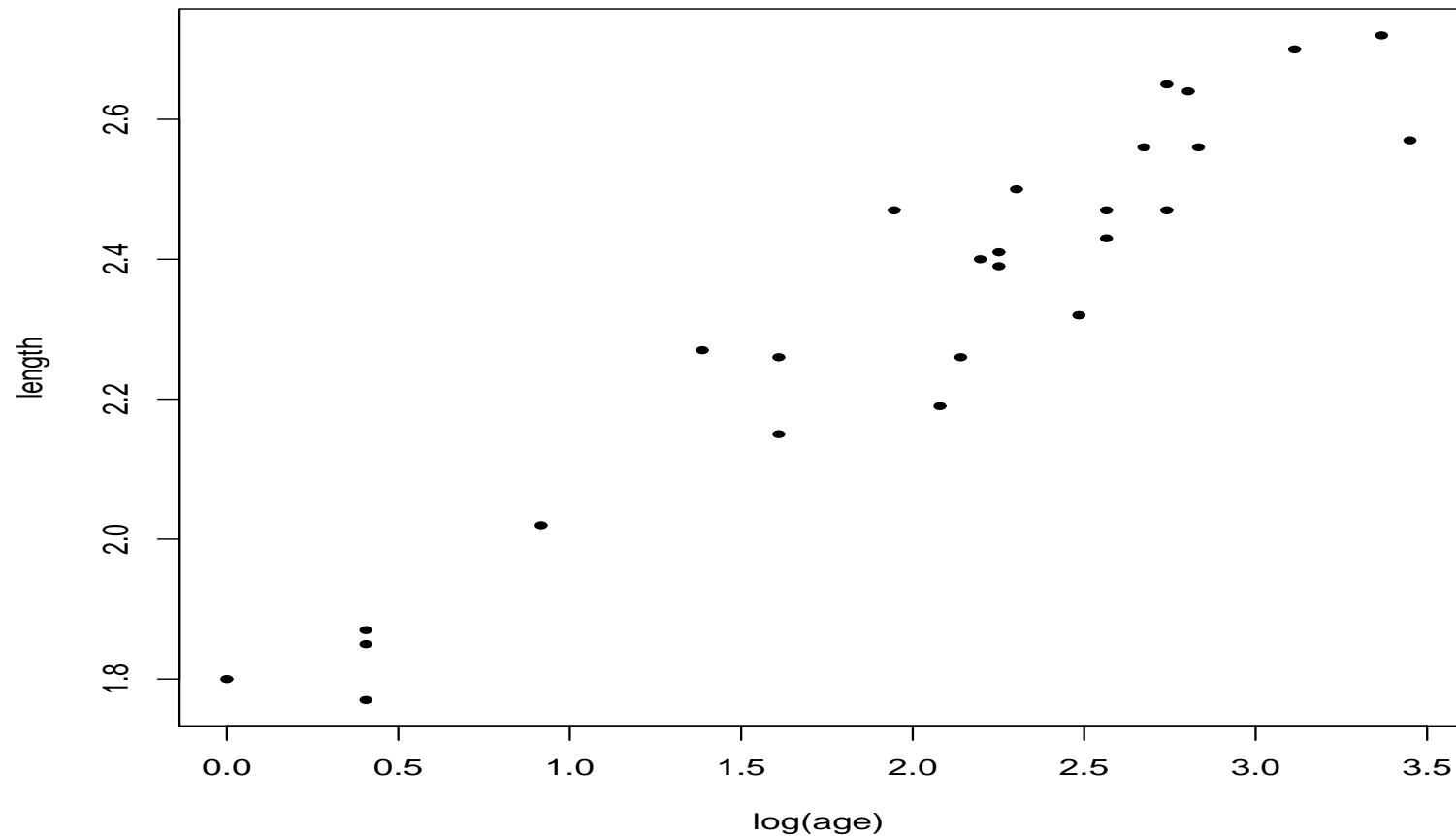
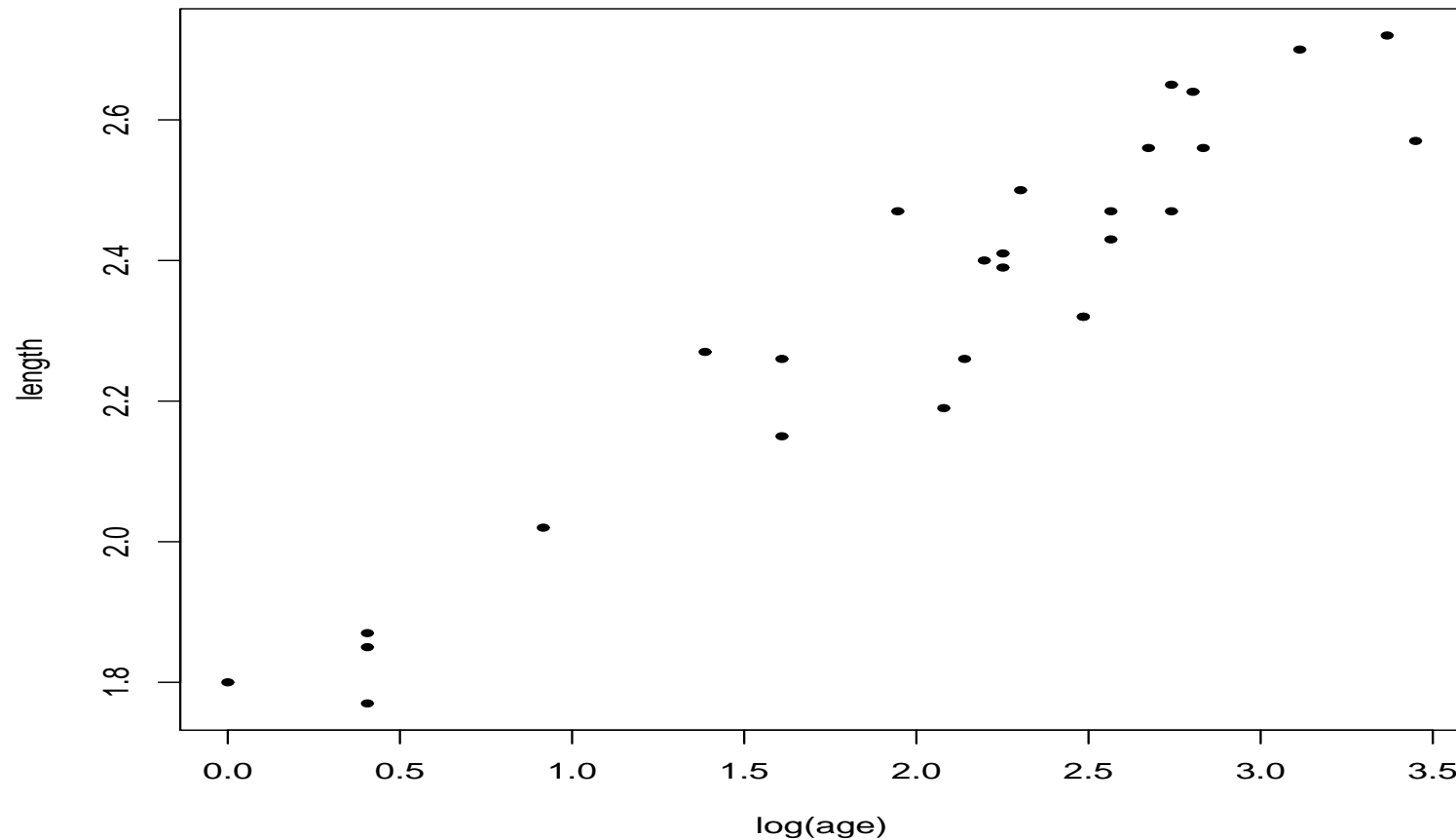


BUGS Example 1: Linear Regression



- For $n = 27$ captured samples of the sirenian species *dugong* (sea cow), relate an animal's length in meters, Y_i , to its age in years, x_i .

BUGS Example 1: Linear Regression



- For $n = 27$ captured samples of the sirenian species *dugong* (sea cow), relate an animal's length in meters, Y_i , to its age in years, x_i .
- To avoid a nonlinear model for now, transform x_i to the log scale; plot of Y versus $\log(x)$ looks fairly **linear**!

Simple linear regression in WinBUGS

$$Y_i = \beta_0 + \beta_1 \log(x_i) + \epsilon_i, \quad i = 1, \dots, n$$

where $\epsilon_i \stackrel{iid}{\sim} N(0, \tau)$ and $\tau = 1/\sigma^2$, the **precision** in the data.

- Prior distributions:
 - flat for β_0, β_1
 - vague gamma on τ (say, **Gamma(0.1, 0.1)**, which has mean 1 and variance 10) is traditional

Simple linear regression in WinBUGS

$$Y_i = \beta_0 + \beta_1 \log(x_i) + \epsilon_i, \quad i = 1, \dots, n$$

where $\epsilon_i \stackrel{iid}{\sim} N(0, \tau)$ and $\tau = 1/\sigma^2$, the **precision** in the data.

- Prior distributions:
 - flat for β_0, β_1
 - vague gamma on τ (say, **Gamma(0.1, 0.1)**, which has mean 1 and variance 10) is traditional
- posterior correlation is reduced by **centering** the $\log(x_i)$ around their own mean

Simple linear regression in WinBUGS

$$Y_i = \beta_0 + \beta_1 \log(x_i) + \epsilon_i, \quad i = 1, \dots, n$$

where $\epsilon_i \stackrel{iid}{\sim} N(0, \tau)$ and $\tau = 1/\sigma^2$, the **precision** in the data.

- Prior distributions:
 - flat for β_0, β_1
 - vague gamma on τ (say, **Gamma(0.1, 0.1)**, which has mean 1 and variance 10) is traditional
- posterior correlation is reduced by **centering** the $\log(x_i)$ around their own mean
- Andrew Gelman suggests placing a **uniform** prior on σ , bounding the prior away from 0 and $\infty \implies U(.01, 100)$?

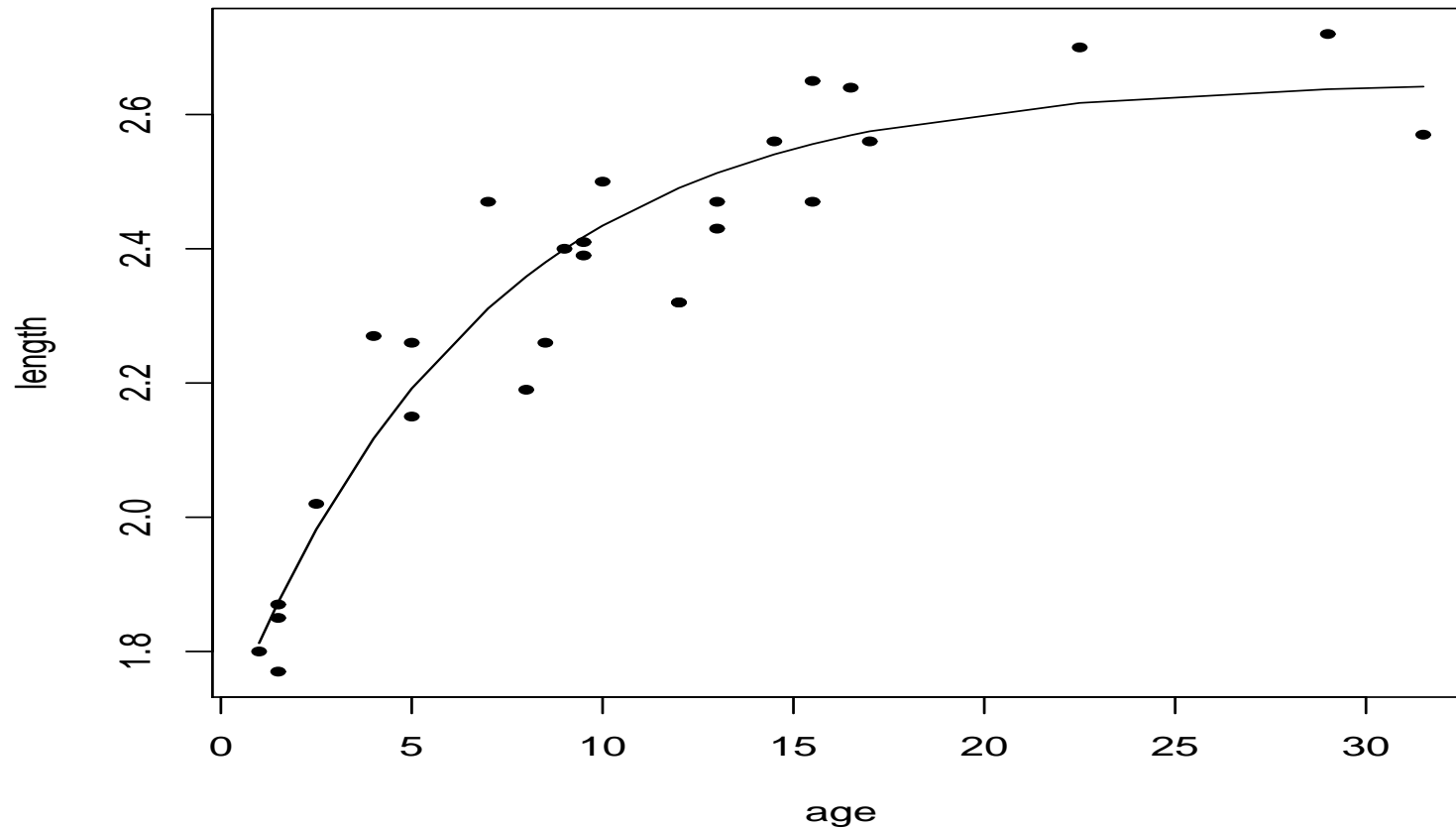
Simple linear regression in WinBUGS

$$Y_i = \beta_0 + \beta_1 \log(x_i) + \epsilon_i, \quad i = 1, \dots, n$$

where $\epsilon_i \stackrel{iid}{\sim} N(0, \tau)$ and $\tau = 1/\sigma^2$, the **precision** in the data.

- Prior distributions:
 - flat for β_0, β_1
 - vague gamma on τ (say, **Gamma(0.1, 0.1)**, which has mean 1 and variance 10) is traditional
- posterior correlation is reduced by **centering** the $\log(x_i)$ around their own mean
- Andrew Gelman suggests placing a **uniform** prior on σ , bounding the prior away from 0 and $\infty \implies U(.01, 100)$?
- **Code:**
www.biostat.umn.edu/~brad/data/dugongs_BUGS.txt

BUGS Example 2: Nonlinear Regression



- Model the **untransformed** dugong data as

$$Y_i = \alpha - \beta\gamma^{x_i} + \epsilon_i, \quad i = 1, \dots, n,$$

where $\alpha > 0$, $\beta > 0$, $0 \leq \gamma \leq 1$, and as usual $\epsilon_i \stackrel{iid}{\sim} N(0, \tau)$ for $\tau \equiv 1/\sigma^2 > 0$.

Nonlinear regression in WinBUGS

- In this model,
 - α corresponds to the average length of a fully grown dugong ($x \rightarrow \infty$)
 - $(\alpha - \beta)$ is the length of a dugong at birth ($x = 0$)
 - γ determines the **growth rate**: lower values produce an initially steep growth curve while higher values lead to gradual, almost linear growth.

Nonlinear regression in WinBUGS

- In this model,
 - α corresponds to the average length of a fully grown dugong ($x \rightarrow \infty$)
 - $(\alpha - \beta)$ is the length of a dugong at birth ($x = 0$)
 - γ determines the **growth rate**: lower values produce an initially steep growth curve while higher values lead to gradual, almost linear growth.
- **Prior distributions**: flat for α and β , $U(.01, 100)$ for σ , and $U(0.5, 1.0)$ for γ (harder to estimate)

Nonlinear regression in WinBUGS

- In this model,
 - α corresponds to the average length of a fully grown dugong ($x \rightarrow \infty$)
 - $(\alpha - \beta)$ is the length of a dugong at birth ($x = 0$)
 - γ determines the **growth rate**: lower values produce an initially steep growth curve while higher values lead to gradual, almost linear growth.
- **Prior distributions**: flat for α and β , $U(.01, 100)$ for σ , and $U(0.5, 1.0)$ for γ (harder to estimate)
- **Code**:
www.biostat.umn.edu/~brad/data/dugongsNL_BUGS.txt

Nonlinear regression in WinBUGS

- In this model,
 - α corresponds to the average length of a fully grown dugong ($x \rightarrow \infty$)
 - $(\alpha - \beta)$ is the length of a dugong at birth ($x = 0$)
 - γ determines the **growth rate**: lower values produce an initially steep growth curve while higher values lead to gradual, almost linear growth.
- **Prior distributions**: flat for α and β , $U(.01, 100)$ for σ , and $U(0.5, 1.0)$ for γ (harder to estimate)
- **Code**:
www.biostat.umn.edu/~brad/data/dugongsNL_BUGS.txt
- Obtain posterior density estimates and autocorrelation plots for α, β, γ , and σ , and investigate the **bivariate posterior** of (α, γ) using the **Correlation** tool on the **Inference** menu!

BUGS Example 3: Logistic Regression

- Consider a binary version of the dugong data,

$$Z_i = \begin{cases} 1 & \text{if } Y_i > 2.4 \text{ (i.e., the dugong is "full-grown")} \\ 0 & \text{otherwise} \end{cases}$$

BUGS Example 3: Logistic Regression

- Consider a binary version of the dugong data,

$$Z_i = \begin{cases} 1 & \text{if } Y_i > 2.4 \text{ (i.e., the dugong is "full-grown")} \\ 0 & \text{otherwise} \end{cases}$$

- A **logistic** model for $p_i = P(Z_i = 1)$ is then

$$\text{logit}(p_i) = \log[p_i/(1 - p_i)] = \beta_0 + \beta_1 \log(x_i) .$$

BUGS Example 3: Logistic Regression

- Consider a binary version of the dugong data,

$$Z_i = \begin{cases} 1 & \text{if } Y_i > 2.4 \text{ (i.e., the dugong is "full-grown")} \\ 0 & \text{otherwise} \end{cases}$$

- A **logistic** model for $p_i = P(Z_i = 1)$ is then

$$\text{logit}(p_i) = \log[p_i/(1 - p_i)] = \beta_0 + \beta_1 \log(x_i) .$$

- Two other commonly used link functions are the **probit**,

$$\text{probit}(p_i) = \Phi^{-1}(p_i) = \beta_0 + \beta_1 \log(x_i) ,$$

and the **complementary log-log** (cloglog),

$$\text{cloglog}(p_i) = \log[-\log(1 - p_i)] = \beta_0 + \beta_1 \log(x_i) .$$

Binary regression in WinBUGS

- Code:
www.biostat.umn.edu/~brad/data/dugongsBin_BUGS.txt

Binary regression in WinBUGS

- Code:
www.biostat.umn.edu/~brad/data/dugongsBin_BUGS.txt
- Code uses flat priors for β_0 and β_1 , and the **phi** function, instead of the less stable **probit** function.

Binary regression in WinBUGS

- Code:
www.biostat.umn.edu/~brad/data/dugongsBin_BUGS.txt
- Code uses flat priors for β_0 and β_1 , and the **phi** function, instead of the less stable **probit** function.
- DIC scores for the three models:

model	\bar{D}	p_D	DIC
logit	19.62	1.85	21.47
probit	19.30	1.87	21.17
cloglog	18.77	1.84	20.61

In fact, these scores can be obtained **from a single run**; see the **“trick version”** at the bottom of the BUGS file!

Binary regression in WinBUGS

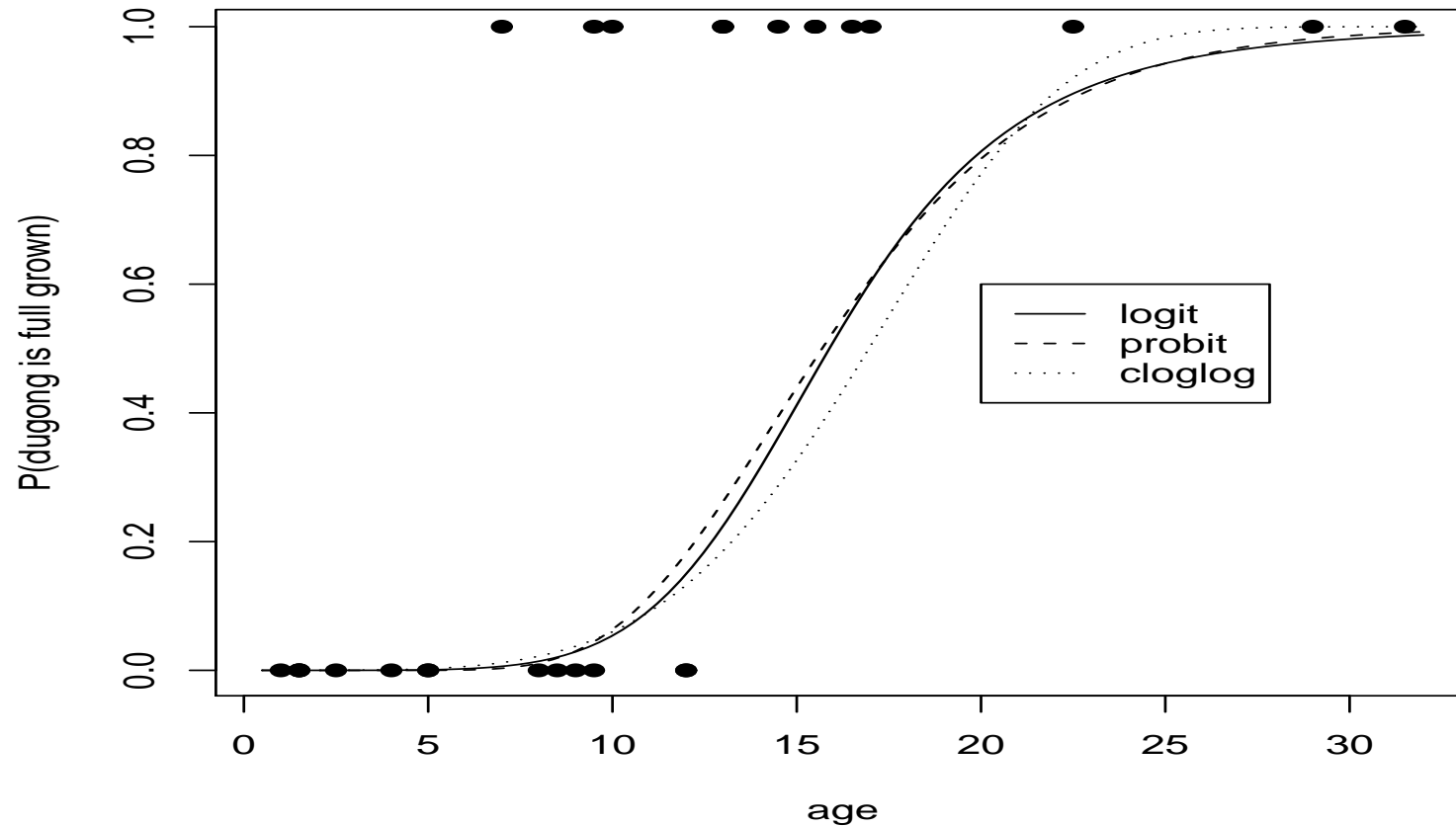
- Code:
www.biostat.umn.edu/~brad/data/dugongsBin_BUGS.txt
- Code uses flat priors for β_0 and β_1 , and the **phi** function, instead of the less stable **probit** function.
- DIC scores for the three models:

model	\bar{D}	p_D	DIC
logit	19.62	1.85	21.47
probit	19.30	1.87	21.17
cloglog	18.77	1.84	20.61

In fact, these scores can be obtained **from a single run**; see the **“trick version”** at the bottom of the BUGS file!

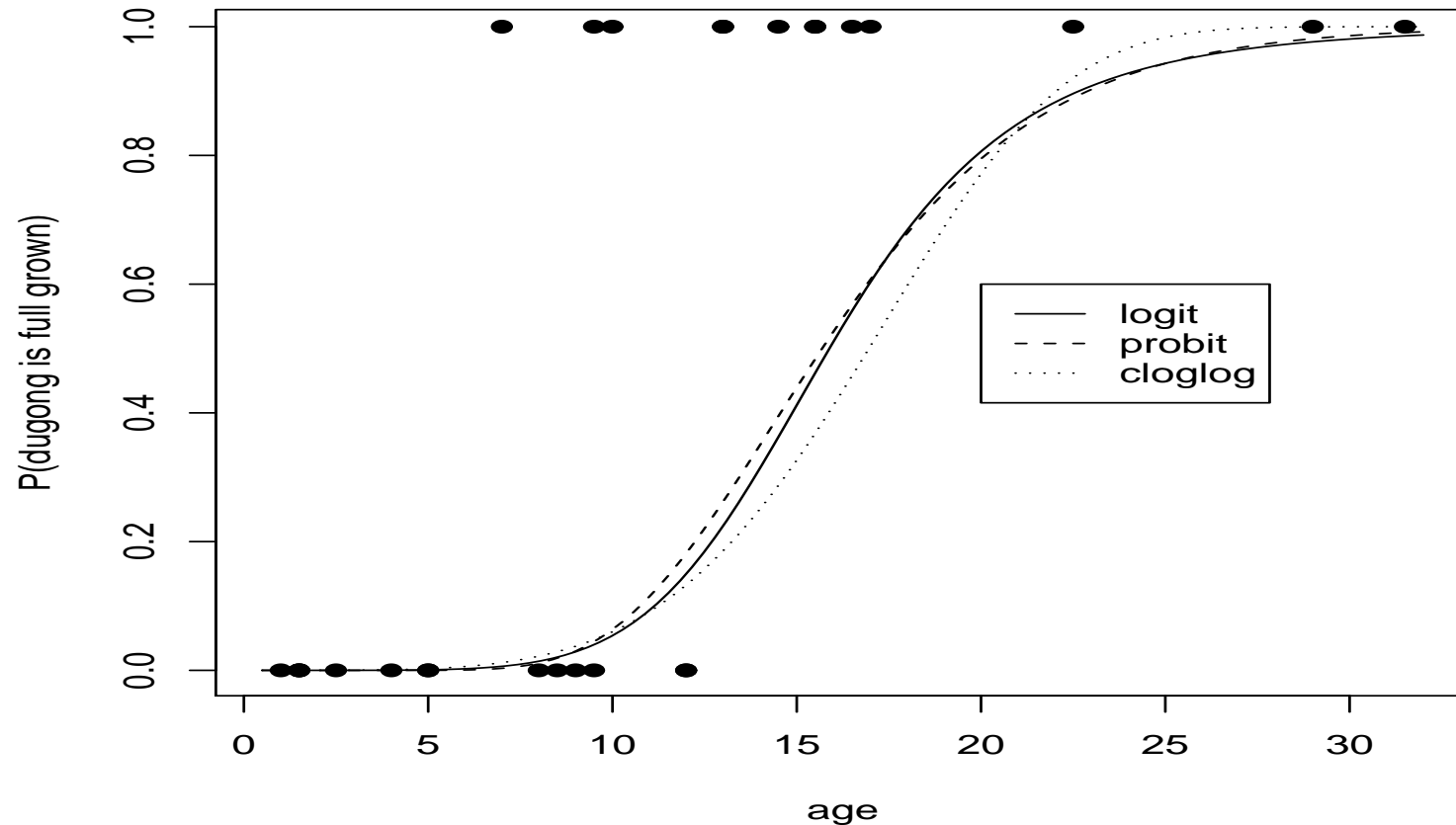
- Use the **Comparison** tool to compare the posteriors of β_1 across models, and the **Correlation** tool to check the bivariate posteriors of (β_0, β_1) across models.

Fitted binary regression models



- The logit and probit fits appear very similar, but the cloglog fitted curve is slightly different

Fitted binary regression models



- The logit and probit fits appear very similar, but the cloglog fitted curve is slightly different
- You can also compare p_i posterior boxplots (induced by the link function and the β_0 and β_1 posteriors) using the **Comparison** tool.

BUGS Example 4: Hierarchical Models

- Extend the usual **two-stage** (likelihood plus prior) Bayesian structure to a hierarchy of L levels, where the joint distribution of the data and the parameters is

$$f(\mathbf{y}|\boldsymbol{\theta}_1)\pi_1(\boldsymbol{\theta}_1|\boldsymbol{\theta}_2)\pi_2(\boldsymbol{\theta}_2|\boldsymbol{\theta}_3)\cdots\pi_L(\boldsymbol{\theta}_L|\boldsymbol{\lambda}).$$

BUGS Example 4: Hierarchical Models

- Extend the usual **two-stage** (likelihood plus prior) Bayesian structure to a hierarchy of L levels, where the joint distribution of the data and the parameters is

$$f(\mathbf{y}|\boldsymbol{\theta}_1)\pi_1(\boldsymbol{\theta}_1|\boldsymbol{\theta}_2)\pi_2(\boldsymbol{\theta}_2|\boldsymbol{\theta}_3)\cdots\pi_L(\boldsymbol{\theta}_L|\boldsymbol{\lambda}).$$

- L is often determined by the number of **subscripts** on the data. For example, suppose Y_{ijk} is the test score of child k in classroom j in school i in a certain city. Model:

$$Y_{ijk}|\theta_{ij} \stackrel{ind}{\sim} N(\theta_{ij}, \tau_\theta) \quad (\theta_{ij} \text{ is the } \mathbf{classroom} \text{ effect})$$

$$\theta_{ij}|\eta_i \stackrel{ind}{\sim} N(\eta_i, \tau_\eta) \quad (\eta_i \text{ is the } \mathbf{school} \text{ effect})$$

$$\eta_i|\lambda \stackrel{iid}{\sim} N(\lambda, \tau_\lambda) \quad (\lambda \text{ is the } \mathbf{grand mean})$$

Priors for λ and the τ 's now complete the specification!

Cross-Study (Meta-analysis) Data

- **Data:** estimated log relative hazards $Y_{ij} = \hat{\beta}_{ij}$ obtained by fitting separate Cox proportional hazards regressions to the data from each of $J = 18$ clinical units participating in $I = 6$ different AIDS studies.

Cross-Study (Meta-analysis) Data

- **Data:** estimated log relative hazards $Y_{ij} = \hat{\beta}_{ij}$ obtained by fitting separate Cox proportional hazards regressions to the data from each of $J = 18$ clinical units participating in $I = 6$ different AIDS studies.
- To these data we wish to fit the **cross-study** model,

$$Y_{ij} = a_i + b_j + s_{ij} + \epsilon_{ij}, \quad i = 1, \dots, I, \quad j = 1, \dots, J,$$

where a_i = study main effect

b_j = unit main effect

s_{ij} = study-unit interaction term, and

$$\epsilon_{ij} \stackrel{iid}{\sim} N(0, \sigma_{ij}^2)$$

and the estimated standard errors from the Cox regressions are used as (known) values of the σ_{ij} .

Cross-Study (Meta-analysis) Data

Estimated Unit-Specific Log Relative Hazards						
Unit	Toxo	ddl/ddC	NuCombo ZDV+ddl	NuCombo ZDV+ddC	Fungal	CMV
A	0.814	NA	-0.406	0.298	0.094	NA
B	-0.203	NA	NA	NA	NA	NA
C	-0.133	NA	0.218	-2.206	0.435	0.145
D	NA	NA	NA	NA	NA	NA
E	-0.715	-0.242	-0.544	-0.731	0.600	0.041
F	0.739	0.009	NA	NA	NA	0.222
G	0.118	0.807	-0.047	0.913	-0.091	0.099
H	NA	-0.511	0.233	0.131	NA	0.017
I	NA	1.939	0.218	-0.066	NA	0.355
J	0.271	1.079	-0.277	-0.232	0.752	0.203
K	NA	NA	0.792	1.264	-0.357	0.807
:	:	:	:	:	:	:
R	1.217	0.165	0.385	0.172	-0.022	0.203

Cross-Study (Meta-analysis) Data

- Note that some values are missing (“NA”) since
 - not all 18 units participated in all 6 studies
 - the Cox estimation procedure did not converge for some units that had few deaths

Cross-Study (Meta-analysis) Data

- Note that some values are missing (“NA”) since
 - not all 18 units participated in all 6 studies
 - the Cox estimation procedure did not converge for some units that had few deaths
- *Goal:* To identify which clinics are **opinion leaders** (strongly agree with overall result across studies) and which are **dissenters** (strongly disagree).

Cross-Study (Meta-analysis) Data

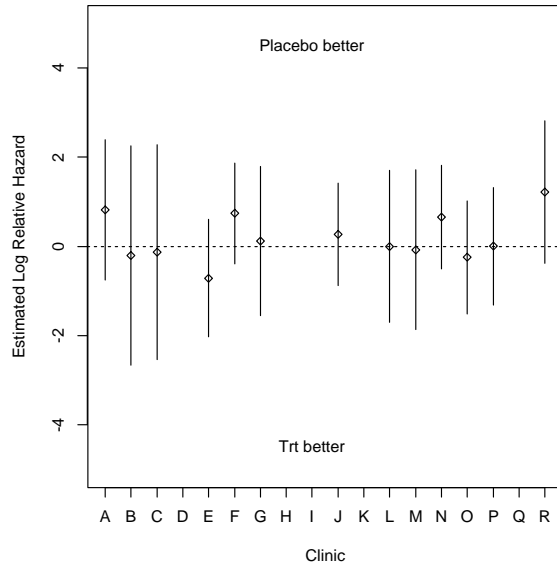
- Note that some values are missing (“NA”) since
 - not all 18 units participated in all 6 studies
 - the Cox estimation procedure did not converge for some units that had few deaths
- *Goal:* To identify which clinics are **opinion leaders** (strongly agree with overall result across studies) and which are **dissenters** (strongly disagree).
- Here, overall results all favor the treatment (i.e. mostly negative Y s) **except in Trial 1** (Toxo). Thus we multiply all the Y_{ij} 's by -1 for $i \neq 1$, so that larger Y_{ij} correspond in all cases to stronger agreement with the overall.

Cross-Study (Meta-analysis) Data

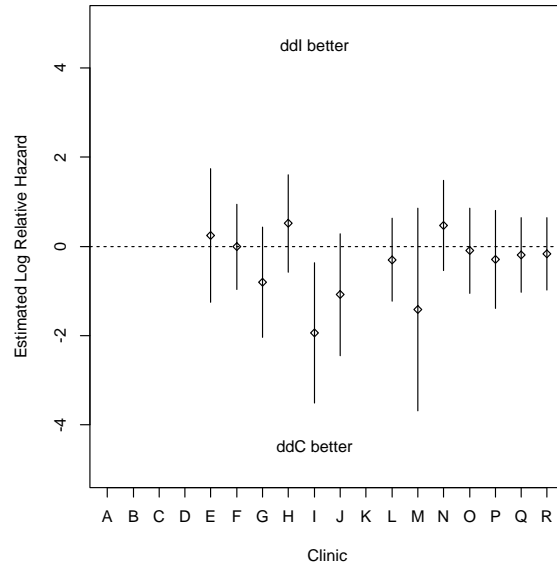
- Note that some values are missing (“NA”) since
 - not all 18 units participated in all 6 studies
 - the Cox estimation procedure did not converge for some units that had few deaths
- *Goal:* To identify which clinics are **opinion leaders** (strongly agree with overall result across studies) and which are **dissenters** (strongly disagree).
- Here, overall results all favor the treatment (i.e. mostly negative Y s) **except in Trial 1** (Toxo). Thus we multiply all the Y_{ij} 's by -1 for $i \neq 1$, so that larger Y_{ij} correspond in all cases to stronger agreement with the overall.
- Next slide shows a plot of the Y_{ij} values and associated approximate 95% CIs...

Cross-Study (Meta-analysis) Data

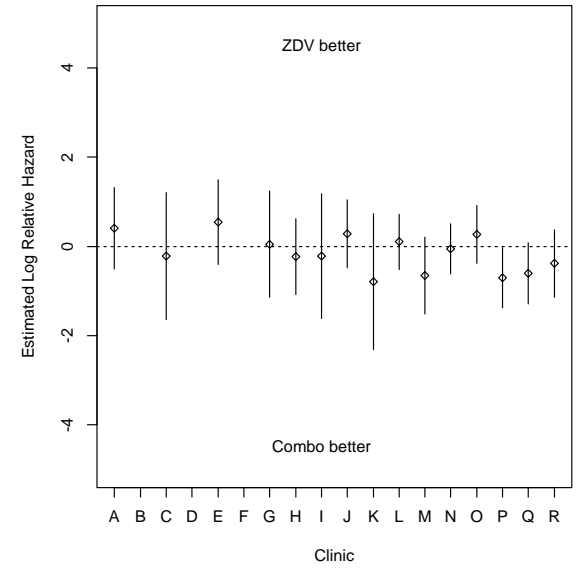
1: Toxo



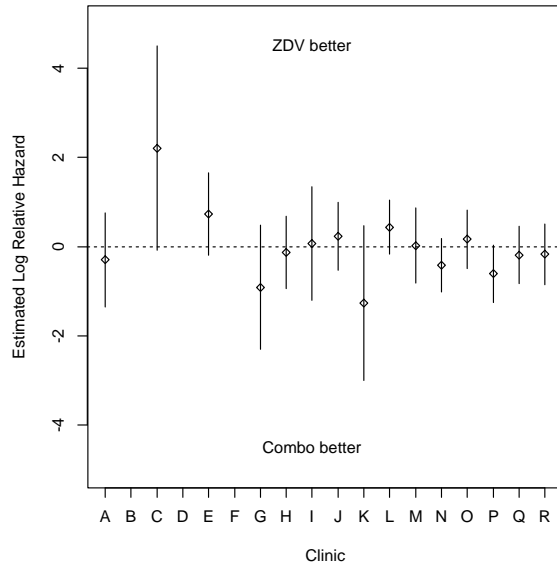
2: ddl/ddC



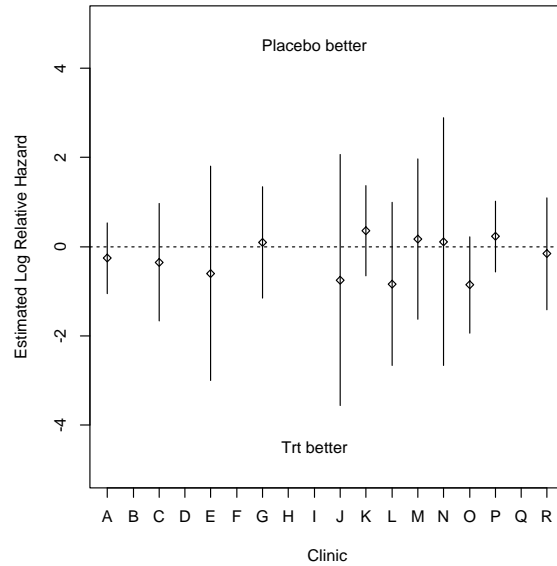
3: NuCombo-ddl



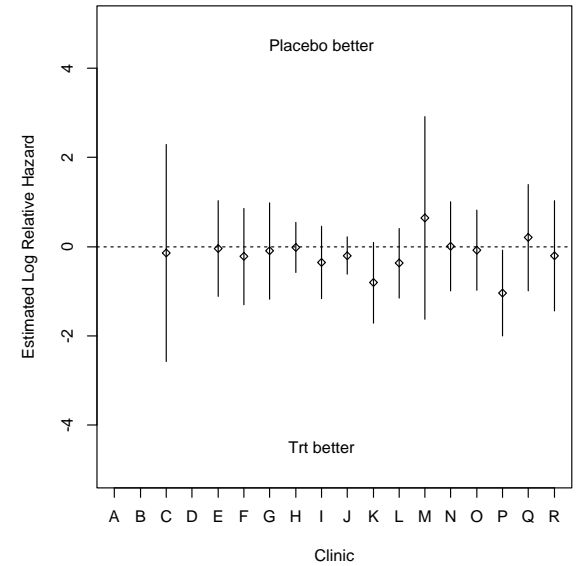
4: NuCombo-ddC



5: Fungal



6: CMV



Cross-Study (Meta-analysis) Data

- Second stage of our model:

$$a_i \stackrel{iid}{\sim} N(0, 100^2), \quad b_j \stackrel{iid}{\sim} N(0, \sigma_b^2), \quad \text{and} \quad s_{ij} \stackrel{iid}{\sim} N(0, \sigma_s^2)$$

Cross-Study (Meta-analysis) Data

- Second stage of our model:

$$a_i \stackrel{iid}{\sim} N(0, 100^2), \quad b_j \stackrel{iid}{\sim} N(0, \sigma_b^2), \quad \text{and} \quad s_{ij} \stackrel{iid}{\sim} N(0, \sigma_s^2)$$

- Third stage of our model:

$$\sigma_b \sim Unif(0.01, 100) \quad \text{and} \quad \sigma_s \sim Unif(0.01, 100)$$

That is, we

- **preclude** borrowing of strength across studies, but
- **encourage** borrowing of strength across units

Cross-Study (Meta-analysis) Data

- Second stage of our model:

$$a_i \stackrel{iid}{\sim} N(0, 100^2), \quad b_j \stackrel{iid}{\sim} N(0, \sigma_b^2), \quad \text{and} \quad s_{ij} \stackrel{iid}{\sim} N(0, \sigma_s^2)$$

- Third stage of our model:

$$\sigma_b \sim Unif(0.01, 100) \quad \text{and} \quad \sigma_s \sim Unif(0.01, 100)$$

That is, we

- **preclude** borrowing of strength across studies, but
- **encourage** borrowing of strength across units
- With $I + J + IJ$ parameters but fewer than IJ data points, **some** effects **must** be treated as random!

Cross-Study (Meta-analysis) Data

- Second stage of our model:

$$a_i \stackrel{iid}{\sim} N(0, 100^2), \quad b_j \stackrel{iid}{\sim} N(0, \sigma_b^2), \quad \text{and} \quad s_{ij} \stackrel{iid}{\sim} N(0, \sigma_s^2)$$

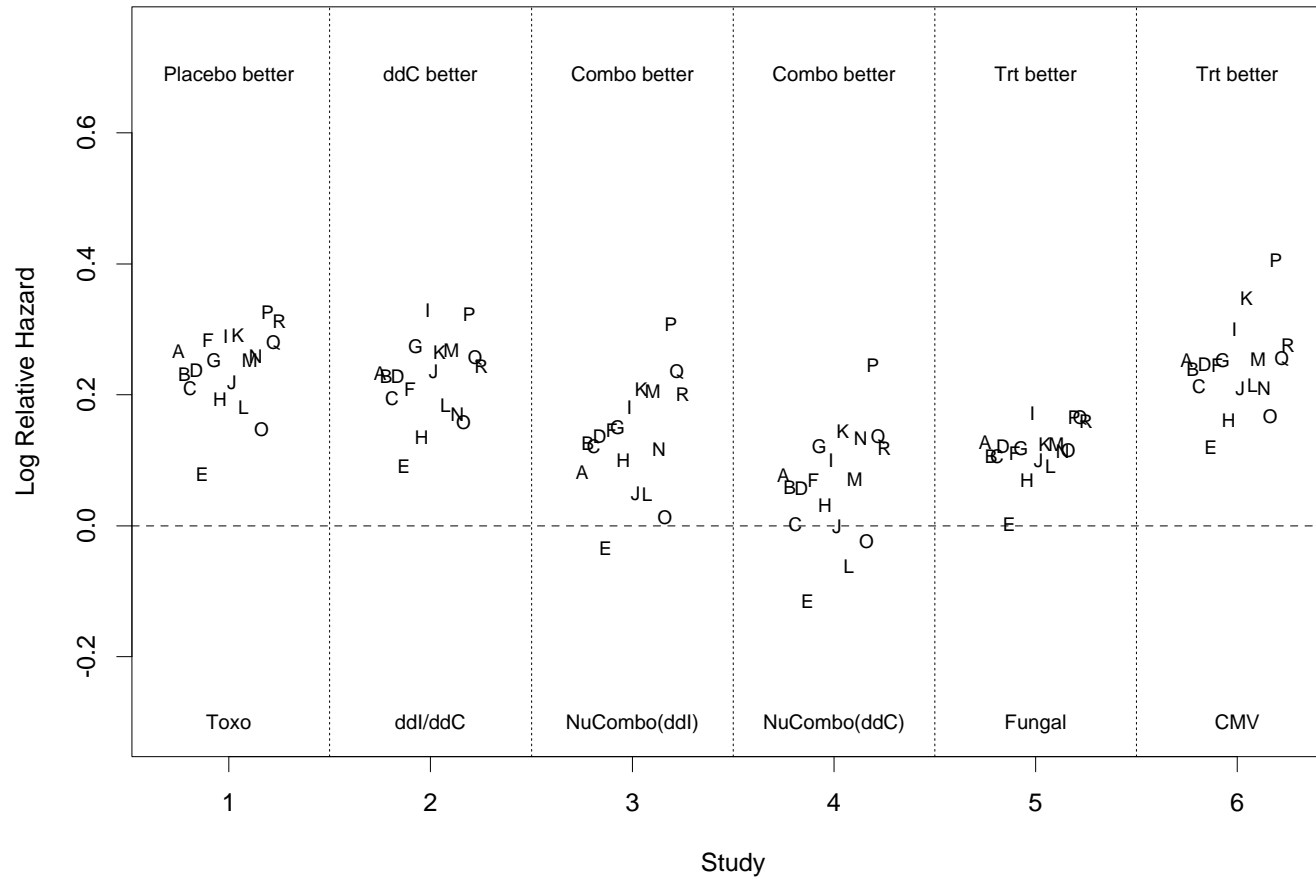
- Third stage of our model:

$$\sigma_b \sim Unif(0.01, 100) \quad \text{and} \quad \sigma_s \sim Unif(0.01, 100)$$

That is, we

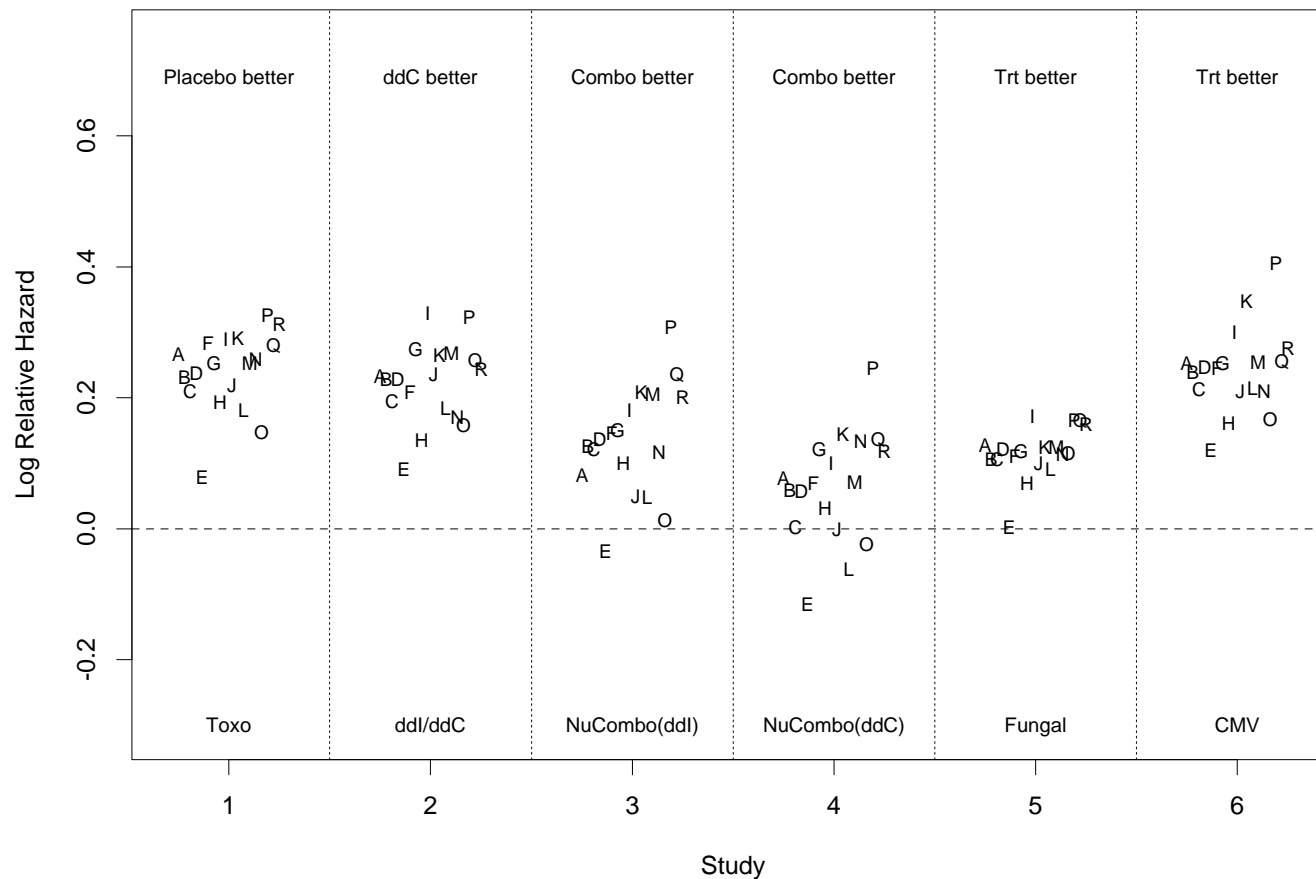
- **preclude** borrowing of strength across studies, but
- **encourage** borrowing of strength across units
- With $I + J + IJ$ parameters but fewer than IJ data points, **some** effects **must** be treated as random!
- **Code:**
www.biostat.umn.edu/~brad/data/crprot_BUGS.txt

Plot of θ_{ij} posterior means



◇ Unit *P* is an opinion leader; Unit *E* is a dissenter

Plot of θ_{ij} posterior means



- ◇ Unit P is an opinion leader; Unit E is a dissenter
- ◇ Substantial shrinkage towards 0 has occurred: mostly positive values; no estimated θ_{ij} greater than 0.6

Model Comparison via DIC

Since we lack replications for each study-unit (i - j) combination, the interactions s_{ij} in this model were only weakly identified, and the model might well be better off without them (or even without the unit effects b_j).

As such, compare a variety of reduced models:

```
Y[i,j] ~ dnorm(theta[i,j],P[i,j])
#   theta[i,j] <- a[i]+b[j]+s[i,j]   # full model
#   theta[i,j] <- a[i] + b[j]       # drop interactions
#   theta[i,j] <- a[i] + s[i,j]     # no unit effect
#   theta[i,j] <- b[j] + s[i,j]     # no study effect
#   theta[i,j] <- a[1] + b[j]       # unit + intercept
#   theta[i,j] <- b[j]              # unit effect only
#   theta[i,j] <- a[i]              # study effect only
```

Investigate p_D values for these models; are they consistent with posterior **boxplots** of the b_i and s_{ij} ?

DIC results for Cross-Study Data:

model	\bar{D}	p_D	DIC
full model	122.0	12.8	134.8
drop interactions	123.4	9.7	133.1
no unit effect	123.8	10.0	133.8
no study effect	121.4	9.7	131.1
unit + intercept	120.3	4.6	124.9
unit effect only	122.9	6.2	129.1
study effect only	126.0	6.0	132.0

The **DIC-best model** is the one with only an intercept (a role played here by a_1) and the unit effects b_j .

These DIC differences are not much larger than their possible Monte Carlo errors, so almost **any** of these models could be justified here.

BUGS Example 5: Survival Modeling

- Our data arises from a clinical trial comparing two treatments for *Mycobacterium avium complex (MAC)*, a disease common in late stage HIV-infected persons. Eleven clinical centers (“units”) have enrolled a total of 69 patients in the trial, of which 18 have died.

BUGS Example 5: Survival Modeling

- Our data arises from a clinical trial comparing two treatments for *Mycobacterium avium complex (MAC)*, a disease common in late stage HIV-infected persons. Eleven clinical centers (“units”) have enrolled a total of 69 patients in the trial, of which 18 have died.
- For $j = 1, \dots, n_i$ and $i = 1, \dots, k$, let
 - t_{ij} = time to death or censoring
 - x_{ij} = treatment indicator for subject j in stratum i

BUGS Example 5: Survival Modeling

- Our data arises from a clinical trial comparing two treatments for *Mycobacterium avium complex (MAC)*, a disease common in late stage HIV-infected persons. Eleven clinical centers (“units”) have enrolled a total of 69 patients in the trial, of which 18 have died.
- For $j = 1, \dots, n_i$ and $i = 1, \dots, k$, let
 - t_{ij} = time to death or censoring
 - x_{ij} = treatment indicator for subject j in stratum i
- Next page gives survival times (in half-days) from the MAC treatment trial, where “+” indicates a censored observation...

MAC Survival Data

unit	drug	time	unit	drug	time	unit	drug	time
A	1	74+	E	1	214	H	1	74+
A	2	248	E	2	228+	H	1	88+
A	1	272+	E	2	262	H	1	148+
A	2	344				H	2	162
			F	1	6			
B	2	4+	F	2	16+	I	2	8
B	1	156+	F	1	76	I	2	16+
			F	2	80	I	2	40
C	2	100+	F	2	202	I	1	120+
			F	1	258+	I	1	168+
D	2	20+	F	1	268+	I	2	174+
D	2	64	F	2	368+	I	1	268+
D	2	88	F	1	380+	I	2	276
D	2	148+	F	1	424+	I	1	286+
...
						K	2	106+

MAC Survival Data

- With **proportional hazards** and a **Weibull** baseline hazard, stratum i 's hazard is

$$\begin{aligned}h(t_{ij}; x_{ij}) &= h_0(t_{ij})\omega_i \exp(\beta_0 + \beta_1 x_{ij}) \\ &= \rho_i t_{ij}^{\rho_i - 1} \exp(\beta_0 + \beta_1 x_{ij} + W_i) ,\end{aligned}$$

where $\rho_i > 0$, $\beta = (\beta_0, \beta_1)' \in \mathbb{R}^2$, and $W_i = \log \omega_i$ is a clinic-specific **frailty** term.

MAC Survival Data

- With **proportional hazards** and a **Weibull** baseline hazard, stratum i 's hazard is

$$\begin{aligned}h(t_{ij}; x_{ij}) &= h_0(t_{ij})\omega_i \exp(\beta_0 + \beta_1 x_{ij}) \\ &= \rho_i t_{ij}^{\rho_i - 1} \exp(\beta_0 + \beta_1 x_{ij} + W_i) ,\end{aligned}$$

where $\rho_i > 0$, $\beta = (\beta_0, \beta_1)' \in \mathbb{R}^2$, and $W_i = \log \omega_i$ is a clinic-specific **frailty** term.

- The W_i capture overall differences among the clinics, while the ρ_i allow differing baseline hazards which either increase ($\rho_i > 1$) or decrease ($\rho_i < 1$) over time. We assume i.i.d. specifications for these random effects,

$$W_i \stackrel{iid}{\sim} N(0, 1/\tau) \quad \text{and} \quad \rho_i \stackrel{iid}{\sim} G(\alpha, \alpha) .$$

MAC Survival Data

- As in the `mice` example (WinBUGS Examples Vol 1),

$$\mu_{ij} = \exp(\beta_0 + \beta_1 x_{ij} + W_i) ,$$

so that

$$t_{ij} \sim Weibull(\rho_i, \mu_{ij}) .$$

MAC Survival Data

- As in the `mice` example (WinBUGS Examples Vol 1),

$$\mu_{ij} = \exp(\beta_0 + \beta_1 x_{ij} + W_i) ,$$

so that

$$t_{ij} \sim Weibull(\rho_i, \mu_{ij}) .$$

- We **recode** the drug covariate from (1,2) to (-1,1) (i.e., set $x_{ij} = 2drug_{ij} - 3$) to **ease collinearity** between the slope β_1 and the intercept β_0 .

MAC Survival Data

- As in the `mice` example (WinBUGS Examples Vol 1),

$$\mu_{ij} = \exp(\beta_0 + \beta_1 x_{ij} + W_i) ,$$

so that

$$t_{ij} \sim Weibull(\rho_i, \mu_{ij}) .$$

- We **recode** the drug covariate from (1,2) to (-1,1) (i.e., set $x_{ij} = 2drug_{ij} - 3$) to **ease collinearity** between the slope β_1 and the intercept β_0 .
- We place vague priors on β_0 and β_1 , a moderately informative $G(1, 1)$ prior on τ , and set $\alpha = 10$.

MAC Survival Data

- As in the `mice` example (WinBUGS Examples Vol 1),

$$\mu_{ij} = \exp(\beta_0 + \beta_1 x_{ij} + W_i) ,$$

so that

$$t_{ij} \sim Weibull(\rho_i, \mu_{ij}) .$$

- We **recode** the drug covariate from (1,2) to (-1,1) (i.e., set $x_{ij} = 2drug_{ij} - 3$) to **ease collinearity** between the slope β_1 and the intercept β_0 .
- We place vague priors on β_0 and β_1 , a moderately informative $G(1, 1)$ prior on τ , and set $\alpha = 10$.
- Data:** www.biostat.umn.edu/~brad/data/MAC.dat
Code: www.biostat.umn.edu/~brad/data/MACfrailty_BUGS.txt

MAC Survival Results

node (unit)	mean	sd	MC error	2.5%	median	97.5%
W_1 (A)	-0.04912	0.835	0.02103	-1.775	-0.04596	1.639
W_3 (C)	-0.1829	0.9173	0.01782	-2.2	-0.1358	1.52
W_5 (E)	-0.03198	0.8107	0.03193	-1.682	-0.02653	1.572
W_6 (F)	0.4173	0.8277	0.04065	-1.066	0.3593	2.227
W_9 (I)	0.2546	0.7969	0.03694	-1.241	0.2164	1.968
W_{11} (K)	-0.1945	0.9093	0.02093	-2.139	-0.1638	1.502
ρ_1 (A)	1.086	0.1922	0.007168	0.7044	1.083	1.474
ρ_3 (C)	0.9008	0.2487	0.006311	0.4663	0.8824	1.431
ρ_5 (E)	1.143	0.1887	0.00958	0.7904	1.139	1.521
ρ_6 (F)	0.935	0.1597	0.008364	0.6321	0.931	1.265
ρ_9 (I)	0.9788	0.1683	0.008735	0.6652	0.9705	1.339
ρ_{11} (K)	0.8807	0.2392	0.01034	0.4558	0.8612	1.394
τ	1.733	1.181	0.03723	0.3042	1.468	4.819
β_0	-7.111	0.689	0.04474	-8.552	-7.073	-5.874
β_1	0.596	0.2964	0.01048	0.06099	0.5783	1.245
RR	3.98	2.951	0.1122	1.13	3.179	12.05

MAC Survival Results

- Units A and E have moderate overall risk ($W_i \approx 0$) but increasing hazards ($\rho > 1$): few deaths, but they occur late

MAC Survival Results

- Units A and E have moderate overall risk ($W_i \approx 0$) but increasing hazards ($\rho > 1$): few deaths, but they occur late
- Units F and I have high overall risk ($W_i > 0$) but decreasing hazards ($\rho < 1$): several early deaths, many long-term survivors

MAC Survival Results

- Units A and E have moderate overall risk ($W_i \approx 0$) but increasing hazards ($\rho > 1$): few deaths, but they occur late
- Units F and I have high overall risk ($W_i > 0$) but decreasing hazards ($\rho < 1$): several early deaths, many long-term survivors
- Units C and K have low overall risk ($W_i < 0$) and decreasing hazards ($\rho < 1$): no deaths at all; a few survivors

MAC Survival Results

- Units A and E have moderate overall risk ($W_i \approx 0$) but increasing hazards ($\rho > 1$): few deaths, but they occur late
- Units F and I have high overall risk ($W_i > 0$) but decreasing hazards ($\rho < 1$): several early deaths, many long-term survivors
- Units C and K have low overall risk ($W_i < 0$) and decreasing hazards ($\rho < 1$): no deaths at all; a few survivors
- Drugs differ significantly: CI for β_1 (RR) excludes 0 (1)

MAC Survival Results

- Units A and E have moderate overall risk ($W_i \approx 0$) but increasing hazards ($\rho > 1$): few deaths, but they occur late
 - Units F and I have high overall risk ($W_i > 0$) but decreasing hazards ($\rho < 1$): several early deaths, many long-term survivors
 - Units C and K have low overall risk ($W_i < 0$) and decreasing hazards ($\rho < 1$): no deaths at all; a few survivors
 - Drugs differ significantly: CI for β_1 (RR) excludes 0 (1)
-
- **Note:** This has all been for two sets of random effects (W_i and ρ_i), called “Model 2” in the BUGS code. You will also see models having three (adding β_{1i}), one (deleting ρ_i), or zero sets of random effects!

BRugs Example 1: Meta-analysis

- This example could be done using only WinBUGS, but then the results are not objects in R (plus it's nice to avoid all the pointing and clicking...)
- We consider a hierarchical **beta-binomial** model for meta-analysis of success proportions
- For nine studies with the same drug, we have the number of successes, x_i , and the number of patients, n_i

study	1	2	3	4	5	6	7	8	9
x_i	20	4	11	10	5	36	9	7	4
n_i	20	10	16	19	14	46	10	9	6
$\hat{p}_i = \frac{x_i}{n_i}$	1.00	0.40	0.69	0.53	0.36	0.78	0.90	0.78	0.67

- We will fit a hierarchical model with **three** levels

Meta-analysis using BRugs

- **BRugs** is a suite of R routines for calling `OpenBUGS` from R, originally written by `WinBUGS` head programmer Andrew Thomas, and refined and maintained by Uwe Ligges
- All necessary programs and instructions can be downloaded from www.biostat.umn.edu/~brad/software/BRugs
- Note that we will now have **two** text files with code:
 - an **R** program that organizes the dataset, contains all the `BRugs` commands, and summarizes the output
 - a piece of **BUGS** code that is sent by R to `OpenBUGS` (must be saved in working directory)
- `BRugs` code and data (for this and all remaining examples in this handout): also available from www.biostat.umn.edu/~brad/software/BRugs

Binomial Meta-analysis

- Suppose that within study i the patients are exchangeable, in the sense that all have the same propensity p_i of success
- **Goal:** To estimate the proportion of success for a **new person** in each study, p_i , **and** for a **new tenth study**
- The three stages of the model are likelihood, prior and hyperprior:

$$\begin{aligned}x_i | p_i &\stackrel{ind}{\sim} \text{Binomial}(n_i, p_i) \\ p_i | a, b &\stackrel{iid}{\sim} \text{Beta}(a, b) \\ a, b &\stackrel{iid}{\sim} \text{Unif}(0, 10)\end{aligned}$$

Warning: Unbounded (improper) flat hyperprior for a, b here leads to **improper** posterior! (Hadjicostas, 1998)

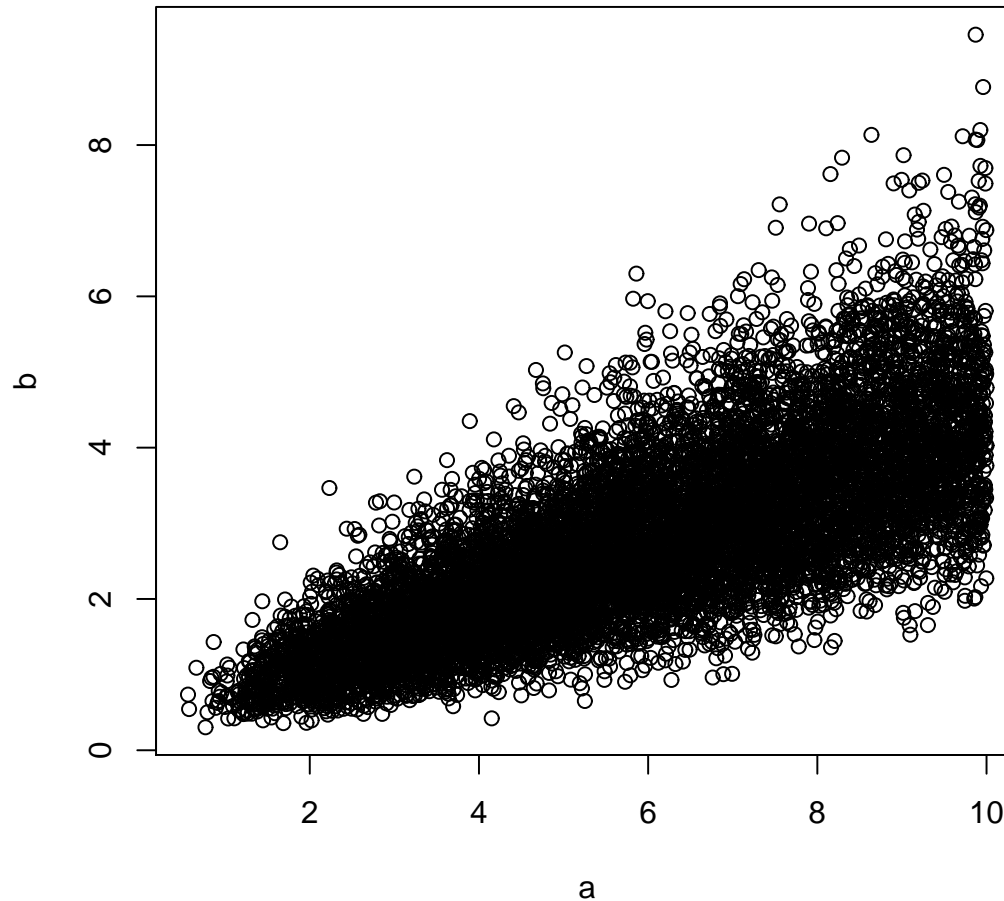
Running BRugs

- Using BRugs, a minimum of four commands are needed to run a Bayesian model:
 - `library(BRugs)`
 - `bugsData`
 - `bugsInits`
 - `BRugsFit`
- Then the MCMC samples for the parameters specified in the `BRugsFit` command are available for analysis with both BUGS commands **and** ordinary R commands
- To get a trace plot of the MCMC chain for new study 10, use either:
 - `samplesHistory("p[10]")` or
 - `plot(samplesSample("p[10]"), type='l')`

Binomial Meta-analysis

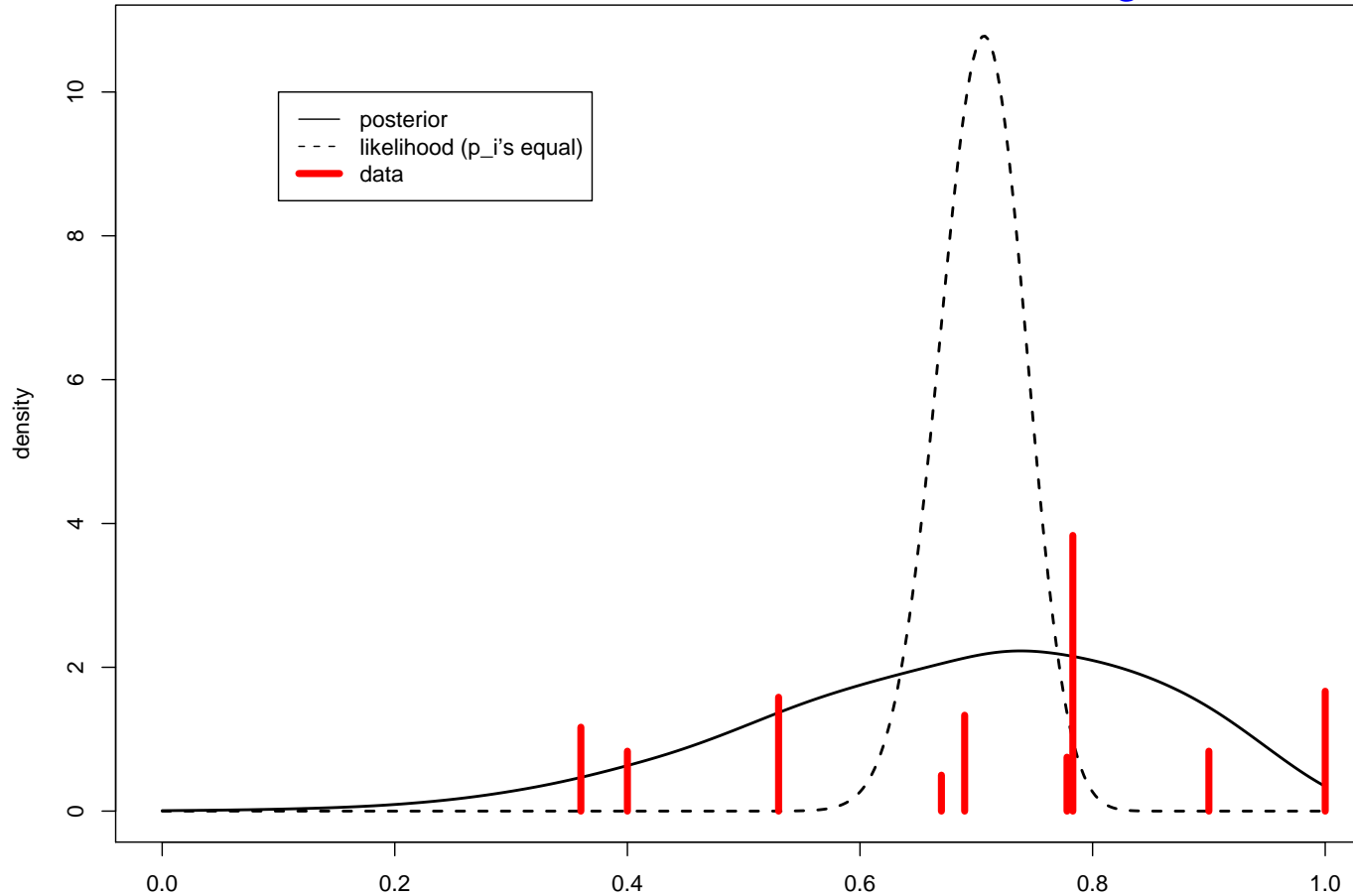
- Estimating the joint posterior of a and b is easily done via a plot of their matched Gibbs pairs:

```
plot(samplesSample("a"), samplesSample("b"), xlab="a", ylab="b")
```



Avoid the “truncated wedge” via: $a, b \stackrel{iid}{\sim} \text{Gamma}(2, 2)$?

Binomial Meta-analysis



- The nine vertical bars correspond to the observed proportions for the nine studies, with bar heights proportional to sample sizes
- The posterior mean of p_{10} is 0.68, which can be found by averaging all 10,000 Gibbs samples for $p[10]$

BRugs Example 2: Model assessment

- Basic tool here is the **cross-validation** residual

$$r_i = y_i - E(y_i | \mathbf{y}_{(i)}) ,$$

where $\mathbf{y}_{(i)}$ denotes the vector of all the data except the i^{th} value, i.e.

$$\mathbf{y}_{(i)} = (y_1, \dots, y_{i-1}, y_{i+1}, \dots, y_n)' .$$

Outliers are indicated by large **standardized** residuals,

$$d_i = r_i / \sqrt{\text{Var}(y_i | \mathbf{y}_{(i)})} .$$

- Also of interest is the **conditional predictive ordinate**, $p(y_i | \mathbf{y}_{(i)}) = \int p(y_i | \boldsymbol{\theta}, \mathbf{y}_{(i)}) p(\boldsymbol{\theta} | \mathbf{y}_{(i)}) d\boldsymbol{\theta}$, the height of the conditional density at the observed value of y_i
 \implies large values indicate good prediction of y_i .

Residuals: Approximate method

- Using MC draws $\boldsymbol{\theta}^{(g)} \sim p(\boldsymbol{\theta}|\mathbf{y})$, we have

$$\begin{aligned} E(y_i|\mathbf{y}_{(i)}) &= \int \int y_i f(y_i|\boldsymbol{\theta})p(\boldsymbol{\theta}|\mathbf{y}_{(i)})dy_id\boldsymbol{\theta} \\ &= \int E(y_i|\boldsymbol{\theta})p(\boldsymbol{\theta}|\mathbf{y}_{(i)})d\boldsymbol{\theta} \\ &\approx \int E(y_i|\boldsymbol{\theta})p(\boldsymbol{\theta}|\mathbf{y})d\boldsymbol{\theta} \\ &\approx \frac{1}{G} \sum_{g=1}^G E(y_i|\boldsymbol{\theta}^{(g)}) . \end{aligned}$$

- Approximation should be adequate unless the dataset is small and y_i is an extreme outlier
- Same** $\boldsymbol{\theta}^{(g)}$'s may be used for each $i = 1, \dots, n$.

Approximate methods in WinBUGS

- The ratio to compute the standardized residuals d_i must be done outside of WinBUGS. Might instead define

$$d_i^* = \frac{y_i - E(y_i|\boldsymbol{\theta})}{\sqrt{Var(y_i|\boldsymbol{\theta})}} .$$

We then find $E(d_i^*|\mathbf{y})$, the posterior average of the ratio (instead of the ratio of the posterior averages).

- For the exact method, we must evaluate $E(y_i|\mathbf{y}_{(i)})$ and $Var(y_i|\mathbf{y}_{(i)})$ separately. For the latter, use the facts that $Var(y_i|\mathbf{y}_{(i)}) = E(y_i^2|\mathbf{y}_{(i)}) - [E(y_i|\mathbf{y}_{(i)})]^2$, and

$$\begin{aligned} E(y_i^2|\mathbf{y}_{(i)}) &= \int E(y_i^2|\boldsymbol{\theta})p(\boldsymbol{\theta}|\mathbf{y}_{(i)})d\boldsymbol{\theta} \\ &= \int \{Var(y_i|\boldsymbol{\theta}) + [E(y_i|\boldsymbol{\theta})]^2\}p(\boldsymbol{\theta}|\mathbf{y}_{(i)})d\boldsymbol{\theta} . \end{aligned}$$

Residuals: Exact method

- An exact solution then arises by calling BUGS n times, once for each “leave one out” dataset!
- This can be easily accomplished in **BRugs**: we simply nest the BUGS calls inside an outer loop in R!

Numerical illustration: Stack Loss data

- An oft-analyzed dataset, featuring the stack loss Y (ammonia escaping), and three covariates X_1 (air flow), X_2 (temperature), and X_3 (acid concentration).
- Fit the linear regression model

$$Y_i \sim N(\beta_0 + \beta_1 z_{i1} + \beta_2 z_{i2} + \beta_3 z_{i3}, \tau),$$

where the z_{ij} are the standardized covariates. We take flat priors on the β s and a Gelman-style noninformative prior on $\sigma = 1/\sqrt{\tau}$.

- WinBUGS code and data for **approximate** method:
www.biostat.umn.edu/~brad/data/stacks_BUGS.txt
- BRugs code and data for **exact** method:
www.biostat.umn.edu/~brad/software/BRugs
- See also “**stacks**” in WinBUGS Examples Volume I!

Approximate vs. Exact Results

obs	sresid		CPO	
	approx	exact	approx	exact
1	0.948	1.098	0.178	0.124
2	-0.566	-0.628	0.224	0.188
3	1.337	1.461	0.122	0.084
4	1.672	1.851	0.078	0.047
5	-0.504	-0.477	0.251	0.244
⋮	⋮	⋮	⋮	⋮
21	-2.126	-3.012	0.046	0.005

- Approximate residuals are too small, especially for the most outlying observations!
- Approximate CPOs also tend to understate lack of fit

BRugs Example 3: Clinical Trial Design

- Following our MAC survival model, let t_i be the time until death for subject i , with corresponding treatment indicator x_i ($= 0$ or 1 for control and treatment, respectively). Suppose

$$t_i \sim \text{Weibull}(r, \mu_i), \quad \text{where } \mu_i = e^{-(\beta_0 + \beta_1 x_i)} .$$

- Then the baseline hazard function is $\lambda_0(t_i) = r t_i^{r-1}$, and the median survival time for subject i is

$$m_i = [(\log 2) e^{\beta_0 + \beta_1 x_i}]^{1/r} .$$

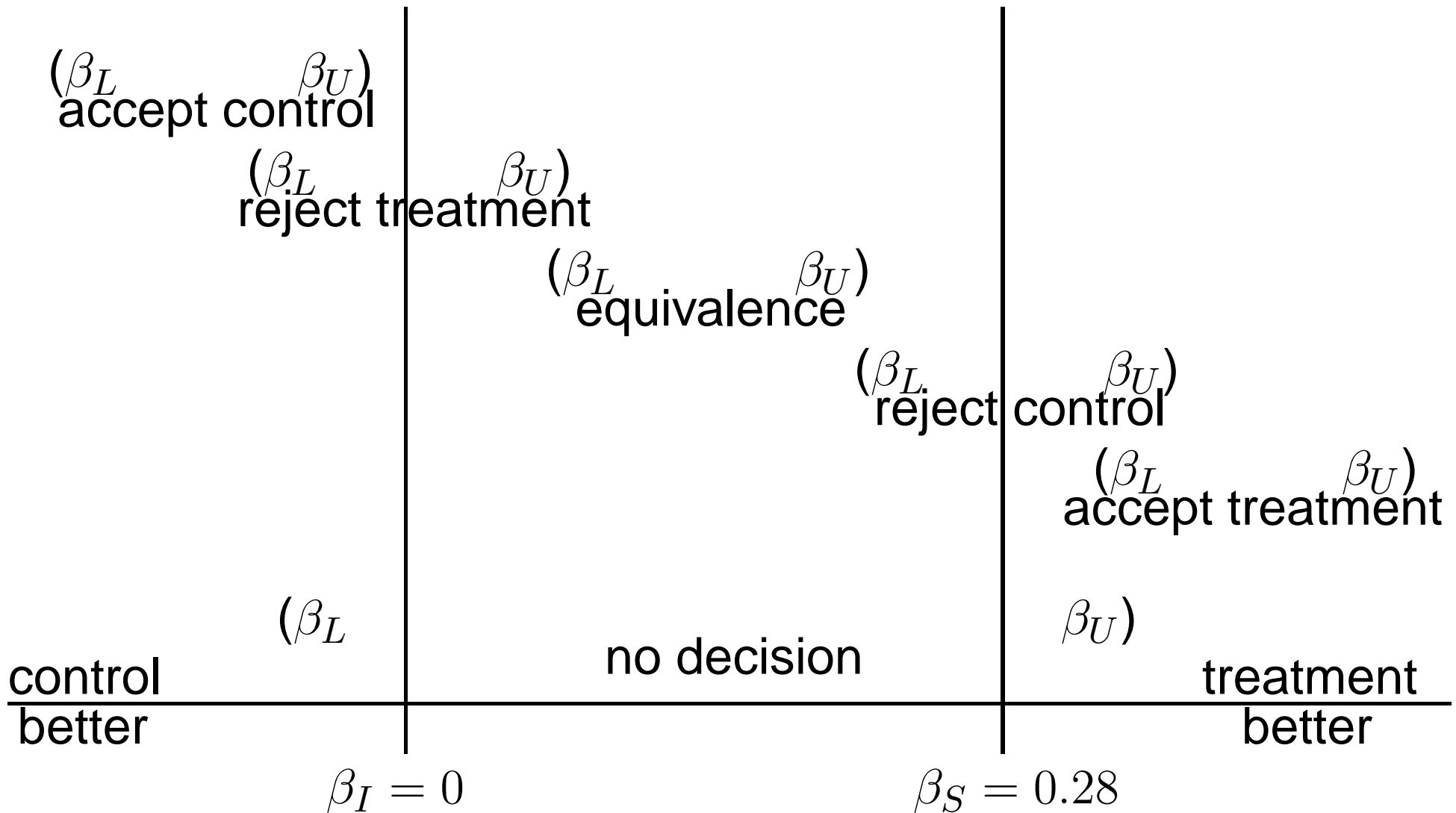
- The value of β_1 corresponding to a **15% increase in median survival in the treatment group** satisfies

$$e^{\beta_1/r} = 1.15 \iff \beta_1 = r \log(1.15) .$$

Range of equivalence

- The range of β_1 values within which we are **indifferent** as to use of treatment or control
- lower limit β_I , the **clinical inferiority** boundary
 - We typically take $\beta_I = 0$, since we would never prefer a harmful treatment
- upper limit β_S , the **clinical superiority** boundary
 - We typically take $\beta_S > 0$, since we may require “clinically significant” improvement under the treatment (due to cost, toxicity, etc.)
 - **Example:** If $r = 2$, then $\beta_S = 2 \log(1.15) \approx 0.28$ corresponds to 15% improvement in median survival
- The outcome of the trial can then be based on the location of the **95% posterior confidence interval** for β_1 , say (β_L, β_U) , relative to the indifference zone!....

The six possible outcomes and decisions



- Note both “acceptance” and “rejection” are possible!

Community of priors

Spiegelhalter et al. (1994) recommend considering several priors, in order to represent the broadest possible audience:

- **Skeptical Prior**

- One that believes the treatment is likely no better than control (as might be believed by the FDA)

- **Enthusiastic (or Clinical) Prior**

- One that believes the treatment will succeed (typical of the clinicians running the trial)

- **Reference (or Noninformative) Prior**

- One that expresses no particular opinion about the treatment's merit
- Often a **improper uniform** (“flat”) prior is permissible

MCMC-based Bayesian design

Simulating the power or other operating characteristics (say, Type I error) in this setting works as follows:

- Sample “true” β values from an assumed “true prior” (skeptical, enthusiastic, or in between)
- Given these, sample fake survival times t_i (say, N from each study group) from the Weibull
- We may also wish to sample fake **censoring** times c_i from a particular distribution (e.g., a normal truncated below 0); for all i such that $t_i > c_i$, replace t_i by “NA”
- Compute (β_L, β_U) **by calling OpenBUGS from R**
- Determine the simulated trial’s outcome based on location of (β_L, β_U) relative to the indifference zone

Results from Power.BRugs

- Assuming:
 - Weibull shape $r = 2$, and $N = 50$ in each group
 - median survival of 36 days with 50% improvement in the treatment group
 - a $N(80, 20)$ censoring distribution
 - the enthusiastic prior as the “truth”

We obtain the following output from $Nrep = 100$ reps:

- Here are simulated outcome frequencies for $N = 50$
 - accept control: 0
 - reject treatment: 0.07
 - equivalence: 0
 - reject control: 0.87
 - accept treatment: 0.06
 - no decision: 0
- End of BRugs power simulation

Homework Problems

● WinBUGS

- PK hierarchical linear model:
www.biostat.umn.edu/~brad/data/PK_BUGS.txt
- PK hierarchical nonlinear model:
www.biostat.umn.edu/~brad/data/PKNL_BUGS.txt
- Interstim multivariate model:
www.biostat.umn.edu/~brad/data/InterStim.odc
- Bayesian p -values (illustrated with stacks data):
www.biostat.umn.edu/~brad/data/stackspval_BUGS.txt

● BRugs

- Design for binary and Cox PH models: Brian Hobbs' webpage:
www.biostat.umn.edu/~brianho/papers/2007/JBS/prac_bayes_design.html

● *Thanks for your attention!*