# Automatic variable selection in hierarchical models: application to durations of untreated psychoses

Dave Lunn [1] James Kirkbride [2]

[1]MRC Biostatistics Unit, Cambridge, UK

[2]Dept. Psychiatry, University of Cambridge, UK

July 1, 2008

# Outline

- Background
- DUP data
- Random effects models
- Mixed effects models
- Variable selection
- Interpretation of results

# Outline

- Background
- DUP data
- Random effects models
- Mixed effects models
- Variable selection
- Interpretation of results

# Outline

- Background
- DUP data
- Random effects models
- Mixed effects models
- Variable selection
- Interpretation of results

# Outline

- Background
- DUP data
- Random effects models
- Mixed effects models
- Variable selection
- Interpretation of results

# Background

- Family involvement in help-seeking associated with shorter duration of untreated psychoses (DUP)
- Are factors at neighbourhood level relevant to DUP?
- Recent study in Netherlands found +ve relationship between hospital admissions and neighbourhood level of informal social control
- Perhaps neighbourhoods with greater social cohesion are associated with shorter DUPs?

# Background

- Family involvement in help-seeking associated with shorter duration of untreated psychoses (DUP)
- Are factors at neighbourhood level relevant to DUP?
- Recent study in Netherlands found +ve relationship between hospital admissions and neighbourhood level of informal social control
- Perhaps neighbourhoods with greater social cohesion are associated with shorter DUPs?

# Background

- Family involvement in help-seeking associated with shorter duration of untreated psychoses (DUP)
- Are factors at neighbourhood level relevant to DUP?
- Recent study in Netherlands found +ve relationship between hospital admissions and neighbourhood level of informal social control
- Perhaps neighbourhoods with greater social cohesion are associated with shorter DUPs?

# Background

- Family involvement in help-seeking associated with shorter duration of untreated psychoses (DUP)
- Are factors at neighbourhood level relevant to DUP?
- Recent study in Netherlands found +ve relationship between hospital admissions and neighbourhood level of informal social control
- Perhaps neighbourhoods with greater social cohesion are associated with shorter DUPs?

# DUP data

- Data from SE London centre of ÆSOP first episode psychoses study
- 314 durations from individuals in 32 electoral wards
- Mode number of durations = 7; min = 1, max = 31
- Durations range from 1 day to 9118 days (25 years)
  median = 70 days, mean = 450 days!!
- Distribution a little skewed perhaps...

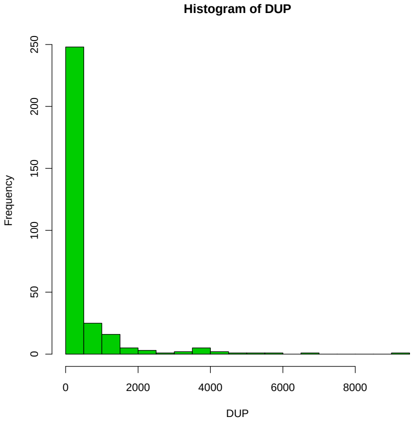
# DUP data

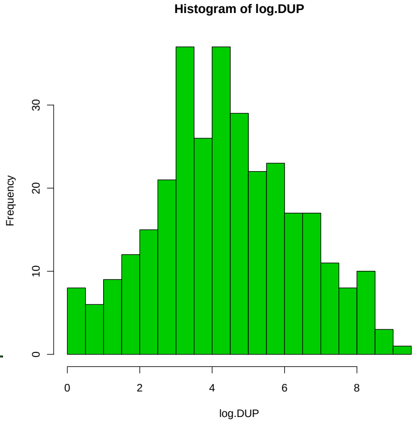
- Data from SE London centre of ÆSOP first episode psychoses study
- 314 durations from individuals in 32 electoral wards
- Mode number of durations = 7; min = 1, max = 31
- Durations range from 1 day to 9118 days (25 years)
  median = 70 days, mean = 450 days!!
- Distribution a little skewed perhaps...

# DUP data

- Data from SE London centre of ÆSOP first episode psychoses study
- 314 durations from individuals in 32 electoral wards
- Mode number of durations = 7; min = 1, max = 31
- Durations range from 1 day to 9118 days (25 years)
  median = 70 days, mean = 450 days!!
- Distribution a little skewed perhaps...

# Distribution of DUPs




- Measurement error model?

# Covariates

| Neighbourhood level | Individual level |
|---|---|
| ◇ population density | ◇ gender |
| ◇ proportion non-white British | ◇ type of disorder (affective or broad schizophrenia) |
| ◇ proportion black Caribbean | ◇ ethnicity, either binary or 7-level |
| ◇ proportion black African | ◇ relationship status (3-level) |
| ◇ voter turnout | ◇ highest ever occupation (6-level) |
| ◇ segregation between white and non-white British | ◇ age |
| ◇ proportion single | ◇ family involvement |
| ◇ proportion of Asian groups | ◇ mode of onset, binary or 6-level |
| ◇ multiple deprivation | |
| ◇ social capital | |
| ◇ social disorganisation | |
| ◇ social cohesion | |

# Basic random effects model

- Let $T_{ij} = \log(\mathrm{DUP}_{ij})$, where $\mathrm{DUP}_{ij}$ is the DUP for individual $j$ in neighbourhood $i$ ($i = 1, ..., N = 32$, $j = 1, ..., n_i$)

$$T_{ij} = \alpha + u_i + e_{ij}$$
$$u_i \sim \mathrm{N}(0, \sigma_u^2), \; e_{ij} \sim \mathrm{N}(0, \sigma_e^2)$$

- Or, equivalently: $T_{ij} \sim \mathrm{N}(\mu_i, \sigma_e^2)$, $\mu_i = \alpha + u_i$
- Could also allow spatially correlated random effects, e.g.

$$u_i = r_i + s_i, \quad r_i \sim \mathrm{N}(0, \sigma_r^2), \quad \mathbf{s} \sim \mathrm{CAR}(\sigma_s^2)$$

# Basic random effects model

- Let $T_{ij} = \log(\mathrm{DUP}_{ij})$, where $\mathrm{DUP}_{ij}$ is the DUP for individual $j$ in neighbourhood $i$ ($i = 1, ..., N = 32$, $j = 1, ..., n_i$)

$$T_{ij} = \alpha + u_i + e_{ij}$$
$$u_i \sim \mathrm{N}(0, \sigma_u^2), \;\; e_{ij} \sim \mathrm{N}(0, \sigma_e^2)$$

- Or, equivalently: $T_{ij} \sim \mathrm{N}(\mu_i, \sigma_e^2)$, $\mu_i = \alpha + u_i$
- Could also allow spatially correlated random effects, e.g.

$$u_i = r_i + s_i, \quad r_i \sim \mathrm{N}(0, \sigma_r^2), \quad \mathbf{s} \sim \mathrm{CAR}(\sigma_s^2)$$

# Basic random effects model

- Let $T_{ij} = \log(\text{DUP}_{ij})$, where $\text{DUP}_{ij}$ is the DUP for individual $j$ in neighbourhood $i$ ($i = 1, ..., N = 32$, $j = 1, ..., n_i$)

$$T_{ij} = \alpha + u_i + e_{ij}$$
$$u_i \sim \text{N}(0, \sigma_u^2),\ \ e_{ij} \sim \text{N}(0, \sigma_e^2)$$

- Or, equivalently: $T_{ij} \sim \text{N}(\mu_i, \sigma_e^2)$, $\mu_i = \alpha + u_i$
- Could also allow spatially correlated random effects, e.g.

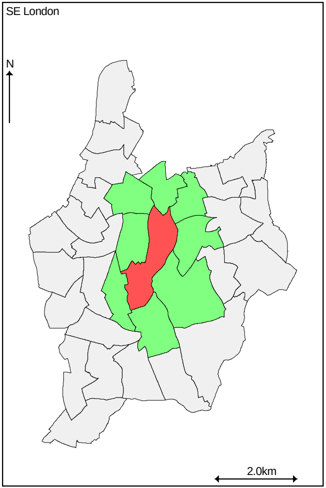
$$u_i = r_i + s_i, \quad r_i \sim \text{N}(0, \sigma_r^2), \quad \mathbf{s} \sim \text{CAR}(\sigma_s^2)$$

# Basic CAR model

$$s_i | \mathbf{s}_{-i} \sim \text{Normal}\left( \frac{1}{\nu_i} \sum_{j \in \delta_i} s_j,\ \frac{\sigma_s^2}{\nu_i} \right)$$

$\delta_i$ = set of neighbours, area $i$

$\nu_i$ = no. of neighbours, area $i$

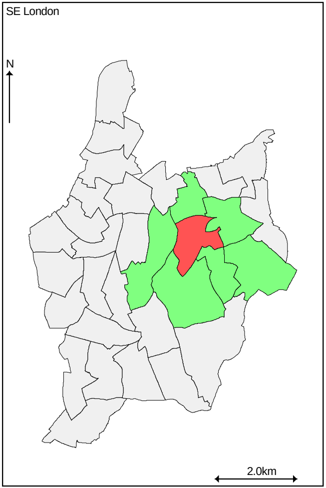


# Basic CAR model

$$s_i \big| \mathbf{s}_{-i} \sim \text{Normal}\left( \frac{1}{\nu_i} \sum_{j \in \delta_i} s_j, \; \frac{\sigma_s^2}{\nu_i} \right)$$

$\delta_i$ = set of neighbours, area $i$

$\nu_i$ = no. of neighbours, area $i$

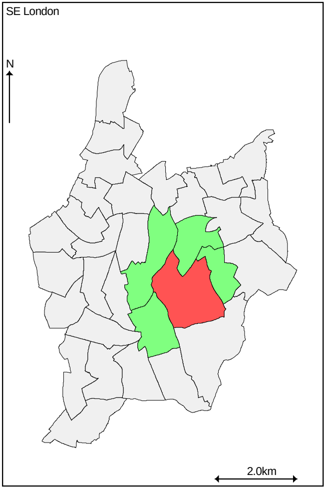


# Basic CAR model

$$s_i \big| \mathbf{s}_{-i} \sim \text{Normal}\left( \frac{1}{\nu_i} \sum_{j \in \delta_i} s_j,\ \frac{\sigma_s^2}{\nu_i} \right)$$

$\delta_i =$ set of neighbours, area $i$

$\nu_i =$ no. of neighbours, area $i$



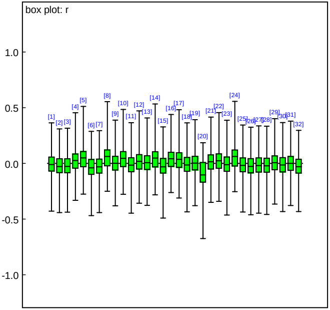
# BUGS code

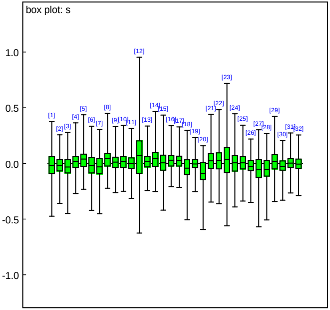
```
model {
  for (i in 1:N) {
    for (j in start[i]:(start[i + 1] - 1)) {
      T[j] ~ dnorm(mu[i], tau.e)
    }
    mu[i] <- alpha + u[i]
    u[i] ~ dnorm(0, tau.u)
  }
  ...
}
```

# BUGS code

```
model {
  for (i in 1:N) {
    for (j in start[i]:(start[i + 1] - 1)) {
      T[j] ~ dnorm(mu[i], tau.e)
    }
    mu[i] <- alpha + r[i] + s[i]
    r[i] ~ dnorm(0, tau.r)
  }
  s[1:N] ~ car.normal(adj[], weights[], nu[], tau.s)
  ...
}
```

# Uncorrelated and spatially correlated random effects

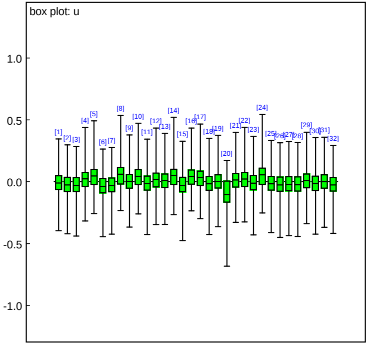



$$\hat{\alpha} = 4.39 \quad (4.15, 4.62), \quad \hat{\sigma_e} = 1.97 \quad (1.82, 2.14)$$

$$\hat{\sigma_r} = 0.129 \quad (0.00715, 0.440), \quad \hat{\sigma_s} = 0.185 \quad (0.00592, 0.696)$$

# Uncorrelated random effects only




$$\hat{\alpha} = 4.39 \quad (4.16, 4.62)$$

$$\hat{\sigma_e} = 1.97 \quad (1.83, 2.14)$$

$$\hat{\sigma_u} = 0.120 \quad (0.0065, 0.419)$$

# Which model is better?

- Spiegelhalter et al (2002, RSSB) propose *Deviance Information Criterion* (DIC) for comparing models
- Model fit measured by posterior mean deviance $\overline{D}$ (deviance $= -2 \times$ log-likelihood)
- Penalized by 'effective number of parameters' $p_D = \overline{D} - D(\overline{\theta})$ (posterior mean deviance $-$ deviance evaluated at posterior mean parameters)
- $\text{DIC} = \overline{D} + p_D = D(\overline{\theta}) + 2p_D$
- Models with smaller DIC are better supported by data...

# Compare DICs

| Model | $\overline{D}$ | $D(\overline{\theta})$ | $p_D$ | DIC |
|---|---|---|---|---|
| Uncorr. + spat. corr. | 1308.14 | 1308.14 | 6.071 | 1314.21 |
| Uncorrelated only | 1308.17 | 1304.01 | 4.158 | 1312.33 |
| No area-level effects | 1308.63 | 1306.66 | 1.97 | 1310.60 |

$\hat{\alpha} = 4.39 \quad (4.17, 4.61)$

$\hat{\sigma_e} = 1.97 \quad (1.83, 2.14)$

# Compare DICs

| Model | $\overline{D}$ | $D(\overline{\theta})$ | $p_D$ | DIC |
|---|---|---|---|---|
| Uncorr. + spat. corr. | 1308.14 | 1308.14 | 6.071 | 1314.21 |
| Uncorrelated only | 1308.17 | 1304.01 | 4.158 | 1312.33 |
| No area-level effects | 1308.63 | 1306.66 | 1.97 | 1310.60 |

```
model {
  for (i in 1:N) {
    for (j in start[i]:(start[i + 1] - 1)) {
      T[j] ~ dnorm(mu[i], tau.e)
    }
    mu[i] <- alpha
  }
  ...
}
```

$$\hat{\alpha} = 4.39 \quad (4.17, 4.61)$$

$$\hat{\sigma_e} = 1.97 \quad (1.83, 2.14)$$

# Variable selection

- Model:

$$T_j \sim \mathrm{N}(\mu_j, \sigma_e^2), \quad \mu_j = \alpha + \mathrm{X}_j\beta, \quad j = 1, ..., N_{\text{tot}} = 314$$

  where $\mathrm{X}_j$ is a row-vector of area-level and/or individual-level covariates for observation $j$ ($X_j$ is $j^{\text{th}}$ row of matrix $\mathbf{X}$)
- Which covariates should form the columns of $\mathbf{X}$?
- Forward/backward selection can be unreliable
- Include all of them?
  - interpretability issues
  - spurious effects

# Variable selection

- Model:

$$T_j \sim \mathrm{N}(\mu_j, \sigma_e^2), \quad \mu_j = \alpha + \mathrm{X}_j\beta, \quad j = 1, ..., N_{\mathrm{tot}} = 314$$

  where $\mathrm{X}_j$ is a row-vector of area-level and/or individual-level covariates for observation $j$ ($X_j$ is $j^{\mathrm{th}}$ row of matrix $\mathbf{X}$)
- Which covariates should form the columns of $\mathbf{X}$?
- Forward/backward selection can be unreliable
- Include all of them?
  - interpretability issues
  - spurious effects

# Variable selection continued

- Let **Z** denote the matrix of *all* area- and individual-level covariates
- Let $\phi$ denote the set of column indices that form **X**:

$$\mathrm{X}_{jk} = \mathrm{Z}_{j\phi_k}, \quad k = 1, ..., d = \dim(\phi)$$

- Perhaps we can 'estimate' $\phi$?
- Note, however, that it has variable dimension, as has $\beta$ ($X_j$ expands and contracts as covariates are included in and discarded from the model)
- Can handle in Bayesian setting using 'reversible jump'

# Variable selection continued

- Let **Z** denote the matrix of *all* area- and individual-level covariates
- Let $\phi$ denote the set of column indices that form **X**:

$$\mathrm{X}_{jk} = \mathrm{Z}_{j\phi_k}, \quad k = 1, ..., d = \mathrm{dim}(\phi)$$

- Perhaps we can 'estimate' $\phi$?
- Note, however, that it has variable dimension, as has $\beta$ ($X_j$ expands and contracts as covariates are included in and discarded from the model)
- Can handle in Bayesian setting using 'reversible jump'

# Priors for variable selection

- Need to specify priors for $\beta$, $\phi$ and $d$
- Typically

$$\beta|d \sim \mathrm{MVN}_d(\mathbf{0}_d, \sigma_\beta^2 \mathbf{I}_d) \quad \text{(e.g. } \sigma_\beta^2 = 50)$$

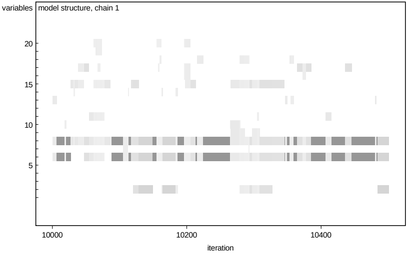
$$\phi|d \sim \text{Uniform} \Rightarrow p(\phi|d) = \frac{d!(Q-d)!}{Q!}$$

$$d \sim \text{Poisson}(.), \quad d \sim \text{Binomial}(p, Q) \quad \text{(e.g. } p = 0.5)$$

# BUGS code

```
model {
  for (i in 1:N) {
    for (j in start[i]:(start[i + 1] - 1)) {
      T[j] ~ dnorm(mu[j], tau.e)
    }
  }
  mu[1:N.tot] <- jump.lin.pred(X[,], d, prec)
  d ~ dbin(0.5, Q)
  ...
}
```

# BUGS output

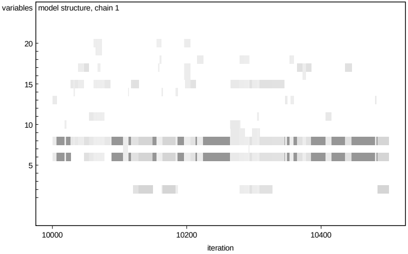


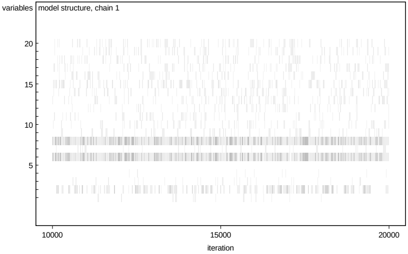

var #2 = affective/broad schizophrenia (binary)

var #6 = age (continuous)

var #8 = mode of onset (6-level)

# BUGS output




var $\#2$ = affective/broad schizophrenia (binary)

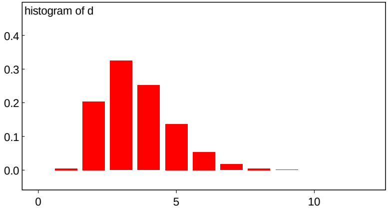
var $\#6$ = age (continuous)

var $\#8$ = mode of onset (6-level)

# BUGS output continued

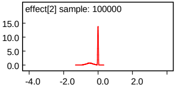
| model structure | posterior prob. | cumulative prob. |
|---|---|---|
| 00000101000000000000 | 0.1943107039 | 0.1943107039 |
| 01000101000000000000 | 0.09628279883 | 0.2905935027 |
| 00000101000000100000 | 0.04008329863 | 0.3306768013 |
| 01000101000000100000 | 0.0231153686 | 0.3537921699 |
| 00000101000010000000 | 0.02144314869 | 0.3752353186 |
| ... | | |

| variable no. | marginal prob. |
|---|---|
| 1 | 0.02665139525 |
| 2 | 0.3389046231 |
| 3 | 0.02906497293 |
| 4 | 0.0168971262 |
| 5 | 0.0 |
| 6 | 0.9702936277 |
| 7 | 0.03944814661 |
| 8 | 1.0 |
| 9 | 0.08227821741 |
| ... | |



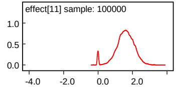
# Effect sizes

$\hat{\sigma_e}$ : $1.97 \longrightarrow 1.50$



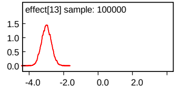

*affective*?

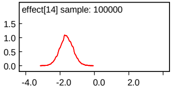
$\mathbb{E}[\beta_2 | 2 \in \phi, T] = -0.488$




*age*

$\mathbb{E}[\beta_{11} | 6 \in \phi, T] = 1.62$



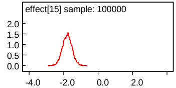



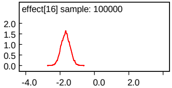

*sudden onset*

$\mathbb{E}[\beta_{13} | 8 \in \phi, T] = -3.00$

*precipitous*

$\mathbb{E}[\beta_{14} | 8 \in \phi, T] = -1.63$



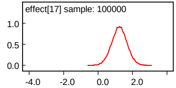




*acute, no symptoms*

$\mathbb{E}[\beta_{15} | 8 \in \phi, T] = -1.79$

*acute, symptoms*

$\mathbb{E}[\beta_{16} | 8 \in \phi, T] = -1.68$

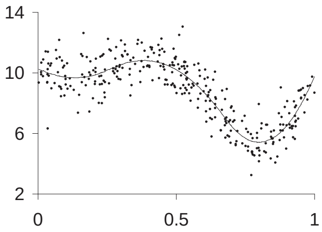



*no clear demarcation*

$\mathbb{E}[\beta_{17} | 8 \in \phi, T] = 1.24$

# Comments

- Can use for selecting 'best' model or for averaging inferences over all (or some) models – *Bayesian Model Averaging*
- Advisable to transform all covariates onto same scale
- Care needed when choosing priors – sensitivity
- Can also use 'reversible jump' for other modelling challenges... e.g. controlling for complex covariate effects



# Acknowledgements


- Joint work with James Kirkbride in Dept. Psychiatry, University of Cambridge
- Thanks to Craig Morgan, Julia Lappin, Paul Fearon and Paola Dazzan (Institute of Psychiatry, King's College London) for the DUP data

# References


- DIC paper: Spiegelhalter, D. J., Best, N. G., Carlin, B. P. and van der Linde, A. (2002) Bayesian measures of model complexity and fit (with discussion), *Journal of the Royal Statistical Society, Series B*, **64**, 583–639
- WinBUGS website: http://www.mrc-bsu.cam.ac.uk/bugs/
- OpenBUGS website: http://mathstat.helsinki.fi/openbugs/
- WinBUGS Development website (e.g. Jump): http://www.winbugs-development.org.uk/