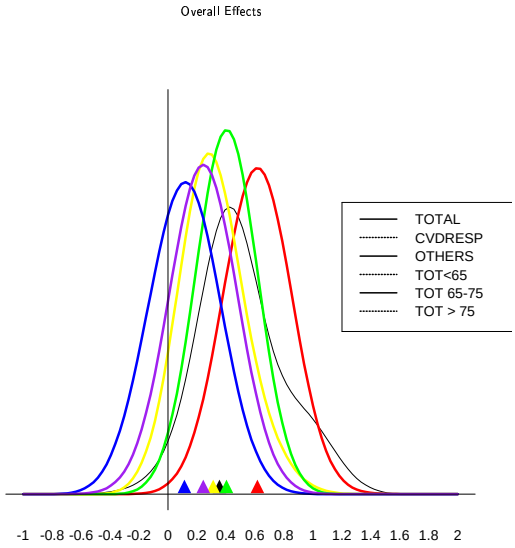


Figure 4: *Marginal Posterior distributions of the overall relative rates of mortality associated with exposure to air pollution for different health outcomes*