Example using R: Heart Valves Study

- Goal: Show that the thrombogenicity rate (TR) is less than two times the objective performance criterion

Example using R: Heart Valves Study

- Goal: Show that the thrombogenicity rate (TR) is less than two times the objective performance criterion
- Data: From both the current study and a previous study on a similar product (St. Jude mechanical valve).

Example using R: Heart Valves Study

- Goal: Show that the thrombogenicity rate (TR) is less than two times the objective performance criterion
- Data: From both the current study and a previous study on a similar product (St. Jude mechanical valve).
- Model: Let T be the total number of patient-years of followup, and θ be the TR per year. We assume the number of thrombogenicity events $Y \sim Poisson(\theta T)$:

$$f(y|\theta) = \frac{e^{-\theta T}(\theta T)^y}{y!} .$$

Example using R: Heart Valves Study

- Goal: Show that the thrombogenicity rate (TR) is less than two times the objective performance criterion
- Data: From both the current study and a previous study on a similar product (St. Jude mechanical valve).
- Model: Let T be the total number of patient-years of followup, and θ be the TR per year. We assume the number of thrombogenicity events $Y \sim Poisson(\theta T)$:

$$f(y|\theta) = \frac{e^{-\theta T}(\theta T)^y}{y!} \; .$$

- Prior: Assume a $Gamma(\alpha, \beta)$ prior for θ:

$$p(\theta) = \frac{\theta^{\alpha-1}e^{-\theta/\beta}}{\Gamma(\alpha)\beta^\alpha} \, , \theta > 0 \; .$$

Heart Valves Study

- The gamma prior is conjugate with the likelihood, so the posterior emerges in closed form:

$$\begin{aligned} p(\theta|y) &\propto \theta^{y+\alpha-1} e^{-\theta(T+1/\beta)} \\ &\propto Gamma(y+\alpha, (T+1/\beta)^{-1}) \ . \end{aligned}$$

The study objective is met if

$$P(\theta < 2 \times OPC \mid y) \geq 0.95 \ ,$$

where $OPC = \theta_0 = 0.038$.

Heart Valves Study

- The gamma prior is conjugate with the likelihood, so the posterior emerges in closed form:

$$\begin{aligned} p(\theta|y) &\propto \theta^{y+\alpha-1} e^{-\theta(T+1/\beta)} \\ &\propto Gamma(y+\alpha, (T+1/\beta)^{-1}) \ . \end{aligned}$$

  The study objective is met if

$$P(\theta < 2 \times OPC \mid y) \geq 0.95 \ ,$$

  where $OPC = \theta_0 = 0.038$.

- Prior selection: Our gamma prior has mean $M = \alpha\beta$ and variance $V = \alpha\beta^2$. This means that if we specify M and V, we can solve for α and β as

$$\alpha = M^2/V \text{ and } \beta = V/M \ .$$

Heart Valves Study

- A few possibilities for prior parameters:

Heart Valves Study

- A few possibilities for prior parameters:
  - Suppose we set $M = \theta_0 = 0.038$ and $\sqrt{V} = 2\theta_0$ (so that 0 is two standard deviations below the mean). Then $\alpha = 0.25$ and $\beta = 0.152$, a rather vague prior.

Heart Valves Study

- A few possibilities for prior parameters:
  - Suppose we set $M = \theta_0 = 0.038$ and $\sqrt{V} = 2\theta_0$ (so that 0 is two standard deviations below the mean). Then $\alpha = 0.25$ and $\beta = 0.152$, a rather vague prior.
  - Suppose we set $M = 98/5891 = .0166$, the overall value from the St. Jude studies, and $\sqrt{V} = M$ (so 0 is one sd below the mean). Then $\alpha = 1$ and $\beta = 0.0166$, a moderate (exponential) prior.

Heart Valves Study

- A few possibilities for prior parameters:
  - Suppose we set $M = \theta_0 = 0.038$ and $\sqrt{V} = 2\theta_0$ (so that 0 is two standard deviations below the mean). Then $\alpha = 0.25$ and $\beta = 0.152$, a rather vague prior.
  - Suppose we set $M = 98/5891 = .0166$, the overall value from the St. Jude studies, and $\sqrt{V} = M$ (so 0 is one sd below the mean). Then $\alpha = 1$ and $\beta = 0.0166$, a moderate (exponential) prior.
  - Suppose we set $M = 98/5891 = .0166$ again, but set $\sqrt{V} = M/2$. This is a rather informative prior.

Heart Valves Study

- A few possibilities for prior parameters:
    - Suppose we set $M = \theta_0 = 0.038$ and $\sqrt{V} = 2\theta_0$ (so that 0 is two standard deviations below the mean). Then $\alpha = 0.25$ and $\beta = 0.152$, a rather vague prior.
    - Suppose we set $M = 98/5891 = .0166$, the overall value from the St. Jude studies, and $\sqrt{V} = M$ (so 0 is one sd below the mean). Then $\alpha = 1$ and $\beta = 0.0166$, a moderate (exponential) prior.
    - Suppose we set $M = 98/5891 = .0166$ again, but set $\sqrt{V} = M/2$. This is a rather informative prior.
- We also consider event counts that are lower (1), about the same (3), and much higher (20) than for St. Jude.

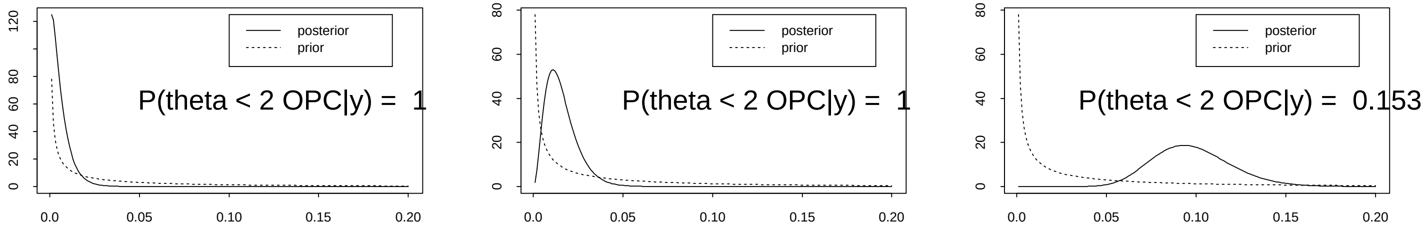
Heart Valves Study

- A few possibilities for prior parameters:
  - Suppose we set $M = \theta_0 = 0.038$ and $\sqrt{V} = 2\theta_0$ (so that 0 is two standard deviations below the mean). Then $\alpha = 0.25$ and $\beta = 0.152$, a rather vague prior.
  - Suppose we set $M = 98/5891 = .0166$, the overall value from the St. Jude studies, and $\sqrt{V} = M$ (so 0 is one sd below the mean). Then $\alpha = 1$ and $\beta = 0.0166$, a moderate (exponential) prior.
  - Suppose we set $M = 98/5891 = .0166$ again, but set $\sqrt{V} = M/2$. This is a rather informative prior.
- We also consider event counts that are lower (1), about the same (3), and much higher (20) than for St. Jude.
- The study objective is not met with the “bad” data – *unless* the posterior is “rescued” by the informative prior (lower right corner, next page).

Heart Valves Study

Priors and posteriors, Heart Valves ADVANTAGE study, Poisson-gamma model
for various prior (M, sd) and data (y) values; T = 200 , 2 OPC = 0.076

vague
prior

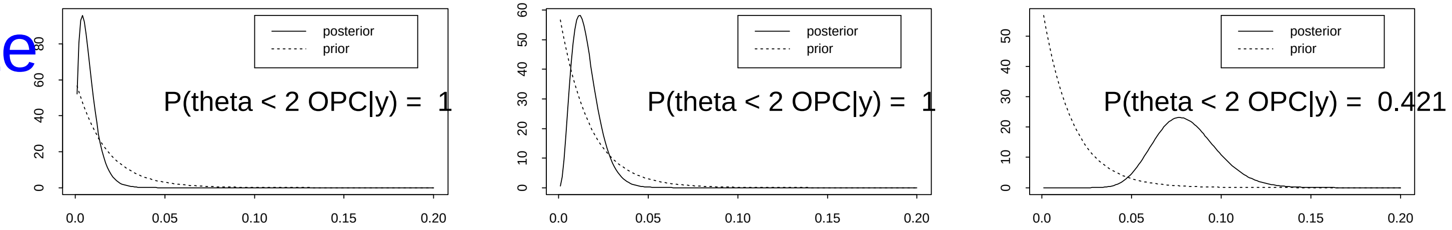


M, sd = 0.038 0.076 ; Y = 1

M, sd = 0.038 0.076 ; Y = 3

M, sd = 0.038 0.076 ; Y = 20

moderate
prior

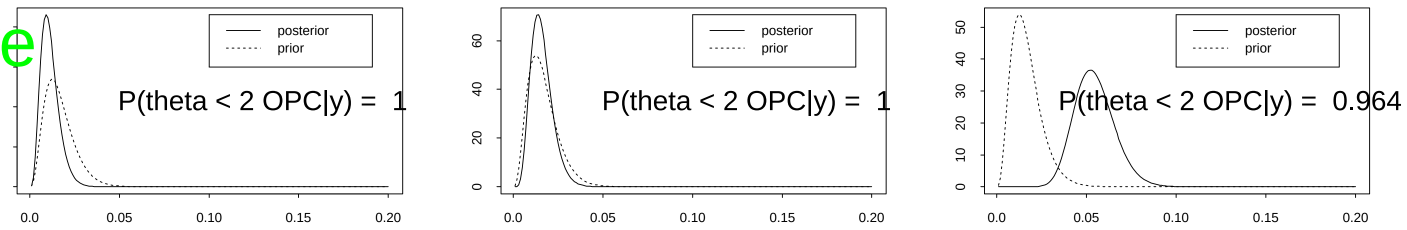


M, sd = 0.017 0.017 ; Y = 1

M, sd = 0.017 0.017 ; Y = 3

M, sd = 0.017 0.017 ; Y = 20

informative
prior



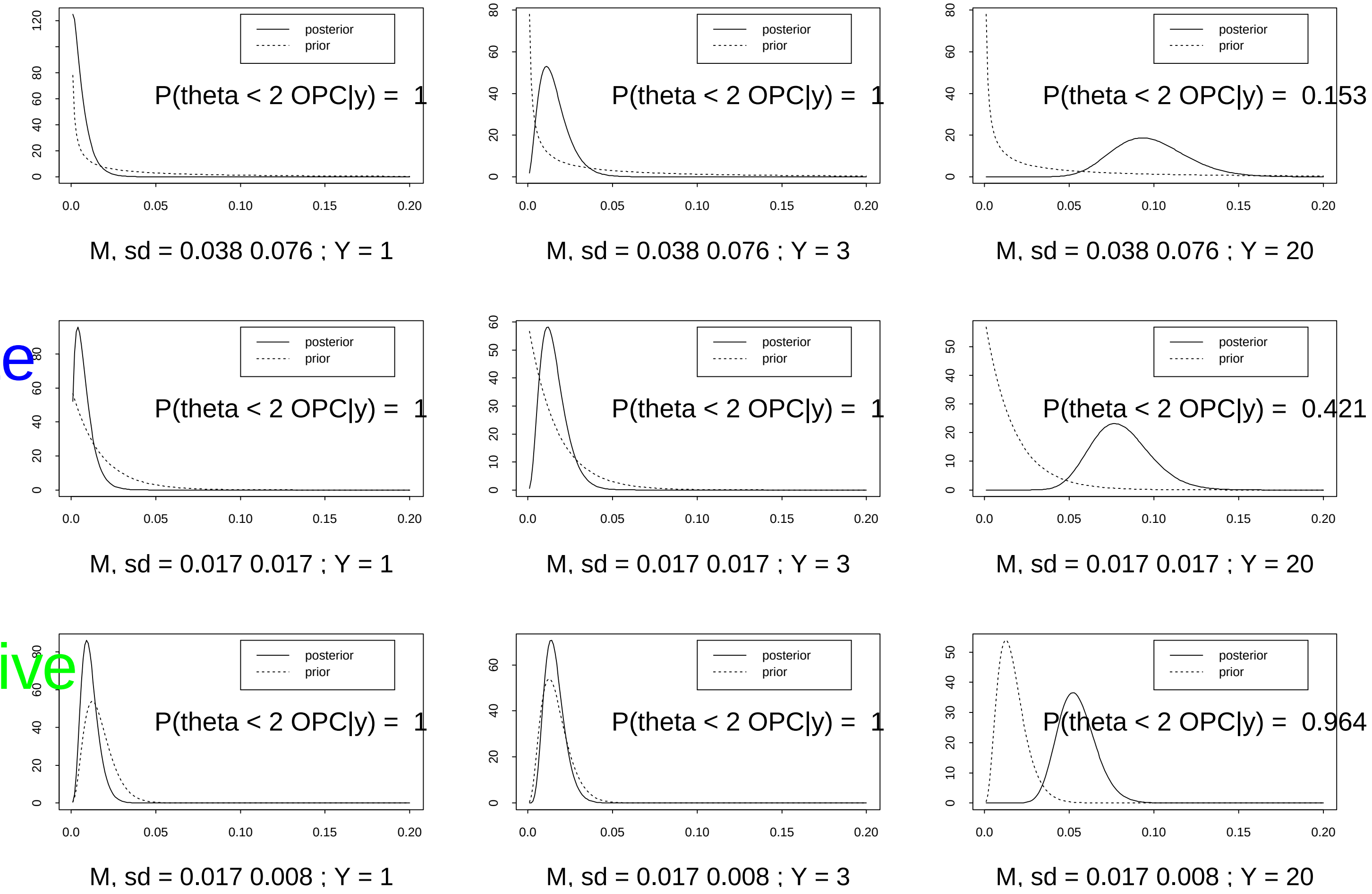
M, sd = 0.017 0.008 ; Y = 1

M, sd = 0.017 0.008 ; Y = 3

M, sd = 0.017 0.008 ; Y = 20

Heart Valves Study

Priors and posteriors, Heart Valves ADVANTAGE study, Poisson-gamma model
for various prior (M, sd) and data (y) values; T = 200 , 2 OPC = 0.076



vague
prior

moderate
prior

informative
prior

- S code to create this plot is available in www.biostat.umn.edu/∼brad/hv.S
– try it yourself in S-plus or R (http://cran.r-project.org)

Alternate hierarchical models

- One might be uncomfortable with our implicit assumption that the TR is the same in both studies. To handle this, extend to a hierarchical model:

  $$Y_i \sim Poisson(\theta_i T_i),\ i = 1, 2,$$

  where $i = 1$ for St. Jude, and $i = 2$ for the new study.

Alternate hierarchical models

- One might be uncomfortable with our implicit assumption that the TR is the same in both studies. To handle this, extend to a hierarchical model:

  $$Y_i \sim Poisson(\theta_i T_i), \; i = 1, 2,$$

  where $i = 1$ for St. Jude, and $i = 2$ for the new study.
- Borrow strength between studies by assuming

  $$\theta_i \overset{iid}{\sim} Gamma(\alpha, \beta),$$

  i.e., the two TR's are exchangeable, but not identical.

Alternate hierarchical models

- One might be uncomfortable with our implicit assumption that the TR is the same in both studies. To handle this, extend to a hierarchical model:

  $$Y_i \sim Poisson(\theta_i T_i),\ i = 1, 2,$$

  where $i = 1$ for St. Jude, and $i = 2$ for the new study.
- Borrow strength between studies by assuming

  $$\theta_i \overset{iid}{\sim} Gamma(\alpha, \beta),$$

  i.e., the two TR's are exchangeable, but not identical.
- We now place a third stage prior on α and β, say

  $$\alpha \sim Exp(a) \text{ and } \beta \sim IG(c, d).$$

Alternate hierarchical models

- One might be uncomfortable with our implicit assumption that the TR is the same in both studies. To handle this, extend to a hierarchical model:

  $$Y_i \sim Poisson(\theta_i T_i),\ i = 1, 2,$$

  where $i = 1$ for St. Jude, and $i = 2$ for the new study.
- Borrow strength between studies by assuming

  $$\theta_i \overset{iid}{\sim} Gamma(\alpha, \beta),$$

  i.e., the two TR's are exchangeable, but not identical.
- We now place a third stage prior on α and β, say

  $$\alpha \sim Exp(a) \text{ and } \beta \sim IG(c, d).$$
- Fit in WinBUGS using the pump example as a guide!

Pump Example

Example 2.5 revisited again!

$$
\begin{gathered}
Y_i|\theta_i \overset{ind}{\sim} Poisson(\theta_i t_i), \\
\theta_i \overset{ind}{\sim} G(\alpha, \beta), \\
\alpha \sim Exp(\mu),\ \beta \sim IG(c, d),
\end{gathered}
$$

$i = 1, \ldots, k$, where μ, c, d, and the t_i are known, and Exp denotes the exponential distribution.

Pump Example

Example 2.5 revisited again!

$$
\begin{gathered}
Y_i|\theta_i \overset{ind}{\sim} Poisson(\theta_i t_i), \\
\theta_i \overset{ind}{\sim} G(\alpha, \beta), \\
\alpha \sim Exp(\mu),\ \beta \sim IG(c, d),
\end{gathered}
$$

$i = 1, \ldots, k$, where μ, c, d, and the t_i are known, and Exp denotes the exponential distribution.

- We apply this model to a dataset giving the numbers of pump failures, Y_i, observed in t_i thousands of hours for $k = 10$ different systems of a certain nuclear power plant.

Pump Example

Example 2.5 revisited again!

$$Y_i|\theta_i \overset{ind}{\sim} Poisson(\theta_i t_i),$$
$$\theta_i \overset{ind}{\sim} G(\alpha, \beta),$$
$$\alpha \sim Exp(\mu),\ \beta \sim IG(c, d),$$

$i = 1, \ldots, k$, where μ, c, d, and the t_i are known, and Exp denotes the exponential distribution.

- We apply this model to a dataset giving the numbers of pump failures, Y_i, observed in t_i thousands of hours for $k = 10$ different systems of a certain nuclear power plant.
- The observations are listed in increasing order of raw failure rate $r_i = Y_i/t_i$, the classical point estimate of the true failure rate θ_i for the i^{th} system.

Pump Data

i	Y_i	t_i	r_i
1	5	94.320	.053
2	1	15.720	.064
3	5	62.880	.080
4	14	125.760	.111
5	3	5.240	.573
6	19	31.440	.604
7	1	1.048	.954
8	1	1.048	.954
9	4	2.096	1.910
10	22	10.480	2.099

Hyperparameters: We choose the values $\mu = 1$, $c = 0.1$, and $d = 1.0$, resulting in reasonably vague hyperpriors for α and β.

Pump Example

- Recall that the full conditional distributions for the θ_i and β are available in closed form (gamma and inverse gamma, respectively), but that no conjugate prior for α exists.

Pump Example

- Recall that the full conditional distributions for the θ_i and β are available in closed form (gamma and inverse gamma, respectively), but that no conjugate prior for α exists.
- However, the full conditional for α,

$$
\begin{aligned}
p(\alpha|\beta, \{\theta_i\}, \mathbf{y}) &\propto \left[\prod_{i=1}^{k} g(\theta_i|\alpha, \beta)\right] h(\alpha) \\
&\propto \left[\prod_{i=1}^{k} \frac{\theta_i^{\alpha-1}}{\Gamma(\alpha)\beta^{\alpha}}\right] e^{-\alpha/\mu}
\end{aligned}
$$

can be shown to be log-concave in α. Thus `WinBUGS` uses adaptive rejection sampling for this parameter.

WinBUGS code to fit this model

```
model {
    for (i in 1:k) {
          theta[i] ~ dgamma(alpha,beta)
          lambda[i] <- theta[i]*t[i]
          Y[i] ~ dpois(lambda[i])
    }
    alpha ~ dexp(1.0)
    beta ~ dgamma(0.1, 1.0)
}

  DATA:
list(k = 10, Y = c(5, 1, 5, 14, 3, 19, 1, 1, 4, 22),
     t = c(94.320, 15.72, 62.88, 125.76, 5.24, 31.44,
            1.048, 1.048, 2.096, 10.48))

  INITS:
list(theta=c(1,1,1,1,1,1,1,1,1,1), alpha=1, beta=1)
```

Pump Example Results

Results from running 1000 burn-in samples, followed by a “production” run of 10,000 samples (single chain):

node	mean	sd	MC error	2.5%	median	97.5%
alpha	0.7001	0.2699	0.004706	0.2851	0.6634	1.338
beta	0.929	0.5325	0.00978	0.1938	0.8315	2.205
theta[1]	0.0598	0.02542	2.68E-4	0.02128	0.05627	0.1195
theta[5]	0.6056	0.315	0.003087	0.1529	0.5529	1.359
theta[6]	0.6105	0.1393	0.0014	0.3668	0.5996	0.9096
theta[10]	1.993	0.4251	0.004915	1.264	1.958	2.916

Pump Example Results

Results from running 1000 burn-in samples, followed by a “production” run of 10,000 samples (single chain):

node	mean	sd	MC error	2.5%	median	97.5%
alpha	0.7001	0.2699	0.004706	0.2851	0.6634	1.338
beta	0.929	0.5325	0.00978	0.1938	0.8315	2.205
theta[1]	0.0598	0.02542	2.68E-4	0.02128	0.05627	0.1195
theta[5]	0.6056	0.315	0.003087	0.1529	0.5529	1.359
theta[6]	0.6105	0.1393	0.0014	0.3668	0.5996	0.9096
theta[10]	1.993	0.4251	0.004915	1.264	1.958	2.916

- Note that while θ_5 and θ_6 have very similar posterior means, the latter posterior is much narrower (smaller sd).

Pump Example Results

Results from running 1000 burn-in samples, followed by a “production” run of 10,000 samples (single chain):

node	mean	sd	MC error	2.5%	median	97.5%
alpha	0.7001	0.2699	0.004706	0.2851	0.6634	1.338
beta	0.929	0.5325	0.00978	0.1938	0.8315	2.205
theta[1]	0.0598	0.02542	2.68E-4	0.02128	0.05627	0.1195
theta[5]	0.6056	0.315	0.003087	0.1529	0.5529	1.359
theta[6]	0.6105	0.1393	0.0014	0.3668	0.5996	0.9096
theta[10]	1.993	0.4251	0.004915	1.264	1.958	2.916

- Note that while θ_5 and θ_6 have very similar posterior means, the latter posterior is much narrower (smaller sd).
- This is because, while the crude failure rates for the two pumps are similar, the latter is based on a far greater number of hours of observation ($t_6 = 31.44$, while $t_5 = 5.24$). Hence we “know” more about pump 6!

PK Example

- Wakefield et al. (1994) consider a dataset for which

  $$\begin{aligned} Y_{ij} &= \text{plasma concentration of the drug Cadralazine} \\ x_{ij} &= \text{time elapsed since dose given} \end{aligned}$$

  where $i = 1, \ldots, 10$ indexes the patient, while $j = 1, \ldots, n_i$ indexes the observations, $5 \leq n_i \leq 8$.

PK Example

- Wakefield et al. (1994) consider a dataset for which

  $$\begin{aligned} Y_{ij} &= \text{plasma concentration of the drug Cadralazine} \\ x_{ij} &= \text{time elapsed since dose given} \end{aligned}$$

  where $i = 1, \ldots, 10$ indexes the patient, while $j = 1, \ldots, n_i$ indexes the observations, $5 \leq n_i \leq 8$.

- Attempt to fit the one-compartment nonlinear pharmacokinetic (PK) model,

  $$\eta_{ij}(x_{ij}) = 30\alpha_i^{-1} \exp(-\beta_i x_{ij}/\alpha_i) \; .$$

  where $\eta_{ij}(x_{ij})$ is the mean plasma concentration at time x_{ij}.

PK Example

- This model is best fit on the log scale, i.e.

$$Z_{ij} \equiv \log Y_{ij} = \log \eta_{ij}(x_{ij}) + \epsilon_{ij} \ ,$$

  where $\epsilon_{ij} \overset{ind}{\sim} N(0, \tau_i)$.

PK Example

- This model is best fit on the log scale, i.e.

  $$Z_{ij} \equiv \log Y_{ij} = \log \eta_{ij}(x_{ij}) + \epsilon_{ij} \ ,$$

  where $\epsilon_{ij} \overset{ind}{\sim} N(0, \tau_i)$.
- The mean structure for the Z_{ij}'s thus emerges as

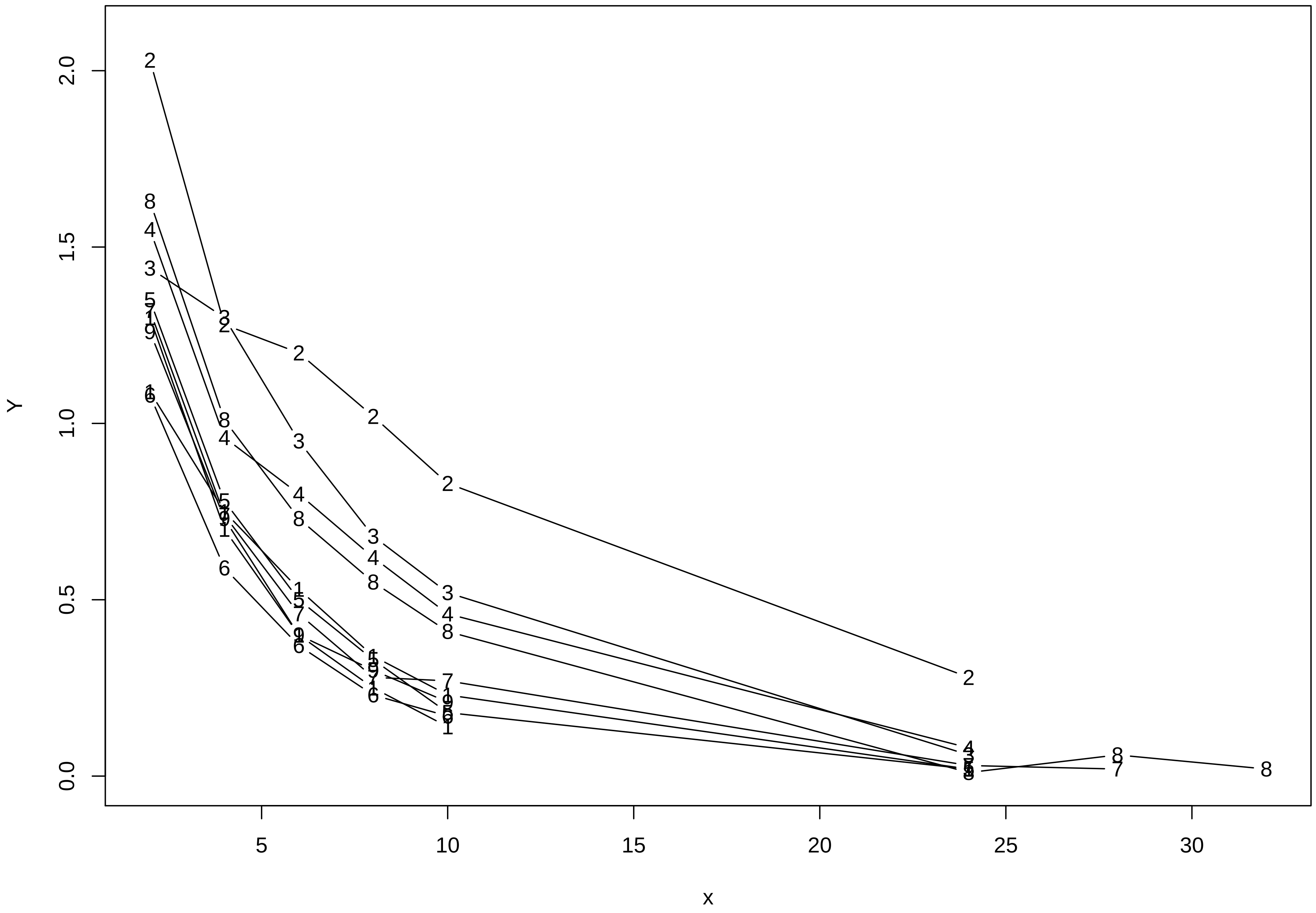
  $$\begin{aligned} \log \eta_{ij}(x_{ij}) &= \log\left[30\alpha_i^{-1} \exp(-\beta_i x_{ij}/\alpha_i)\right] \\ &= \log 30 - \log \alpha_i - \beta_i x_{ij}/\alpha_i \\ &= \log 30 - a_i - \exp(b_i - a_i) x_{ij} \ , \end{aligned}$$

  where $a_i = \log \alpha_i$ and $b_i = \log \beta_i$.

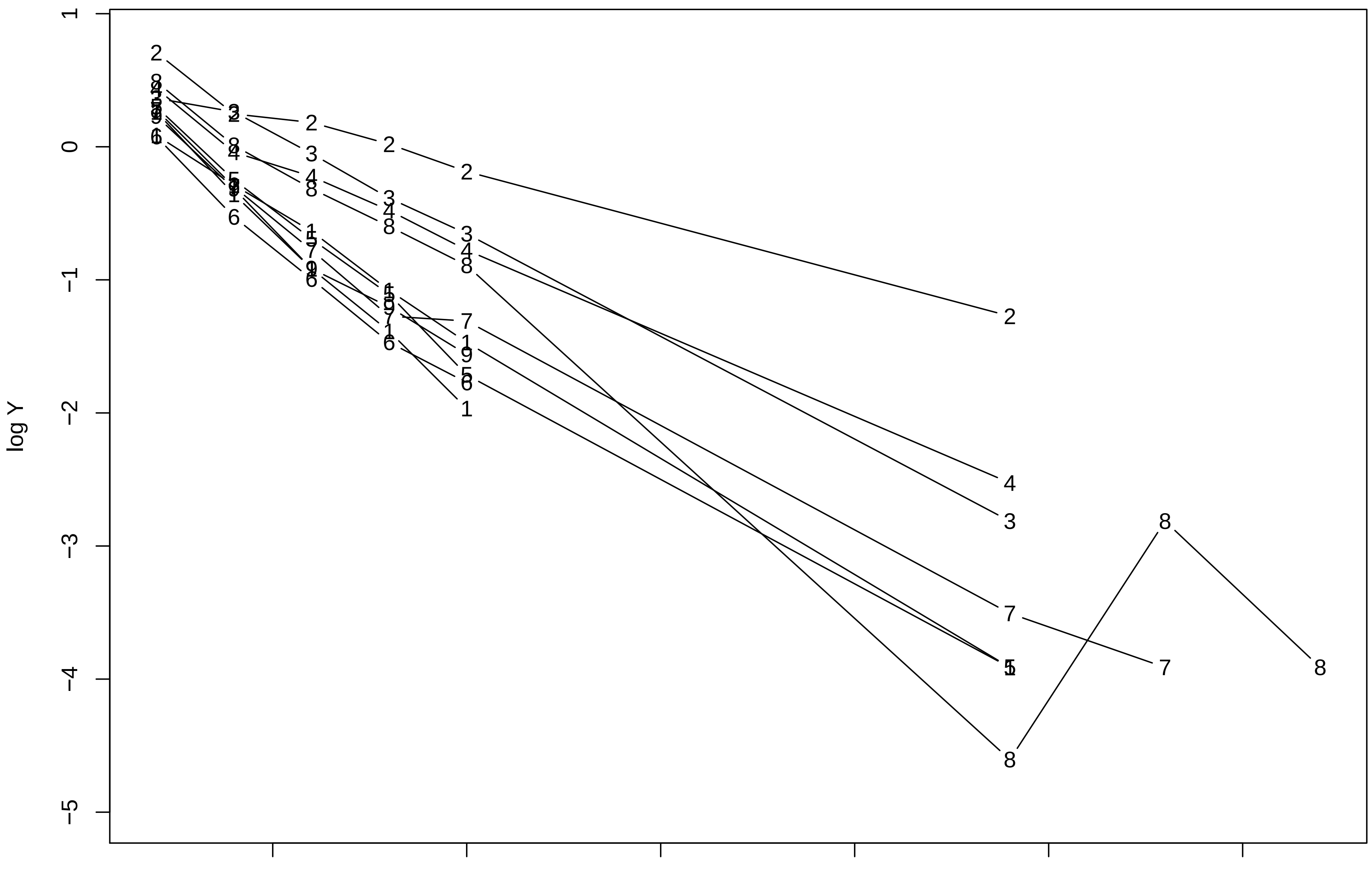
PK Data

	no. of hours following drug administration, x							
patient	2	4	6	8	10	24	28	32
1	1.09	0.75	0.53	0.34	0.23	0.02	–	–
2	2.03	1.28	1.20	1.02	0.83	0.28	–	–
3	1.44	1.30	0.95	0.68	0.52	0.06	–	–
4	1.55	0.96	0.80	0.62	0.46	0.08	–	–
5	1.35	0.78	0.50	0.33	0.18	0.02	–	–
6	1.08	0.59	0.37	0.23	0.17	–	–	–
7	1.32	0.74	0.46	0.28	0.27	0.03	0.02	–
8	1.63	1.01	0.73	0.55	0.41	0.01	0.06	0.02
9	1.26	0.73	0.40	0.30	0.21	–	–	–
10	1.30	0.70	0.40	0.25	0.14	–	–	–

PK Data, original scale



PK Data, log scale



PK Example

- For the subject-specific random effects $\boldsymbol{\theta}_i \equiv (a_i, b_i)'$, assume

$$\boldsymbol{\theta}_i \overset{iid}{\sim} N_2(\boldsymbol{\mu}, \Omega)\ ,\ \text{where } \boldsymbol{\mu} = (\mu_a, \mu_b)\ .$$

PK Example

- For the subject-specific random effects $\boldsymbol{\theta}_i \equiv (a_i, b_i)'$, assume

$$\boldsymbol{\theta}_i \overset{iid}{\sim} N_2(\boldsymbol{\mu}, \Omega)\ , \text{ where } \boldsymbol{\mu} = (\mu_a, \mu_b)\ .$$

- Usual conjugate prior specification:

$$\begin{aligned} \boldsymbol{\mu} &\sim N_2(\boldsymbol{\lambda}, C) \\ \tau_i &\overset{iid}{\sim} G(\nu_0/2\ ,\ \nu_0\tau_0/2) \\ \Omega &\sim \text{Wishart}((\rho R)^{-1}, \rho) \end{aligned}$$

PK Example

- For the subject-specific random effects $\boldsymbol{\theta}_i \equiv (a_i, b_i)'$, assume

$$\boldsymbol{\theta}_i \overset{iid}{\sim} N_2(\boldsymbol{\mu}, \Omega) \ , \text{ where } \boldsymbol{\mu} = (\mu_a, \mu_b) \ .$$

- Usual conjugate prior specification:

$$\begin{aligned} \boldsymbol{\mu} &\sim N_2(\boldsymbol{\lambda}, C) \\ \tau_i &\overset{iid}{\sim} G(\nu_0/2 \ , \ \nu_0\tau_0/2) \\ \Omega &\sim \text{Wishart}((\rho R)^{-1}, \rho) \end{aligned}$$

- Note that the $\boldsymbol{\theta}_i$ full conditional distributions are:
  - not simple conjugate forms
  - not guaranteed to be log-concave

  Thus, the Metropolis capability of `WinBUGS` is required.

PK Results (WinBUGS vs. Fortran)

parameter	BUGS			Sargent et al. (2000)		
	mean	sd	lag 1 acf	mean	sd	lag 1 acf
a_1	2.956	0.0479	0.969	2.969	0.0460	0.947
a_2	2.692	0.0772	0.769	2.708	0.0910	0.808
a_7	2.970	0.1106	0.925	2.985	0.1360	0.938
a_8	2.828	0.1417	0.828	2.838	0.1863	0.934
b_1	1.259	0.0335	0.972	1.268	0.0322	0.951
b_2	0.234	0.0648	0.661	0.239	0.0798	0.832
b_7	1.157	0.0879	0.899	1.163	0.1055	0.925
b_8	0.936	0.1458	0.759	0.941	0.1838	0.932
τ_1	362.4	260.4	0.313	380.8	268.8	0.220
τ_2	84.04	57.60	0.225	81.40	58.41	0.255
τ_7	18.87	12.07	0.260	15.82	11.12	0.237
τ_8	2.119	1.139	0.085	1.499	0.931	0.143
$Y_{2,8}$	0.1338	0.0339	0.288	0.1347	0.0264	–
$Y_{7,8}$	0.00891	0.00443	0.178	0.00884	0.00255	–

Lip Cancer Example

- Consider the spatial disease mapping model:

$$\begin{aligned} Y_i \mid \mu_i &\overset{ind}{\sim} Po\left(E_i\, e^{\mu_i}\right)\ , \quad \text{where} \\ Y_i &= \text{observed disease count,} \\ E_i &= \text{expected count (known), and} \\ \mu_i &= \mathbf{x}_i'\boldsymbol{\beta} + \theta_i + \phi_i \end{aligned}$$

Lip Cancer Example

- Consider the spatial disease mapping model:

$$\begin{array}{rcl} Y_i \mid \mu_i & \overset{ind}{\sim} & Po\left(E_i\, e^{\mu_i}\right)\ , \quad \text{where} \\ Y_i & = & \text{observed disease count,} \\ E_i & = & \text{expected count (known), and} \\ \mu_i & = & \mathbf{x}_i'\boldsymbol{\beta} + \theta_i + \phi_i \end{array}$$

- The $\mathbf{x}_i$ are explanatory spatial covariates; typically $\boldsymbol{\beta}$ is assigned a flat prior.

Lip Cancer Example

- Consider the spatial disease mapping model:

$$\begin{aligned} Y_i \mid \mu_i &\overset{ind}{\sim} Po\left(E_i\, e^{\mu_i}\right)\ , \quad \text{where} \\ Y_i &= \text{observed disease count,} \\ E_i &= \text{expected count (known), and} \\ \mu_i &= \mathbf{x}_i'\boldsymbol{\beta} + \theta_i + \phi_i \end{aligned}$$

- The $\mathbf{x}_i$ are explanatory spatial covariates; typically $\boldsymbol{\beta}$ is assigned a flat prior.
- Note the mean structure also contains two sets of random effects! The first, θ_i, capture heterogeneity among the regions via

$$\theta_i \overset{iid}{\sim} N(0\ ,\ 1/\tau_h)\ .$$

Lip Cancer Example

- The second set, ϕ_i, capture regional clustering via a conditionally autoregressive (CAR) prior,

  $$\phi_i \,|\, \phi_{j \neq i} \sim N(\bar{\phi}_i \,,\, 1/(\tau_c m_i)) \;,$$

  where m_i is the number of “neighbors” of region i, and $\bar{\phi}_i = m_i^{-1} \Sigma_{j \in \partial_i} \, \phi_j$.

Lip Cancer Example

- The second set, ϕ_i, capture regional clustering via a conditionally autoregressive (CAR) prior,

  $$\phi_i \,|\, \phi_{j \neq i} \sim N(\bar{\phi}_i \;,\; 1/(\tau_c m_i)) \;,$$

  where m_i is the number of “neighbors” of region i, and $\bar{\phi}_i = m_i^{-1} \Sigma_{j \in \partial_i} \, \phi_j$.
- The CAR prior is translation invariant, so typically we insist $\Sigma_{i=1}^{I} \phi_i = 0$ (imposed numerically after each MCMC iteration). Still, Y_i cannot inform about θ_i or ϕ_i, but only about their sum $\eta_i = \theta_i + \phi_i$.

Lip Cancer Example

- The second set, ϕ_i, capture regional clustering via a conditionally autoregressive (CAR) prior,

  $$\phi_i \,|\, \phi_{j \neq i} \sim N(\bar{\phi}_i \,,\, 1/(\tau_c m_i)) \,,$$

  where m_i is the number of "neighbors" of region i, and $\bar{\phi}_i = m_i^{-1} \Sigma_{j \in \partial_i} \phi_j$.
- The CAR prior is translation invariant, so typically we insist $\Sigma_{i=1}^{I} \phi_i = 0$ (imposed numerically after each MCMC iteration). Still, Y_i cannot inform about θ_i or ϕ_i, but only about their sum $\eta_i = \theta_i + \phi_i$.
- Making the reparametrization from $(\boldsymbol{\theta}, \boldsymbol{\phi})$ to $(\boldsymbol{\theta}, \boldsymbol{\eta})$, we have the joint posterior

  $$p(\boldsymbol{\theta}, \boldsymbol{\eta} | \mathbf{y}) \propto L(\boldsymbol{\eta}; \mathbf{y}) p(\boldsymbol{\theta}) p(\boldsymbol{\eta} - \boldsymbol{\theta}).$$

Lip Cancer Example

- This means that

$$p(\theta_i \mid \theta_{j\neq i}, \boldsymbol{\eta}, \mathbf{y}) \propto p(\theta_i)\, p(\eta_i - \theta_i \mid \{\eta_j - \theta_j\}_{j\neq i}) \ .$$

Lip Cancer Example

- This means that

$$p(\theta_i \mid \theta_{j \neq i}, \boldsymbol{\eta}, \mathbf{y}) \propto p(\theta_i)\, p(\eta_i - \theta_i \mid \{\eta_j - \theta_j\}_{j \neq i}) \ .$$

- Since this distribution is free of the data $\mathbf{y}$, the θ_i are Bayesianly unidentified (and so are the ϕ_i).

Lip Cancer Example

- This means that

$$p(\theta_i \mid \theta_{j\neq i}, \boldsymbol{\eta}, \mathbf{y}) \propto p(\theta_i)\, p(\eta_i - \theta_i \mid \{\eta_j - \theta_j\}_{j\neq i}) \ .$$

- Since this distribution is free of the data $\mathbf{y}$, the θ_i are Bayesianly unidentified (and so are the ϕ_i).
- BUT this does not preclude Bayesian learning about θ_i; this would instead require

$$p(\theta_i|\mathbf{y}) = p(\theta_i) \ .$$

[Stronger condition: data have no impact on the marginal (not conditional) posterior.]

Lip Cancer Example

- Dilemma: Though unidentified, the θ_i and ϕ_i are interesting in their own right, as is

  $$\psi = \frac{sd(\boldsymbol{\phi})}{sd(\boldsymbol{\theta}) + sd(\boldsymbol{\phi})} ,$$

  where $sd(\cdot)$ is the empirical marginal standard deviation function. Can we specify vague but proper prior values τ_h and τ_c that

Lip Cancer Example

- Dilemma: Though unidentified, the θ_i and ϕ_i are interesting in their own right, as is

  $$\psi = \frac{sd(\boldsymbol{\phi})}{sd(\boldsymbol{\theta}) + sd(\boldsymbol{\phi})} \; ,$$

  where $sd(\cdot)$ is the empirical marginal standard deviation function. Can we specify vague but proper prior values τ_h and τ_c that
  - lead to acceptable convergence behavior, and

Lip Cancer Example

- Dilemma: Though unidentified, the θ_i and ϕ_i are interesting in their own right, as is

  $$\psi = \frac{sd(\boldsymbol{\phi})}{sd(\boldsymbol{\theta}) + sd(\boldsymbol{\phi})} \ ,$$

  where $sd(\cdot)$ is the empirical marginal standard deviation function. Can we specify vague but proper prior values τ_h and τ_c that

  - lead to acceptable convergence behavior, and
  - still allow Bayesian learning?

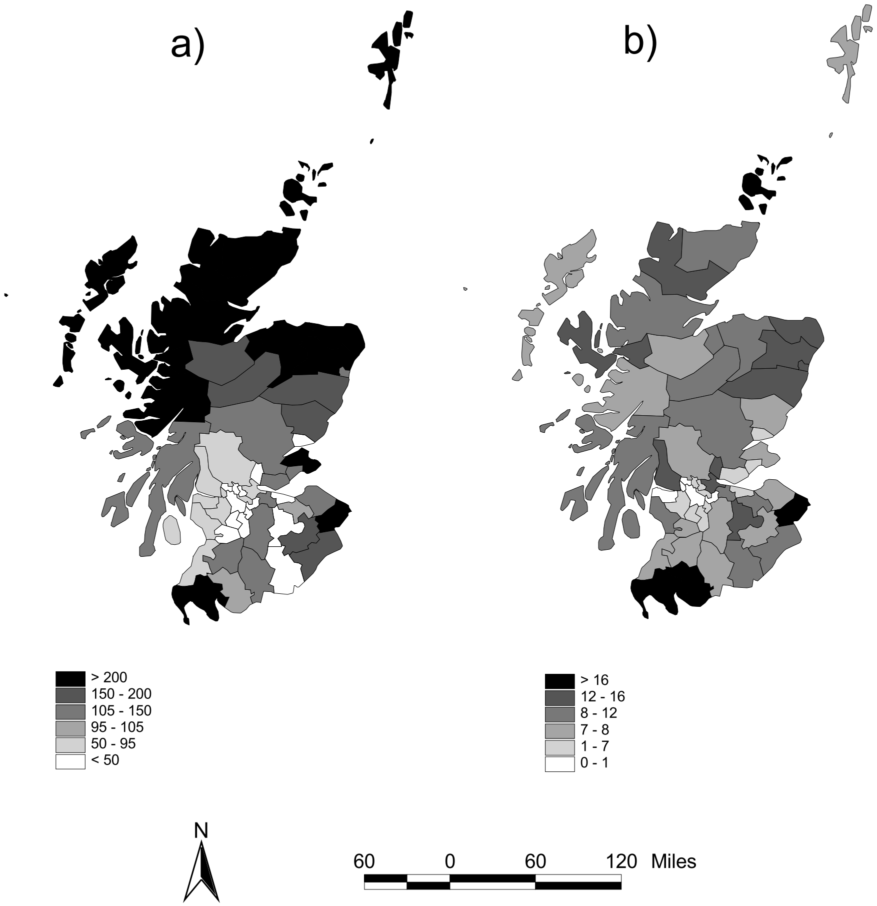
Lip Cancer Example

- Dilemma: Though unidentified, the θ_i and ϕ_i are interesting in their own right, as is

  $$\psi = \frac{sd(\boldsymbol{\phi})}{sd(\boldsymbol{\theta}) + sd(\boldsymbol{\phi})} \,,$$

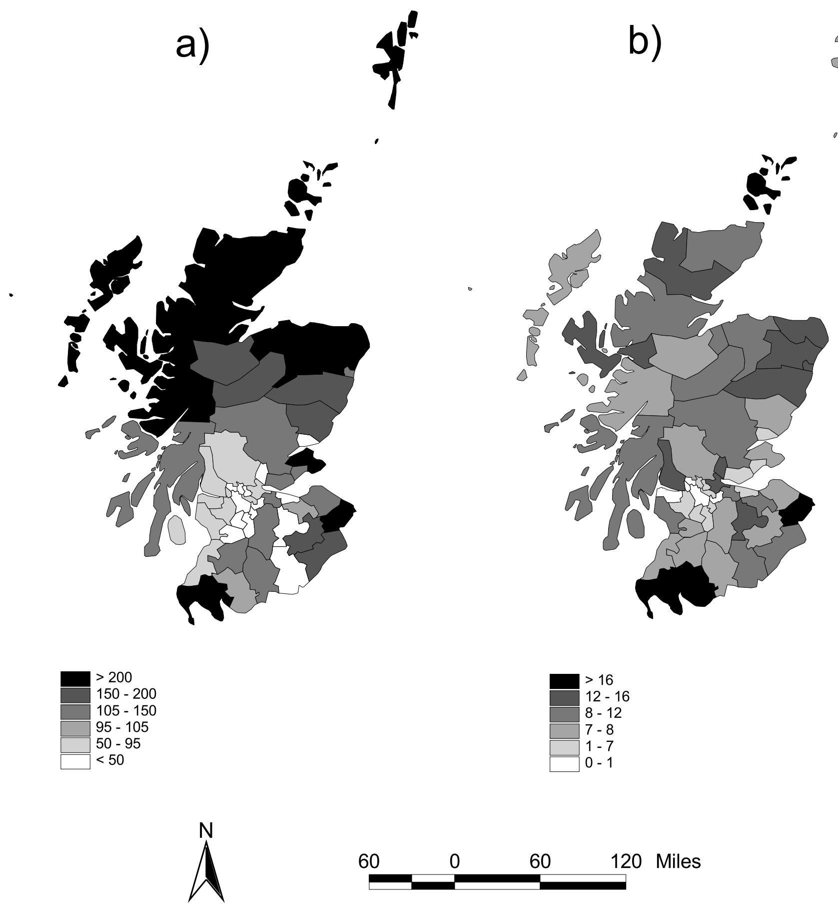
  where $sd(\cdot)$ is the empirical marginal standard deviation function. Can we specify vague but proper prior values τ_h and τ_c that
  - lead to acceptable convergence behavior, and
  - still allow Bayesian learning?
- Tricky to specify a “fair” prior balance between heterogeneity and clustering (e.g., one for which $\psi \approx 1/2$) since θ_i prior is specified marginally while the ϕ_i prior is specified conditionally.

Dataset: Scottish lip cancer data



a) $SMR_i = 100Y_i/E_i$, standardized mortality ratio for lip cancer in $I = 56$ districts, 1975–1980

Dataset: Scottish lip cancer data



a) $SMR_i = 100Y_i/E_i$, standardized mortality ratio for lip cancer in $I = 56$ districts, 1975–1980

b) one covariate, x_i = percentage of the population engaged in agriculture, fishing or forestry (AFF)

WinBUGS code to fit this model

```
model {
   for (i in 1 : regions) {
      O[i] ˜ dpois(mu[i])
      log(mu[i]) <- log(E[i]) + beta*aff[i]/10 + phi[i] + theta[i]
      theta[i] ˜ dnorm(0.0,tau.h)
      eta[i] <- theta[i] + phi[i]
   }
   phi[1:regions] ˜ {\red car.normal}(adj[], weights[], num[], tau.c)

   beta ˜ dnorm(0.0, 1.0E-5)  # vague prior on covariate effect

   tau.h ˜ dgamma(1.0E-3,1.0E-3)  #  ‘‘fair’’ prior from Best et al.
   tau.c ˜ dgamma(1.0E-1,1.0E-1)  #   (1999, Bayesian Statistics 6)

   sd.h <- sd(theta[]) # marginal SD of heterogeneity effects
   sd.c <- sd(phi[])   # marginal SD of clustering (spatial) effects
   psi <- sd.c / (sd.h + sd.c)
}

(See WinBUGS Map manual for DATA and INITS for this example)
```

Lip Cancer Results

	posterior for ψ			posterior for β		
priors for τ_c, τ_h	mean	sd	l1acf	mean	sd	l1acf
G(1.0, 1.0), G(3.2761, 1.81)	.57	.058	.80	.43	.17	.94
G(.1, .1), G(.32761, .181)	.65	.073	.89	.41	.14	.92
G(.1, .1), G(.001, .001)	.82	.10	.98	.38	.13	.91
	posterior for η_1			posterior for η_{56}		
priors for τ_c, τ_h	mean	sd	l1acf	mean	sd	l1acf
G(1.0, 1.0), G(3.2761, 1.81)	.92	.40	.33	–.96	.52	.12
G(.1, .1), G(.32761, .181)	.89	.36	.28	–.79	.41	.17
G(.1, .1), G(.001, .001)	.90	.34	.31	–.70	.35	.21

- AFF covariate is significantly $\neq 0$ under all 3 priors

Lip Cancer Results

	posterior for ψ			posterior for β		
priors for τ_c, τ_h	mean	sd	l1acf	mean	sd	l1acf
G(1.0, 1.0), G(3.2761, 1.81)	.57	.058	.80	.43	.17	.94
G(.1, .1), G(.32761, .181)	.65	.073	.89	.41	.14	.92
G(.1, .1), G(.001, .001)	.82	.10	.98	.38	.13	.91

	posterior for η_1			posterior for η_{56}		
priors for τ_c, τ_h	mean	sd	l1acf	mean	sd	l1acf
G(1.0, 1.0), G(3.2761, 1.81)	.92	.40	.33	–.96	.52	.12
G(.1, .1), G(.32761, .181)	.89	.36	.28	–.79	.41	.17
G(.1, .1), G(.001, .001)	.90	.34	.31	–.70	.35	.21

- AFF covariate is significantly $\neq 0$ under all 3 priors
- convergence is slow for ψ and β, but rapid for η_i (and μ_i)

Lip Cancer Results

	posterior for ψ			posterior for β		
priors for τ_c, τ_h	mean	sd	l1acf	mean	sd	l1acf
G(1.0, 1.0), G(3.2761, 1.81)	.57	.058	.80	.43	.17	.94
G(.1, .1), G(.32761, .181)	.65	.073	.89	.41	.14	.92
G(.1, .1), G(.001, .001)	.82	.10	.98	.38	.13	.91

	posterior for η_1			posterior for η_{56}		
priors for τ_c, τ_h	mean	sd	l1acf	mean	sd	l1acf
G(1.0, 1.0), G(3.2761, 1.81)	.92	.40	.33	–.96	.52	.12
G(.1, .1), G(.32761, .181)	.89	.36	.28	–.79	.41	.17
G(.1, .1), G(.001, .001)	.90	.34	.31	–.70	.35	.21

- AFF covariate is significantly $\neq 0$ under all 3 priors
- convergence is slow for ψ and β, but rapid for η_i (and μ_i)
- Excess variability in the data is mostly due to clustering ($E(\psi|\mathbf{y}) > .50$), but the posterior distribution for ψ does not seem robust to changes in the prior.

InterStim Example

Device that uses electrical stimulation of the brain to prevent urinary incontinences. For patients $i = 1, \ldots, 49$:

$$\begin{aligned} X_{1i} &= \text{number of incontinences per week at baseline} \\ X_{2i} &= \text{number of incontinences per week at 3 months} \\ X_{3i} &= \text{number of incontinences per week at 6 months} \\ X_{4i} &= \text{number of incontinences per week at 12 months} \end{aligned}$$

patient	X_{1i}	X_{2i}	X_{3i}	X_{4i}
1	60	0.7	0	16
2	8	0	0	0
...				
8	9	0.7	12	NA
9	3	0	0.7	0
...				
49	16	NA	NA	NA

InterStim Example

- Goal 1: Obtain full predictive inference for all missing X values (point and interval estimates)

InterStim Example

- Goal 1: Obtain full predictive inference for all missing X values (point and interval estimates)
- Goal 2: Obtain measure of percent improvement (relative to baseline) due to InterStim at 6 and 12 months

InterStim Example

- Goal 1: Obtain full predictive inference for all missing X values (point and interval estimates)
- Goal 2: Obtain measure of percent improvement (relative to baseline) due to InterStim at 6 and 12 months
- Model: Let $\mathbf{X}_i = (X_{1i}, X_{2i}, X_{3i}, X_{4i})'$ and $\boldsymbol{\theta} = (\theta_1, \theta_2, \theta_3, \theta_4)'$. Clearly the X_{ij}'s are correlated, but an ordinary longitudinal model does not seem appropriate (we can't just use a linear model here). So instead, maintain the generality:

$$\begin{aligned} \mathbf{X}_i | \boldsymbol{\theta}, \Upsilon &\overset{iid}{\sim} N_4(\boldsymbol{\theta}, \Upsilon^{-1}) \\ \boldsymbol{\theta} &\sim N_4(\boldsymbol{\mu}, \Omega^{-1}) \\ \Upsilon &\sim Wishart_4(R, \rho) \end{aligned}$$

InterStim Example

- `WinBUGS` will generate all missing X's ("NA"s in the dataset) from their full conditional distributions as part of the Gibbs algorithm. Thus we will obtain samples from $p(X_{ij}|\mathbf{X}_{obs})$ for all missing X_{ij} (achieving Goal 1).

InterStim Example

- `WinBUGS` will generate all missing X's ("NA"s in the dataset) from their full conditional distributions as part of the Gibbs algorithm. Thus we will obtain samples from $p(X_{ij}|\mathbf{X}_{obs})$ for all missing X_{ij} (achieving Goal 1).
- Re: Goal 2 (percent improvement from baseline), let

$$\alpha = \frac{\theta_1 - \theta_3}{\theta_1} \text{ and } \beta = \frac{\theta_1 - \theta_4}{\theta_1}$$

Then $p(\alpha|\mathbf{X}_{obs})$ and $p(\beta|\mathbf{X}_{obs})$ address this issue!

InterStim Example

- `WinBUGS` will generate all missing X's ("NA"s in the dataset) from their full conditional distributions as part of the Gibbs algorithm. Thus we will obtain samples from $p(X_{ij}|\mathbf{X}_{obs})$ for all missing X_{ij} (achieving Goal 1).
- Re: Goal 2 (percent improvement from baseline), let

  $$\alpha = \frac{\theta_1 - \theta_3}{\theta_1} \text{ and } \beta = \frac{\theta_1 - \theta_4}{\theta_1}$$

  Then $p(\alpha|\mathbf{X}_{obs})$ and $p(\beta|\mathbf{X}_{obs})$ address this issue!
- Hyperparameters: all vague: $\Omega = Diag(10^{-6}, \ldots, 10^{-6})$, $\boldsymbol{\mu} = \mathbf{0}$, $\rho = 4$ (the smallest value for which the Wishart prior for Υ is proper) and $R = Diag(10^{-1}, \ldots, 10^{-1})$.

InterStim Example

- `WinBUGS` will generate all missing X's ("NA"s in the dataset) from their full conditional distributions as part of the Gibbs algorithm. Thus we will obtain samples from $p(X_{ij}|\mathbf{X}_{obs})$ for all missing X_{ij} (achieving Goal 1).
- Re: Goal 2 (percent improvement from baseline), let

  $$\alpha = \frac{\theta_1 - \theta_3}{\theta_1} \text{ and } \beta = \frac{\theta_1 - \theta_4}{\theta_1}$$

  Then $p(\alpha|\mathbf{X}_{obs})$ and $p(\beta|\mathbf{X}_{obs})$ address this issue!
- Hyperparameters: all vague: $\Omega = Diag(10^{-6}, \ldots, 10^{-6})$, $\boldsymbol{\mu} = \mathbf{0}$, $\rho = 4$ (the smallest value for which the Wishart prior for Υ is proper) and $R = Diag(10^{-1}, \ldots, 10^{-1})$.
- Code and Data:
  www.biostat.umn.edu/∼brad/data/InterStim.odc