The consequences of misspecifying the random effects distribution when fitting generalized linear mixed models

John M. Neuhaus Charles E. McCulloch

Division of Biostatistics
University of California, San Francisco

Outline

1. Motivation/Background
2. Parameter estimation
3. Prediction

Antibotic Rx Study in acute care
(Gonzales et al, Academic Emergency Medicine 2006)

- Study objectives: 1) assess intervention to reduce inappropriate antibiotic prescription rates - Rx for antibiotic-nonresponsive conditions (e.g. acute bronchitis); 2) identify poorly performing providers
- Design: multiple responses from 720 providers (clusters)
- Large between-provider variability in RX rates & cluster sizes - borrow strength between providers

Popular approach: Use estimated regression coefficients & predicted random effects from GLMMs

Generalized Linear Mixed Models

$$Y_{ij} \mid b_i, X_{ij} \sim \text{GLM}$$

$$E(Y_{ij}|X_{ij}, b_i) = h^{-1}(\beta_0 + b_i + \beta_1 X_{ij})$$

$h(\cdot)$ link function, e.g. identity, logit, probit

i=1,$\ldots, m$: subjects; j=1,$\ldots, n_i$: repeats per cluster

When $b \sim \mathrm{G_T}$, $Y_{i1}, \ldots, Y_{n_i}$ cond'ly indep given b_i & b_i indep of X

$$\text{Likelihood: } \mathrm{L}(\beta, \mathrm{G_T}) = \prod_{i=1}^{m} \int \prod_{j=1}^{n_i} f(Y_{ij} \mid b,\ X_{ij})\ dG_T(b)$$

Misspecified random effects distributions

G_T typically unknown & we might assume $b \sim G_F \neq G_T$. Often $G_F = N(0, \sigma_b^2)$ (e.g. Stata default)

Two general forms of misspecified G_T:

- Incorrect distributional shape
- Incorrectly assuming b indep of X

1. $E(b \mid X) = \mu_M(X)$, Neuhaus & McCulloch (2006)
2. $var(b \mid X) = \mu_V(X)$, Heagerty & Kurland (2001)

Worry about correctly specifying the shape of G_T?

- Some say yes - Motivation for more flexible G_F, e.g. mixture of normals (Chen et al 2002), nonparametric G_F (Lesperance & Kalbfleisch 1992) & specification tests (Tchetgen & Coull 2006)
- Others show little bias in “slopes", $\hat{\beta}_1$ (Neuhaus et al 1992)
- Little work on random effects prediction under misspecification

Bias with misspecified shape of G?

Linear mixed effects model: $Y_{ij} \mid b_i = \beta_0 + \sigma_b\, b_i + \beta_1 X_{ij} + e_{ij}$

$b \perp e$, $b \sim G_T$, $E(b) = 0$, $var(b) = 1$, $var(e) = \sigma_e^2$

$cov(Y_{i1}, \ldots, Y_{in_i}) = V = \sigma_e^2 I + \sigma_b^2 J$, indep of G_T

$E\{\hat{\beta}_{GLS}\} = E\{(X^T V^{-1} X)^{-1} X^T V^{-1} Y\} = \beta_1$, indep of G_F

Unbiased estimators of slopes, β_1, with misspecified shape

Assessing consequences of random effects misspecification

Follow theory on inference with misspecified models (e.g. White 1994) - let $\xi = (\beta, \theta)$

$$\text{True} : f_T(Y_i \mid X_i;\ \xi) \;=\; \int \prod_{j=1}^{n_i} f(Y_{ij} \mid b,\ X_{ij}; \beta)\, dG_T(b; \theta)$$

$$\text{Fitted} : f_F(Y_i \mid X_i;\ \xi^*) \;=\; \int \prod_{j=1}^{n_i} f(Y_{ij} \mid b,\ X_{ij}; \beta^*)\, dG_F(b; \theta^*)$$

"MLE" $\hat{\xi}^* \rightarrow \xi^*$ minimizes $E_X E_{Y|X}\ log\ \{f_T(y|X;\xi)/f_F(y|X;\xi^*)\}$
Kullback-Leibler Divergence

Misspecified models

Minimizing K-L Divergence wrt $\xi^* = (\xi_1^*, \ldots, \xi_q^*)$ yields

$$E_X \int_y \lambda(y|X, \xi^*) \frac{\partial}{\partial \xi_k{}^*} \{ \int \prod_{j=1}^{n_i} f(Y_{ij} \mid b,\ X_{ij}; \beta^*)\ dG_F(b; \theta^*)\} dy = 0$$

where $\lambda(y|X, \xi^*) = f_T(Y = y|X, \xi) / f_F(Y = y|X, \xi^*)$

Note: ξ^* that yield $\lambda(y|X, \xi^*) = 1\ \forall X$ solve system

Can solve for ξ^* analytically in some simple cases, i.e. match fitted & true densities at all points X

Matched pairs, solutions under misspecification

- For some link functions can find analytic solution with $\beta_1^* = \beta_1$, but $\beta_0^* \neq \beta_0$ & $\sigma^* \neq \sigma$
- Special cases: $\hat{\beta}_1^*$ consistent, but $\hat{\beta}_0^*$ & $\hat{\sigma}^*$ inconsistent when $G_F \neq G_T$

Logistic link

- Binary matched pairs - when G_F generates a wide range of $pr(y_1, y_2)$ (e.g. $G_F =$ Normal) $\hat{\beta}_1^*$ is consistent (& $\hat{\beta}_1^* \equiv \hat{\beta}_{CML}$) (Neuhaus et al 1994)
- G_F unspecified, $\hat{\beta}_1^* \equiv \hat{\beta}_{CML}$ (Lindsay et al 1991)
- General X - when true $\beta_1 = 0$, $\beta_1^* = 0$ solves Kullback-Leibler minimizing equations $\Rightarrow \hat{\beta}_1^*$ consistent

Typically cannot solve Kullback-Leibler system analytically

Numerical solution of Kullback-Leibler equations

- Numerically find ξ^* that minimizes $E_X E_{Y|X}\, log\ \{f_T(y|\xi, X)/f_F(y|\xi^*, X)\}$ using routines in R
- Numerically evaluate integrals
- Numerically minimize Kullback-Leibler divergence
- Allows evaluation of bias over wide range of parameter values, random effect & covariate distributions

Numerical assessment of bias

True: $logit\ \{pr(Y_{ij} = 1|X_{ij}, b_i)\} = \beta_0 + \sigma b_i + \beta_1 X_{ij}$

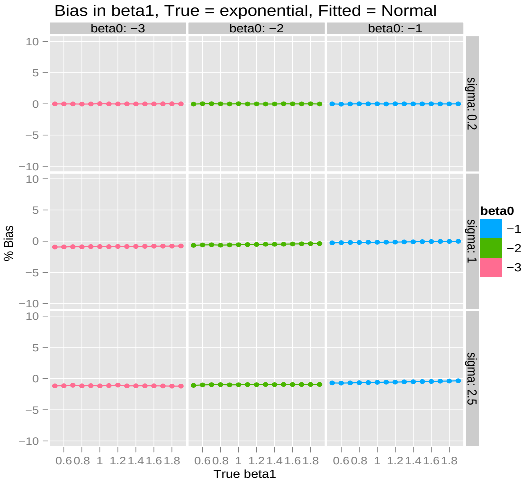
True G_T: $b \sim exp(1)$, shifted to $E(b) = 0$

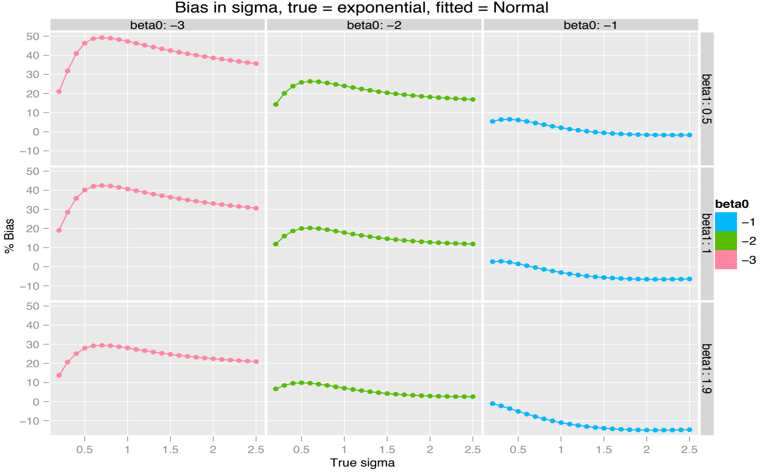
Fitted G_F: $b \sim N(0, 1)$

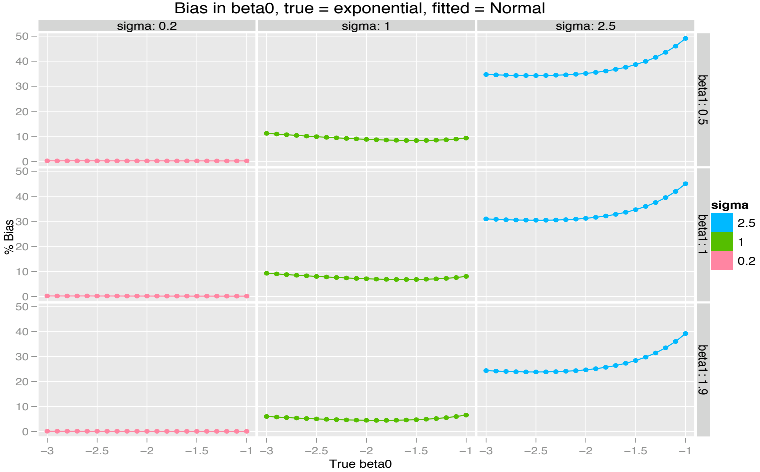
X varied within clusters $X = (0, .25, .5, .75, 1)$, $n_i = 5$

Range of parameter values:
β_0: (-3,-1) by 0.1 β_1: (0.5,2) by 0.1 σ: (0.2,2.5) by 0.1

Numerically solved for β_0^*, β_1^*, σ^* that minimize Kullback-Leibler divergence


Bias in beta1, True = exponential, Fitted = Normal
beta0: −3
beta0: −2
beta0: −1
sigma: 0.2
sigma: 1
sigma: 2.5
% Bias
10
5
0
−5
−10
True beta1
0.6 0.8 1 1.2 1.4 1.6 1.8
beta0
−1
−2
−3


Bias in sigma, true = exponential, fitted = Normal
beta0: −3
beta0: −2
beta0: −1
beta1: 0.5
beta1: 1
beta1: 1.9
% Bias
True sigma
beta0
−1
−2
−3


Bias in beta0, true = exponential, fitted = Normal
sigma: 0.2
sigma: 1
sigma: 2.5
beta1: 0.5
beta1: 1
beta1: 1.9
% Bias
True beta0
sigma
2.5
1
0.2

Logistic link - further results

- Find approximate solution of Kullback-Leibler equations using Taylor series methods (Neuhaus et al 1992)
- Approximate solution & simulations indicate $E(\hat{\beta}_1) \approx \beta_1$ when G_T misspecified, but biased $\hat{\beta}_0$ & $\hat{\sigma}_b$

Antibiotic Rx of acute respiratory infections

- Study objective: assess intervention to reduce inappropriate antibiotic prescription rates - Rx for antibiotic-nonresponsive conditions (e.g. acute bronchitis)
- Cluster: visits to a provider for antibiotic-nonresponsive conditions - number of visits varied from 1 to 71
- Design: Baseline Y $\rightarrow$ intervention $\rightarrow$ Post-Y
- Binary outcome: prescribed antibiotics for nonresponsive conditions (yes/no) - measured at each visit
- Covariates: intervention, time, time*intervention, provider type, illness duration prior to visit

Antibiotic Rx study - Fitted models

- logit $pr(Y_{ij} = 1 | b, X_{ij}) = \beta_0 + exp(log\ \sigma_b)\ b + \beta_1 X_{ij}$

  where $b \sim G$ with $E(b) = 0$, $Var(b) = 1$
- G_1 = N(0,1)
- G_2 = Exponential (1) standardized to $E(b) = 0$
- G_3 = Nonparametric, 4 support points (GLLAMM)

Antibiotic Rx study - Results

Estimated parameters from mixed-effects logistic models with different random effects distributions

$\hat{\beta}$, (SE)

G	β_0	TREAT	TIME	TRT*TIME	*log* σ_b
Normal	-1.06	0.66	0.56	-0.52	0.09
	(0.53)	(0.32)	(0.32)	(0.20)	(0.08)
Exponential	-0.92	0.72	0.55	-0.53	0.14
	(0.52)	(0.31)	(0.31)	(0.19)	(0.09)
NP (4)	-1.01	0.66	0.59	-0.55	0.02
	(0.51)	(0.32)	(0.31)	(0.19)	

Summary for fixed effects estimation

1. Misspecifying shape of G_T yields accurate "slopes" $\hat{\beta}_1$ over wide range of β_0, σ_b^2
2. Misspecifying shape of G_T can yield biased estimates of β_0 & σ_b^2
   - %bias in $\hat{\beta}_0$ $\uparrow$ with σ_b^2 & can be substantial
   - % bias in $\hat{\sigma}$ can also be large
3. When interest focuses on slope parameters, misspecifying shape has little effect on bias

Prediction of random effects

- Useful method: Predict b by $\tilde{b}$ that minimize

$$E[\tilde{b} - b]^2 = \int \int (\tilde{b} - b)^2 \, f(b, y) \, dy \, db,$$

  where $f(b, y)$ is the joint density of b & $y_1, \ldots, y_n$
- Can show: $\tilde{b}_i = E(b_i | y_{i1}, \ldots, y_{in_i})$
- Depends on $f(b, y)$, hence on G
- Misspecifying G may produce inaccurate $\tilde{b}$

Linear Mixed-Effects Model

$$Y_{ij} = \beta_0 + b_i + \beta_1 X_{ij} + \epsilon_{ij}, \quad i = 1, \ldots, m; \; j = 1, \ldots, n_i$$

$$b_i \sim N(0, \sigma_b^2), \; \epsilon_{ij} \sim N(0, \sigma_\epsilon^2)$$

Estimated BLUP: Estimated $\tilde{b} = \hat{b}_i = \hat{D}_i \; Z_i^T \; \hat{V}_i^{-1} \; (y_i - X_i \hat{\beta})$

$$\hat{D}_i = \hat{Cov}(b_i) \; \hat{V}_i = \hat{var}(y_i)$$

Expression depends on joint normality of b & y

Misspecification of distribution of b may produce inaccurate $\tilde{b}$

Theory for Linear Mixed-Effects Model (simple case)

$$Y_{ij} = \mu + b_i + \epsilon_{ij}, \quad i = 1, \ldots, m; \; j = 1, \ldots, n_i$$

$b_i \sim N(0, \sigma_b^2), \; \epsilon_{ij} \sim N(0, \sigma_\epsilon^2), \epsilon_{ij} \perp b_i, \; \mu, \; \sigma_b^2, \; \sigma_\epsilon^2$ known

Best Linear Unbiased Predictor (BLUP) is $\tilde{b}_i$ that minimizes $E[(\tilde{b}_i - b_i)^2]$

Simple calculations show that

$$\tilde{b}_i = E[b_i|y] = \frac{\sigma_b^2}{\sigma_b^2 + \sigma_\epsilon^2/n_i}(\bar{Y}_{i\cdot} - \mu) = \frac{\sigma_b^2}{\sigma_b^2 + \sigma_\epsilon^2/n_i}(b_i + \bar{\epsilon}_{i\cdot})$$

Theory for LME (cont.)

Conditional on b_i, the Y_{ij} are independent $N(\mu + b_i, \sigma_\epsilon^2)$

$$\tilde{b}_i \mid b_i \sim N\left\{\mu_{\tilde{b}} = \frac{\sigma_b^2}{\sigma_b^2 + \sigma_\epsilon^2/n_i}\, b_i,\ \left(\frac{\sigma_b^2}{\sigma_b^2 + \sigma_\epsilon^2/n_i}\right)^2 \frac{\sigma_\epsilon^2}{n_i}\right\}$$

Thus, $\tilde{b}_i$ is conditionally biased

However, since calculations are $|b_i$, result does not depend on distribution of b_i

i.e. conditional bias does not depend on distribution of b_i

Features of $\tilde{b}_i$

As $n_i \to \infty$,

$$\tilde{b}_i = \frac{\sigma_b^2}{\sigma_b^2 + \sigma_\epsilon^2 / n_i} (b_i + \bar{\epsilon}_{i.}) \to \frac{\sigma_b^2}{\sigma_b^2} (b_i + 0) = b_i$$

$\Rightarrow \tilde{b}_i$ converges to true value as $n_i \to \infty$

Such asymptotics typically not of interest, rather as $m \to \infty$

Density of $\tilde{b}_i$

What does the density of $\tilde{b}_i$ look like?

Also, what if $\tilde{b}_i$ misspecified to be normal?

If n_i large, then each $\tilde{b}_i \approx b_i \Rightarrow$ density is approximately correct, irrespective of assumed density

What if n_i not large, usual case of interest?

Then density of $\tilde{b}_i$ is convolution of true density of b_i with the conditional density of $\tilde{b}_i$ given b_i

Density of $\tilde{b}_i$

What does the density of $\tilde{b}_i$ look like?

Also, what if $\tilde{b}_i$ misspecified to be normal?

If n_i large, then each $\tilde{b}_i \approx b_i \Rightarrow$ density is approximately correct, irrespective of assumed density

What if n_i not large, usual case of interest?

Then density of $\tilde{b}_i$ is convolution of true density of b_i with the conditional density of $\tilde{b}_i$ given b_i

Density of $\tilde{b}_i$

What does the density of $\tilde{b}_i$ look like?

Also, what if $\tilde{b}_i$ misspecified to be normal?

If n_i large, then each $\tilde{b}_i \approx b_i \Rightarrow$ density is approximately correct, irrespective of assumed density

What if n_i not large, usual case of interest?

Then density of $\tilde{b}_i$ is convolution of true density of b_i with the conditional density of $\tilde{b}_i$ given b_i

Density of $\tilde{b}_i$

What does the density of $\tilde{b}_i$ look like?

Also, what if $\tilde{b}_i$ misspecified to be normal?

If n_i large, then each $\tilde{b}_i \approx b_i \Rightarrow$ density is approximately correct, irrespective of assumed density

What if n_i not large, usual case of interest?

Then density of $\tilde{b}_i$ is convolution of true density of b_i with the conditional density of $\tilde{b}_i$ given b_i

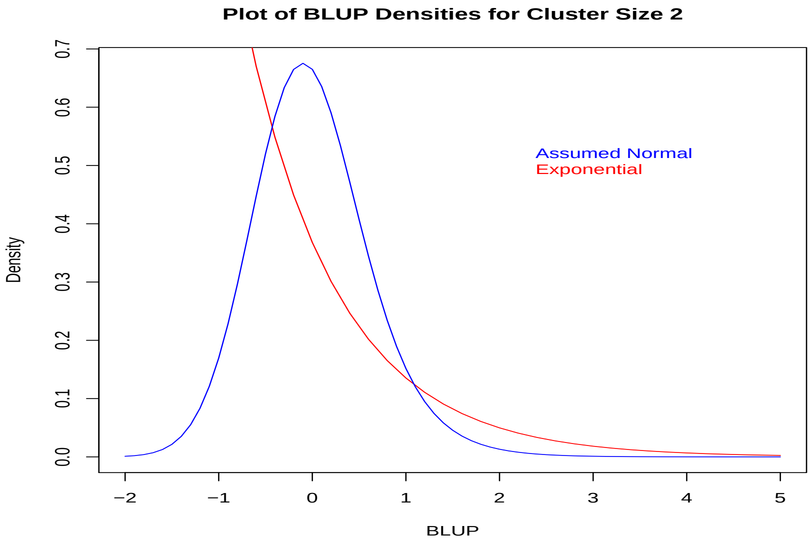
Density of $\tilde{b}_i$ (cont)

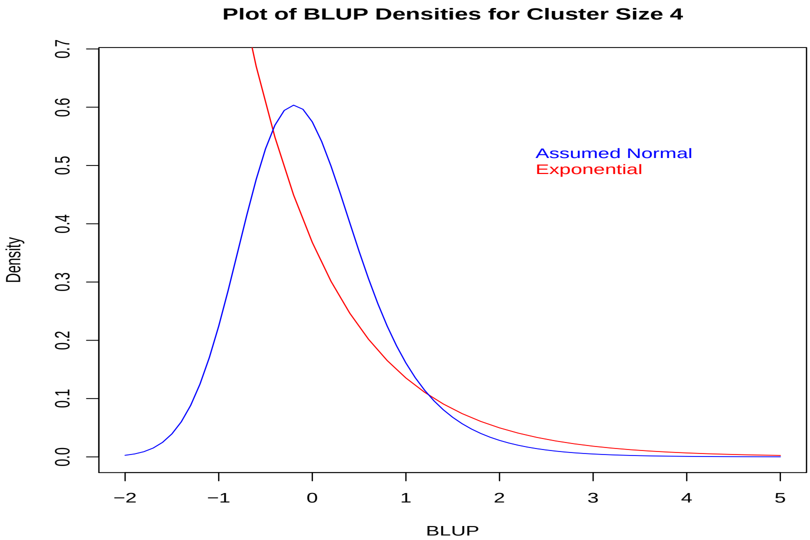
Suppose $b_i \sim$ Exponential(1), shifted so that $E(b_i) = 0$

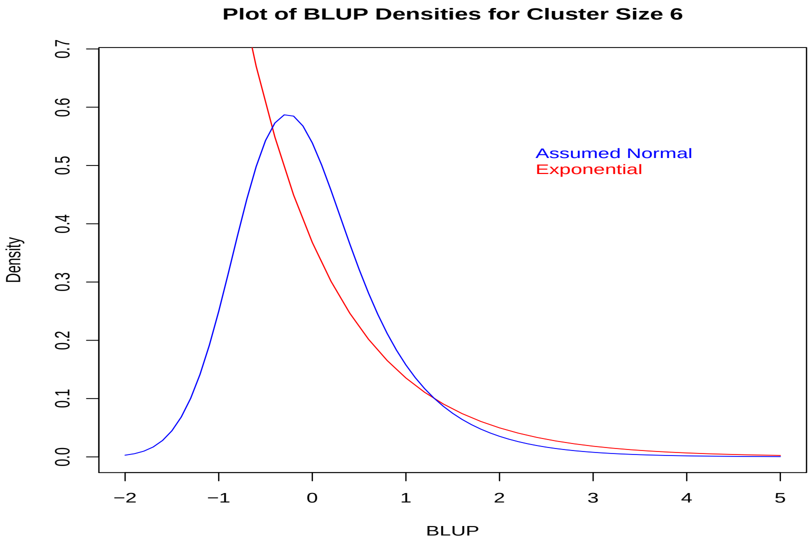
Density of $\tilde{b}_i$ (under normal assumption) is

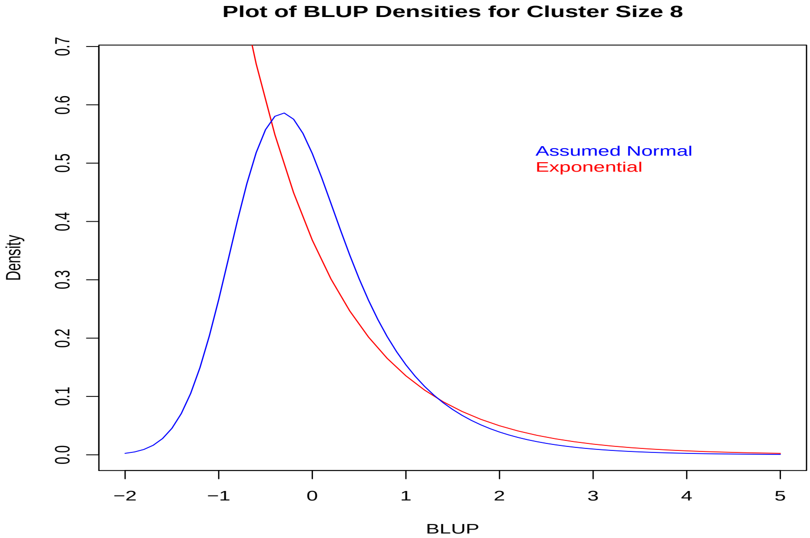
$$\int_0^\infty exp\{-(\tilde{b} - \mu_{\tilde{b}})^2 n_i/(2\sigma_\epsilon^2)\}\ exp(-\tilde{b} - 1)\ d\tilde{b}$$

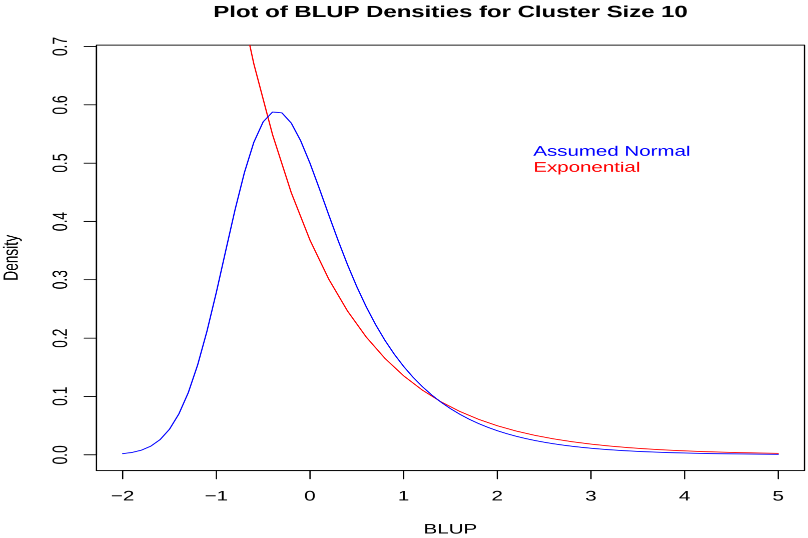
Straightforward to evaluate numerically

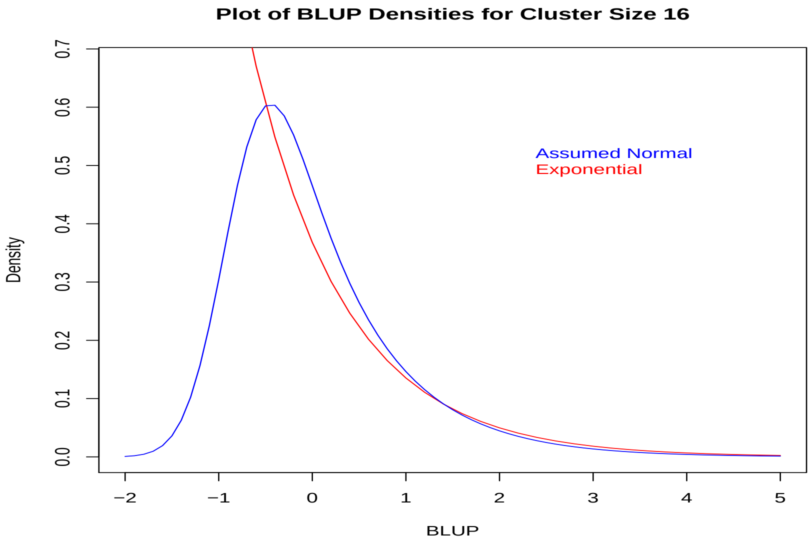

Plot of BLUP Densities for Cluster Size 2
Density
0.0
0.1
0.2
0.3
0.4
0.5
0.6
0.7
−2
−1
0
1
2
3
4
5
BLUP
Assumed Normal
Exponential

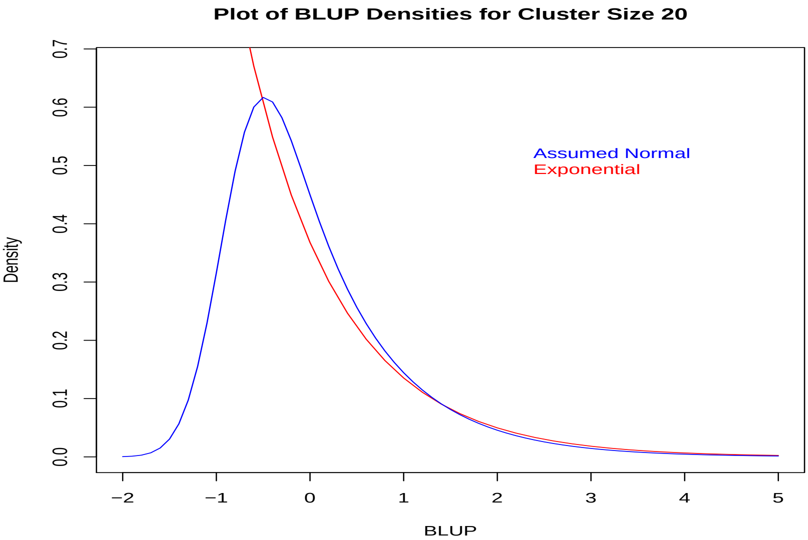

Plot of BLUP Densities for Cluster Size 4
Density
0.0
0.1
0.2
0.3
0.4
0.5
0.6
0.7
−2
−1
0
1
2
3
4
5
BLUP
Assumed Normal
Exponential


Plot of BLUP Densities for Cluster Size 6
Density
0.0
0.1
0.2
0.3
0.4
0.5
0.6
0.7
−2
−1
0
1
2
3
4
5
BLUP
Assumed Normal
Exponential


Plot of BLUP Densities for Cluster Size 8
Density
0.0
0.1
0.2
0.3
0.4
0.5
0.6
0.7
−2
−1
0
1
2
3
4
5
BLUP
Assumed Normal
Exponential


Plot of BLUP Densities for Cluster Size 10
Density
0.0
0.1
0.2
0.3
0.4
0.5
0.6
0.7
−2
−1
0
1
2
3
4
5
BLUP
Assumed Normal
Exponential


Plot of BLUP Densities for Cluster Size 16
Density
0.0
0.1
0.2
0.3
0.4
0.5
0.6
0.7
−2
−1
0
1
2
3
4
5
BLUP
Assumed Normal
Exponential


Plot of BLUP Densities for Cluster Size 20
Density
0.0
0.1
0.2
0.3
0.4
0.5
0.6
0.7
−2
−1
0
1
2
3
4
5
BLUP
Assumed Normal
Exponential

BLUP density plot findings

- Density of BLUPs inherits much of its shape from assumed density
- Doesn't reflect shape of true density of the random effects until cluster sizes get large

Best Predicted values

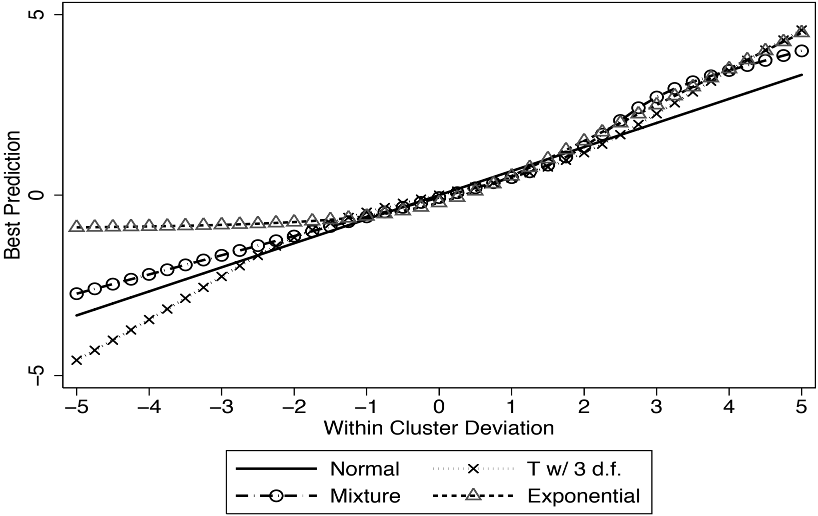
Under LME, with $e_{ij} \sim N(0, \sigma_e^2)$, & true density $f_{b_i}(b_i)$,

$$\tilde{b}_i = \mathrm{E}[b_i|\mathbf{Y}_i] = \frac{\int_{-\infty}^{\infty} b_i \exp\{-\frac{n_i}{2\sigma_\epsilon^2}(\nu_i - b_i)^2\} f_{b_i}(b_i) db_i}{\int_{-\infty}^{\infty} \exp\{-\frac{n_i}{2\sigma_\epsilon^2}(\nu_i - b_i)^2\} f_{b_i}(b_i) db_i}$$

where $\nu_i = (\bar{Y}_{i.} - \bar{\mathbf{x}}_{i.}'\beta)$

Given ν_i, compute $\tilde{b}_i$ for various true $f_{b_i}(b_i)$: Normal, T_3, Exponential (1), Mixture of Normals

BP plots


Best Prediction
5
0
−5
−5
−4
−3
−2
−1
0
1
2
3
4
5
Within Cluster Deviation
Normal
Mixture
T w/ 3 d.f.
Exponential

Order preservation for LME

Can show $\frac{\partial \tilde{b}_i}{\partial \nu_i} > 0$ for any $f_{b_i}(b_i) \Rightarrow$ transformation $\nu_i \rightarrow \tilde{b}_i$ monotone

Given n_i, BP's under any assumed f ordered based on ν_i

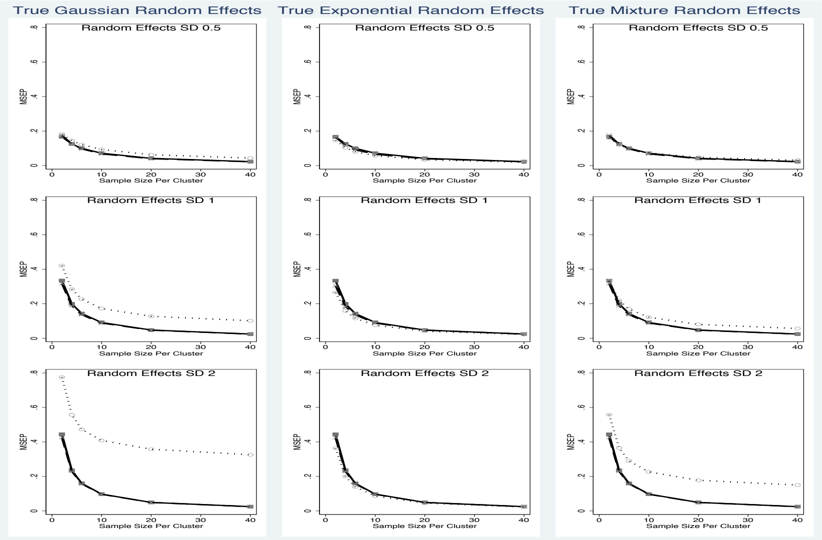
- Let $f_1(b_i)$, $f_2(b_i)$ denote assumed random effects densities
- $\tilde{b}_{i1}(\nu_i)$ & $\tilde{b}_{i2}(\nu_i)$, predictions from each value ν_i
- Order the pairs $[\tilde{b}_{i1}(\nu_i), \tilde{b}_{i2}(\nu_i)]$ by ν_i
- Since $\frac{\partial \tilde{b}_i}{\partial \nu_i} > 0$, pairs also ordered by $\tilde{b}_{i1}(\nu_i)$ & $\tilde{b}_{i2}(\nu_i)$.
- Thus, all pairs concordant & Kendall's τ between $\tilde{b}_{i1}(\nu_i)$ and $\tilde{b}_{i2}(\nu_i)$ is 1

Mean square error of prediction

Given $\sigma_b^2,\ \sigma_e^2$, we calculated $\tilde{b}$ assuming

- $b\ \sim$N(0,1)
- $b \sim$ Exponential
- $b \sim$ Mixture of two N(0,1)

Using expressions for $\tilde{b}$ we calculated MSE $=$ E$[(\tilde{b} - b)^2]$ for various true $f_b(b)$ using numerical integration



Assumed distributions: Solid line/square=Gaussian, dotted line/circle=Exponential, dashed line/X=Mixture

Simulations

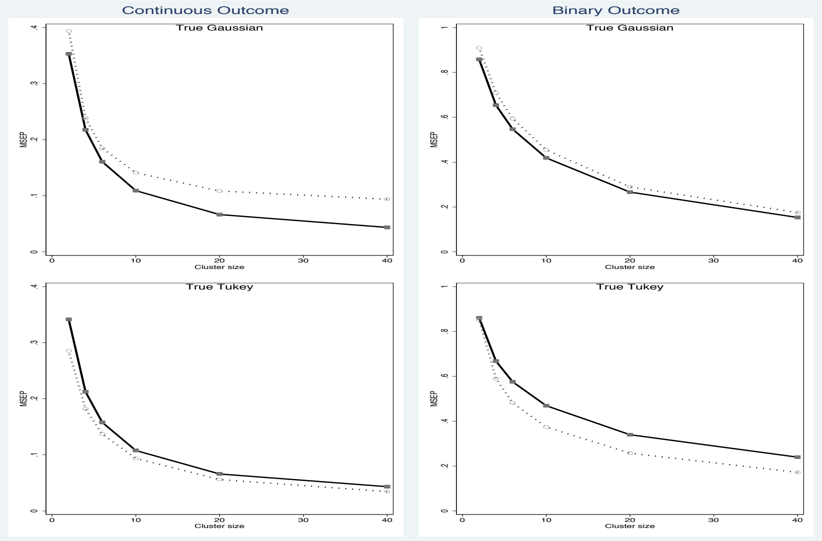
Evaluate performance of BP's when all parameters estimated

Two true & assumed $f(b)$, i.e. 4 true/assumed settings

- $b \sim \mathrm{N}(0, 1)$
- $b \sim \mathrm{Tukey}(g = 0.5, h = 0.1)$
- Linear mixed effects & mixed effects logistic models
- Linear predictor: $\eta_{ij} = -2 + b_i + x_{between} + x_{within}$
- $\sigma_b^2 = 1$, $\sigma_e^2 = 1$ for LME
- Calculated $M\hat{S}E = (1/mK) \sum_{k=1}^{K} \sum_{i=1}^{m} (\tilde{b}_{ki} - b_{ki})^2$
- $K = 1000$, $m = 100$, cluster sizes=2, 4, 6, 10, 20, 40

Percentiles of N(0,1) & standardized Tukey(0.5,0.1)

Percentile	N(0,1)	Standardized Tukey(0.5,0,1)
.1%	-3.09	-1.89
1%	-2.33	-1.40
2.5%	-1.96	-1.21
5%	-1.64	-1.06
10%	-1.28	-0.89
50%	0	-0.21
90%	1.28	1.09
95%	1.64	1.73
97.5%	1.96	2.47
99%	2.33	3.62
99.9%	3.09	7.68



Assumed distributions: Solid line/square=Gaussian, dotted line/circle=Tukey

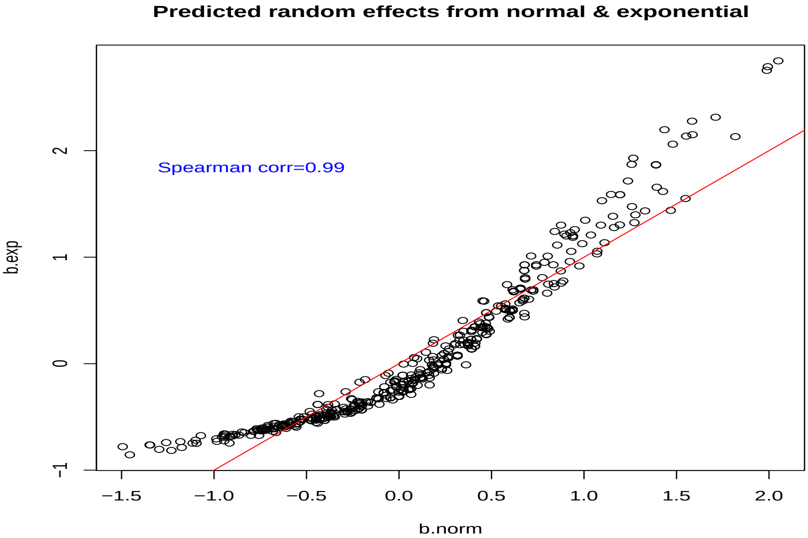
Antibiotic Rx study - Fitted models

Objective: identify poorly performing providers (i.e. large predicted random effects)

- $\text{logit}\, pr(Y_{ij} = 1|b, X_{ij}) = \beta_0 + exp(log\, \sigma_b)\, b + \beta_1 X_{ij}$
- where $b \sim G$ with $E(b) = 0,\ Var(b) = 1$
- G_1 = N(0,1)
- G_2 = Exponential (1) standardized to $E(b) = 0$
- Compute $\tilde{b}$ that maximizes

$$log\{f(y_{i1}, \ldots, y_{in_i} \mid X_{i1}, \ldots, X_{in_i}, \beta, b_i)\; g(b_i \mid \theta)\}$$

- Fit models using PROC NLMIXED in SAS


Predicted random effects from normal & exponential
Spearman corr=0.99
b.exp
b.norm
−1
0
1
2
−1.5
−1.0
−0.5
0.0
0.5
1.0
1.5
2.0

Summary of effects on prediction

1. Predicted values of random effects show modest sensitivity to assumed distributional shape.
2. Distribution shape of BLUPs often not reflective of true random effects distribution.
3. Ranking of predicted values is little affected.
4. Misspecified shape can produce modest increases in MSE of prediction.

Summary

Misspecifying the shape of the random effects produces

- little bias in estimated covariate effects
- slight deterioration in random effects prediction.

Stata default of $b \sim \mathrm{N}(0, \sigma_b^2)$ yields accurate estimates of covariate effects & reasonably precise predicted random effects