

# **Prediction of Random Effects and Effects of Misspecification of Their Distribution**

Charles E. McCulloch and John Neuhaus  
Division of Biostatistics,  
Department of Epidemiology and Biostatistics  
University of California, San Francisco

West Coast Stata Users Group Meeting, October 2007

# Outline

- 1) Introduction and motivating examples
- 2) Prediction of random effects
  - a) Are parametric assumptions important?
- 3) Brief review of effects of misspecification (more generally) in mixed models
- 4) Theoretical calculations (Linear Mixed Model)
- 5) Theoretical calculations (Binary Matched Pairs)
- 6) Simulations (Linear Mixed Model)
- 7) Example: Hormone and Estrogen Replacement Study
- 8) Summary

# 1. Introduction: Examples

Example 1: Measuring cognitive decline in elderly women (Women Who Maintain Optimal Cognitive Function into Old Age. Barnes DE, Cauley JA, Lui L-Y, Fink HA, McCulloch CE, Stone KL, Yaffe K. *J Amer Geriatrics Soc*, 2007). A modified Mini-Mental status examination was given at baseline and years 6, 8, 10 and 15 in a prospective cohort study (Study of Osteoporotic Fractures). Which participants are thought to be in mental decline and what predicts that decline?

Example 2: Effect of pre-hypertension at an early age in the CARDIA study. (Prehypertension During Young Adulthood and Presence of Coronary Calcium Later in Life: The Coronary Artery Risk Development In Young Adults (CARDIA) Study. MJ Pletcher, K Bibbins-Domingo, CE Lewis, G Wei, S Sidney, JJ Carr, E Vittinghoff, CE McCulloch, SB Hulley, submitted). Blood pressure measured every five years since 1986. How to approximate previous and cumulative blood pressure exposure?

Example 3: Predicting those at risk for developing high blood pressure in HERS (The Heart and Estrogen Replacement Study - Hulley, et al, J. American Medical Association, 1998). HERS was a randomized, blinded, placebo controlled trial for women with previous coronary disease. We will use it as a prospective cohort study for prediction of high blood pressure. 2,763 women were enrolled and followed yearly for 5 subsequent visits. We will consider only the subset that were not diabetic and with systolic blood pressure less than 140 at the beginning of the study.

## 2. Mixed models and prediction of random effects

One way to address the questions above is to utilize mixed models and derive predicted values of the random effects.

Example 1: (cognitive decline):

$$\begin{aligned} MMSE_{it} &= \text{cognitive measure for participant } i \text{ at time } t \\ &= b_{0i} + b_{1i}t + \text{covariates} + \varepsilon_{it}, \end{aligned}$$

$$\begin{pmatrix} b_{0i} \\ b_{1i} \end{pmatrix} \sim \text{indep. } N \left( \begin{pmatrix} \beta_0 \\ \beta_1 \end{pmatrix}, \begin{pmatrix} \sigma_{00} & \sigma_{01} \\ \sigma_{01} & \sigma_{11} \end{pmatrix} \right)$$

calculate  $\tilde{b}_{1i}$  = predicted decline for participant  $i$ .

Some realistic but made up data:

```
. table visit, c(mean mmse n mmse sd mmse)
```

---

visit	mean(mmse)	N(mmse)	sd(mmse)
0	27.08	2,031	2.2
1	27.17	1,931	2.3
2	27.10	1,850	2.3
3	27.08	1,750	2.3
4	27.04	1,361	2.3
5	27.10	269	2.2

---

So little change in average MMSE over time.

```
xi: xtmixed mmse visit exercise avgdrpwk || pptid: visit, cov(uns)
```

```
Performing EM optimization:
```

```
Performing gradient-based optimization:
```

```
Iteration 0:    log restricted-likelihood = -11662.158  
Iteration 1:    log restricted-likelihood =  -11662.14  
Iteration 2:    log restricted-likelihood =  -11662.14
```

```
Computing standard errors:
```

```
Mixed-effects REML regression      Number of obs      =      9110  
Group variable: pptid              Number of groups   =      2032  
  
Obs per group: min =              1  
                  avg =             4.5  
                  max =              6  
  
Wald chi2(3)                       =      27.24  
Log restr-likelihood = -11662.14    Prob > chi2        =      0.0000
```



mmse	Coef.	Std. Err.	z	P> z	[95% Conf. Interval]	
visit	-.0060353	.0059123	-1.02	0.307	-.0176231	.0055526
exercise	.0773954	.0179999	4.30	0.000	.0421162	.1126746
avgdrpwk	-.0097331	.0037005	-2.63	0.009	-.0169859	-.0024803
_cons	27.11017	.0495455	547.18	0.000	27.01307	27.20728

Random-effects Parameters	Estimate	Std.Err.	[95% Conf.Interval]	
pptid: Unstructured				
sd(visit)	.1942305	.005721	.183333	.2057752
sd(_cons)	2.158639	.034978	2.09115	2.228296
corr(visit,_cons)	-.0426722	.031013	-.103225	.0181959
sd(Residual)	.481975	.004743	.472767	.4913616

LR test vs. lin regression:  $\chi^2(3) = 17033.5$  Prob >  $\chi^2 = 0.000$

```

. predict rslopedev rintdev, reffects
. gen predslope=_b[visit]+rslopedev
. collapse rslopedev rintdev predslope, by(pptid)
. gen deltammse=6*predslope

```

```

. summarize

```

Variable	Obs	Mean	Std. Dev.	Min	Max
pptid	2032	1394.65	794.41	1	2761
rslopedev	2032	1.3e-10	.1421	-.7615	.9239
rintdev	2032	3.9e-10	2.133	-9.8275	3.0384
predslope	2032	-.006035	.1421	-.7676	.9178
deltammse	2032	-.036211	.8530	-4.6057	5.5071

```

. summarize deltammse predslope if deltammse<-2

```

Variable	Obs	Mean	Std. Dev.	Min	Max
deltammse	40	-2.634	.4937	-4.6057	-2.0331
predslope	40	-.4390	.0822	-.7676	-.3388

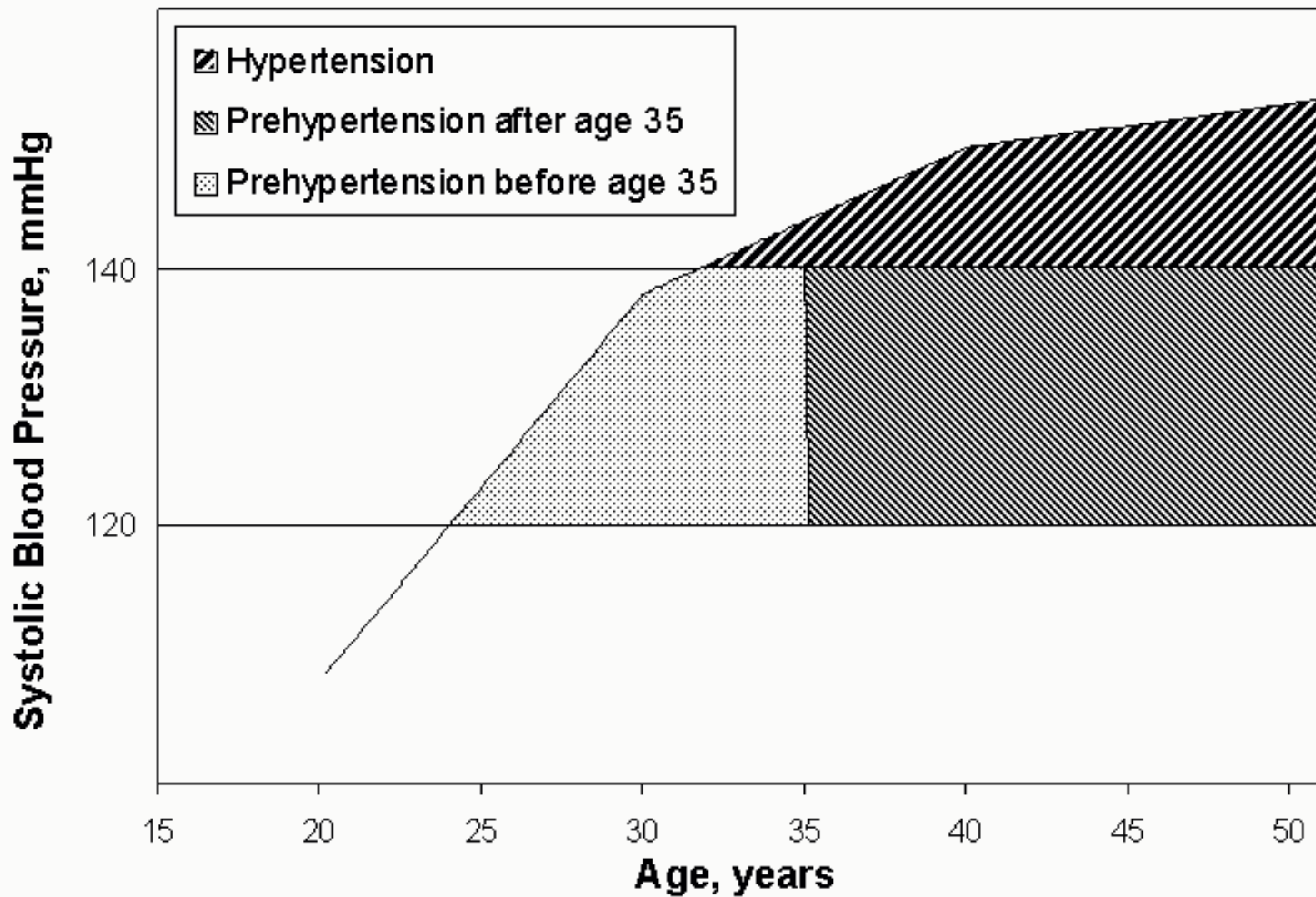
Example 2: (pre-hypertension):

$$\begin{aligned}BP_{it} &= \text{blood pressure for participant } i \text{ at time } t \\ &= \text{spline}_i(t) + \text{covariates} + \varepsilon_{it},\end{aligned}$$

(*spline terms*)  $\sim$  indep.  $N(\boldsymbol{\mu}, \boldsymbol{\Sigma})$

calculate predicted spline for participant  $i$ .

The area under the predicted blood pressure trajectory between 120 and 140 mmHg was integrated over time as a cumulative pre-hypertension exposure (in years of mmHg). This was then used as a predictor of coronary calcification.



Example 3: (high blood pressure):

$BP_{it} = 1$  if blood pressure is high for subject  $i$  at time  $t$ , and 0 o/w

$$\text{logit}(P\{BP_{it} = 1\}) = b_{0i} + \text{covariates} + \varepsilon_{it},$$

$$b_{0i} \sim \text{i.i.d. } N(\beta_0, \sigma_b^2)$$

calculate  $\tilde{b}_{0i}$  = predicted intercept for participant  $i$ .

Predicted values of random effects available from `gllamm` or the new (Ver 10) multilevel logit command `xtmelogit`

The `xtmixed`, `xtmelogit`, `xt-etc.` and `gllamm` commands fit the models using regular or restricted maximum likelihood. So they use a parametric assumption for both the distribution of the outcome and the distribution of the random effects, the latter typically that the distributions are normal.

**Key question: Is the parametric assumption of the random effects distribution important?**

This is especially crucial since we don't get to directly observe the random effects. Unfortunately, the predicted random effects may not reflect the shape of the underlying distribution. (More on this point later).

### 3. Review of impact of misspecification in mixed models

A number of investigations have focused on the effect of misspecifications in parametric mixed models. They can be grouped as:

1. Getting the distributional shape wrong.
2. Falsely assuming the random effect is independent of the cluster size.
3. Falsely assuming the random effect is independent of covariates, e.g.,
  - a. Mean of random effects distribution could be associated with a covariate.
  - b. Variance of random effects distribution could be associated with a covariate.

Most investigations have concentrated on the impact on estimation of the fixed effects portion of the model.

General assessment:

1) Getting the distributional shape wrong has little impact on inferences about the fixed effects.

2) Incorrectly assuming the random effects distribution is independent of the cluster size may affect inferences about the intercept, but does not seriously impact inferences about the regression parameters.

3) However, assuming the random effects distribution is independent of the covariates when it is not is potentially serious. (Related to mean: Neuhaus and McCulloch, JRSSB, 2006; related to the variance: Heagerty and Kurland, Biometrika, 2001).



## **What about inference about the predictions of the random effects?**

We'll concentrate on the issue of wrong distributional shape, where fixed effects inferences seem largely unaffected.

Intuition: the assumed form of the random effects distribution may be a more crucial assumption in this case.

## 4. Theoretical calculations (Linear Mixed Model)

First consider an easy situation. Assumed model is a one-way random effects model with known intercept and variance components and normally distributed random effects:

$$Y_{it} = \mu + b_i + \varepsilon_{it}, t = 1, \dots, n_i; i = 1, \dots, q$$

$$b_i \sim i.i.d. N(0, \sigma_b^2)$$

$$\varepsilon_{it} \sim i.i.d. N(0, \sigma_\varepsilon^2)$$

$$\varepsilon_{it} \perp b_i, \mu, \sigma_\varepsilon^2, \text{ and } \sigma_b^2 \text{ known}$$

In which case the Best Linear Unbiased Predictor is given by

$$\tilde{b}_i = \frac{\sigma_b^2}{\sigma_b^2 + \sigma_\varepsilon^2 / n_i} (\bar{Y}_{i\cdot} - \mu)$$

Writing this out:

$$\begin{aligned}\tilde{b}_i &= \frac{\sigma_b^2}{\sigma_b^2 + \sigma_\varepsilon^2 / n_i} (\bar{Y}_{i.} - \mu) \\ &= \frac{\sigma_b^2}{\sigma_b^2 + \sigma_\varepsilon^2 / n_i} (\mu + b_i + \bar{\varepsilon}_{i.} - \mu) \\ &= \frac{\sigma_b^2}{\sigma_b^2 + \sigma_\varepsilon^2 / n_i} (b_i + \bar{\varepsilon}_{i.})\end{aligned}$$

Conditional on  $b_i$ , the  $Y_{it}$  are independent  $N(\mu + b_i, \sigma_\varepsilon^2)$ . So

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

and  $\tilde{b}_i$  is conditionally biased. Since the calculations are conditional on  $b_i$ , results do not depend on the distribution of the  $b_i$  and so the conditional bias does *not* depend on the distribution.

$$\tilde{b}_i = \frac{\sigma_b^2}{\sigma_b^2 + \sigma_\varepsilon^2 / n_i} (b_i + \bar{\varepsilon}_i.)$$
$$\rightarrow \frac{\sigma_b^2}{\sigma_b^2} (b_i + 0) = b_i \text{ as } n_i \rightarrow \infty$$

So  $\tilde{b}_i$  converges in probability to the true value as  $n_i \rightarrow \infty$ . But asymptotic calculations as  $n_i \rightarrow \infty$  are not usually of interest for a random effects model.

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

And what if the assumption of normality for the  $b_i$  is incorrect, i.e., not normal?

If  $n_i$  is large then each  $\tilde{b}_i$  is close to  $b_i$  and hence the distribution is approximately correct.

But what about the case when  $n_i$  is not large, the usual case of interest?

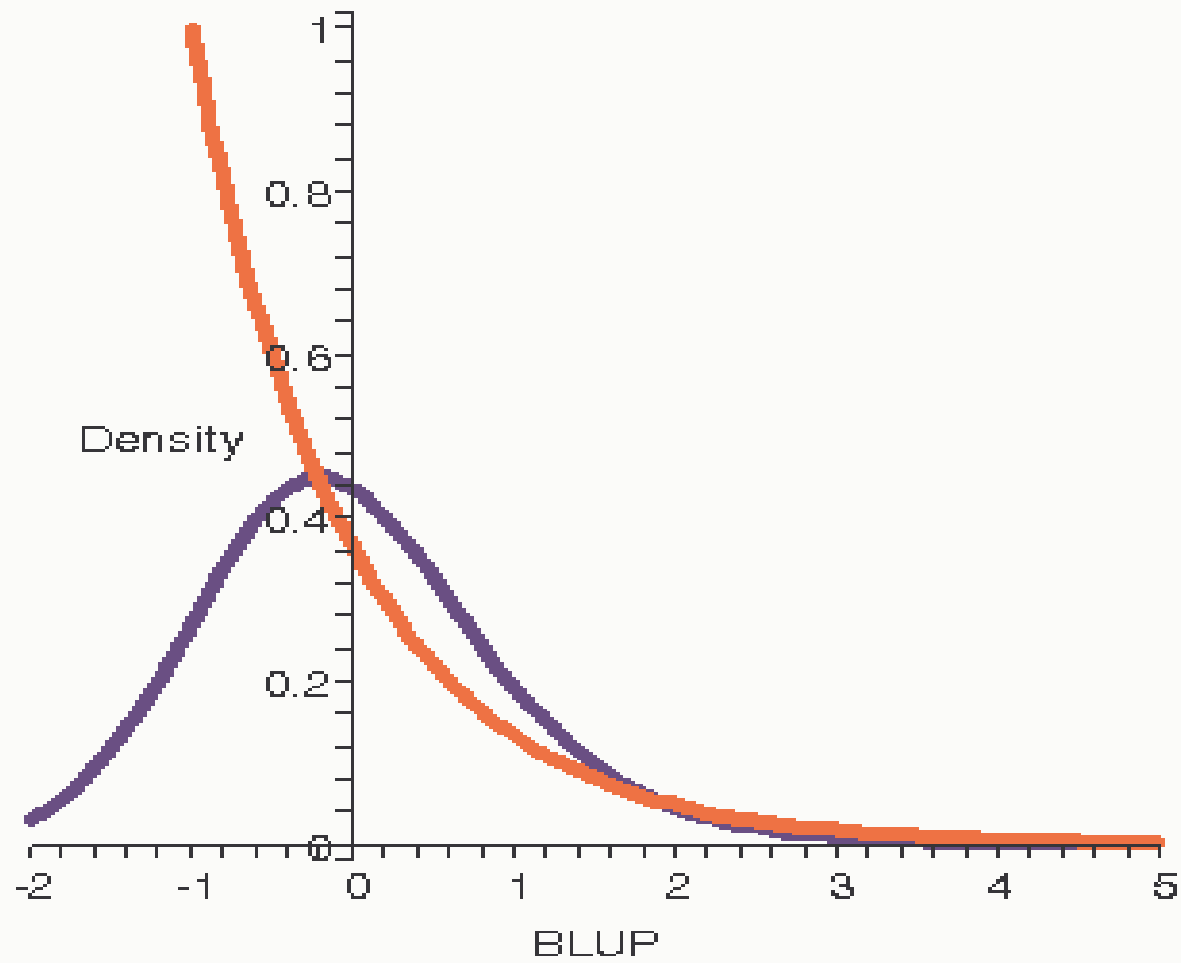
Then the distribution of  $\tilde{b}_i$  is the convolution of the true density with the conditional density of  $\tilde{b}_i$  given  $b_i$ .

For example, suppose the true density is exponential(1), shifted to have mean 0. Then the density of  $\tilde{b}_i$  is given by

$$\int_0^{\infty} \exp\left\{-\frac{(\tilde{b} - \mu_{\tilde{b}})^2 n_i}{2\sigma_{\varepsilon}^2}\right\} \exp(-\tilde{b} - 1) d\tilde{b},$$

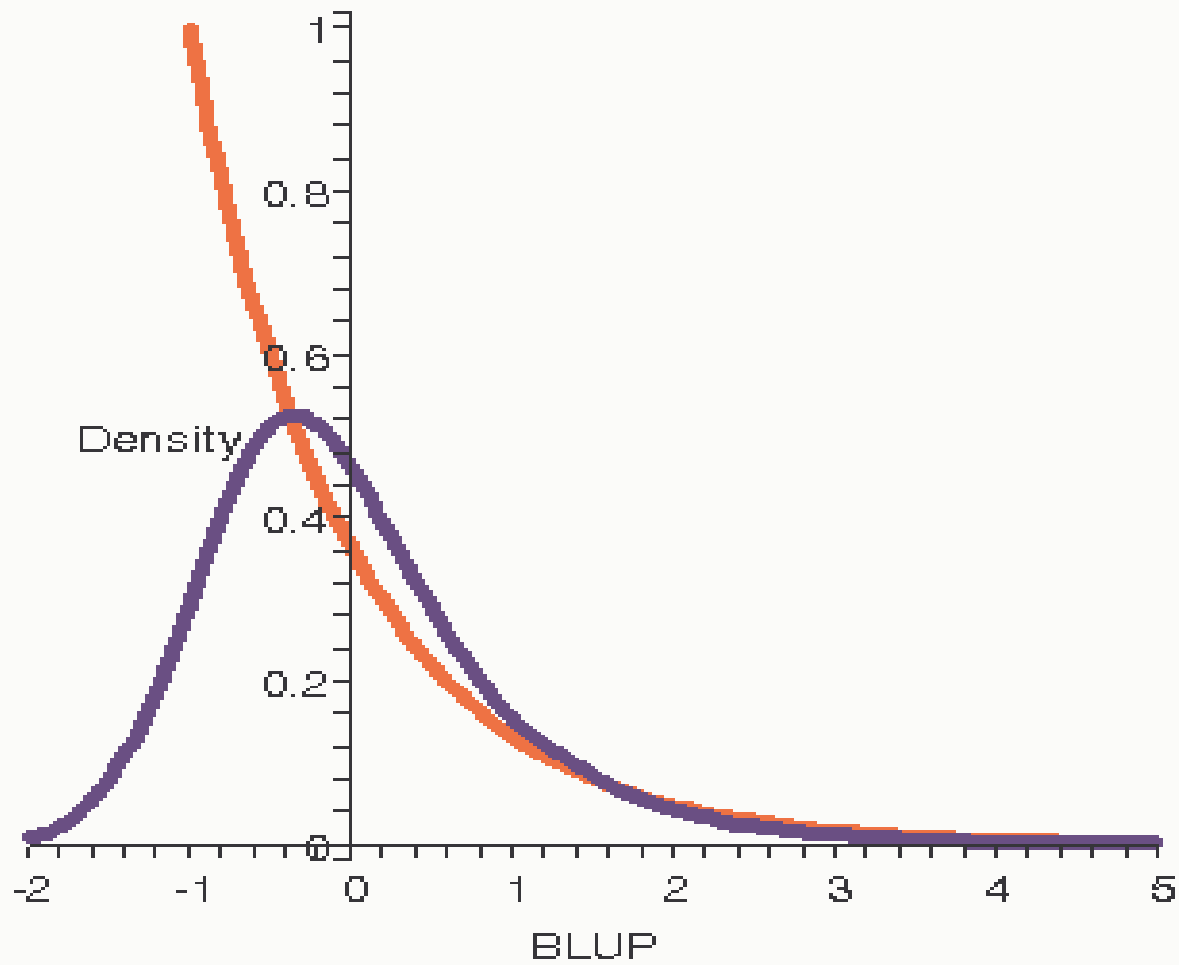
which is straightforward to evaluate numerically:

Plot of BLUP Densities for Cluster Size 2  
Assumed Normal - Blue, Exponential - Red

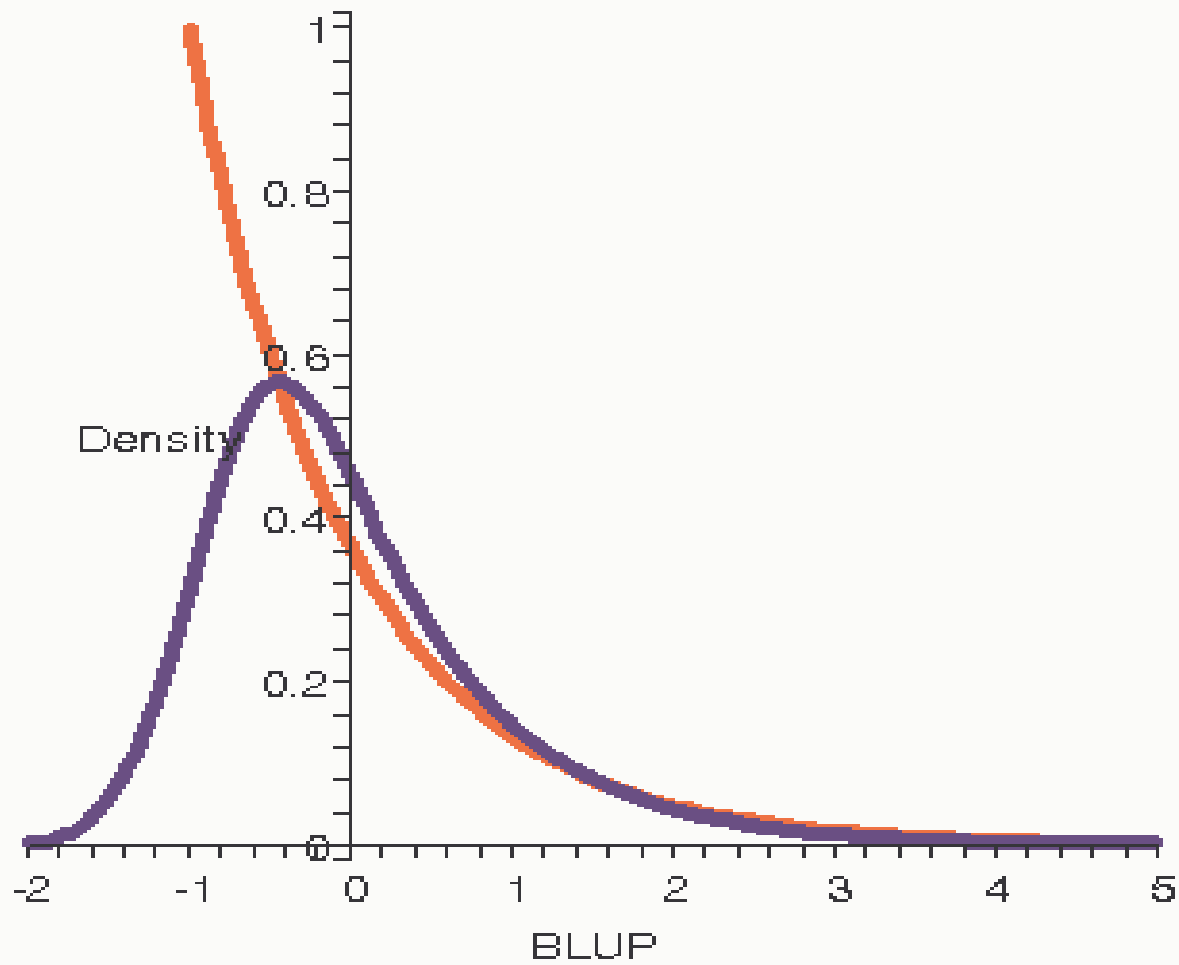




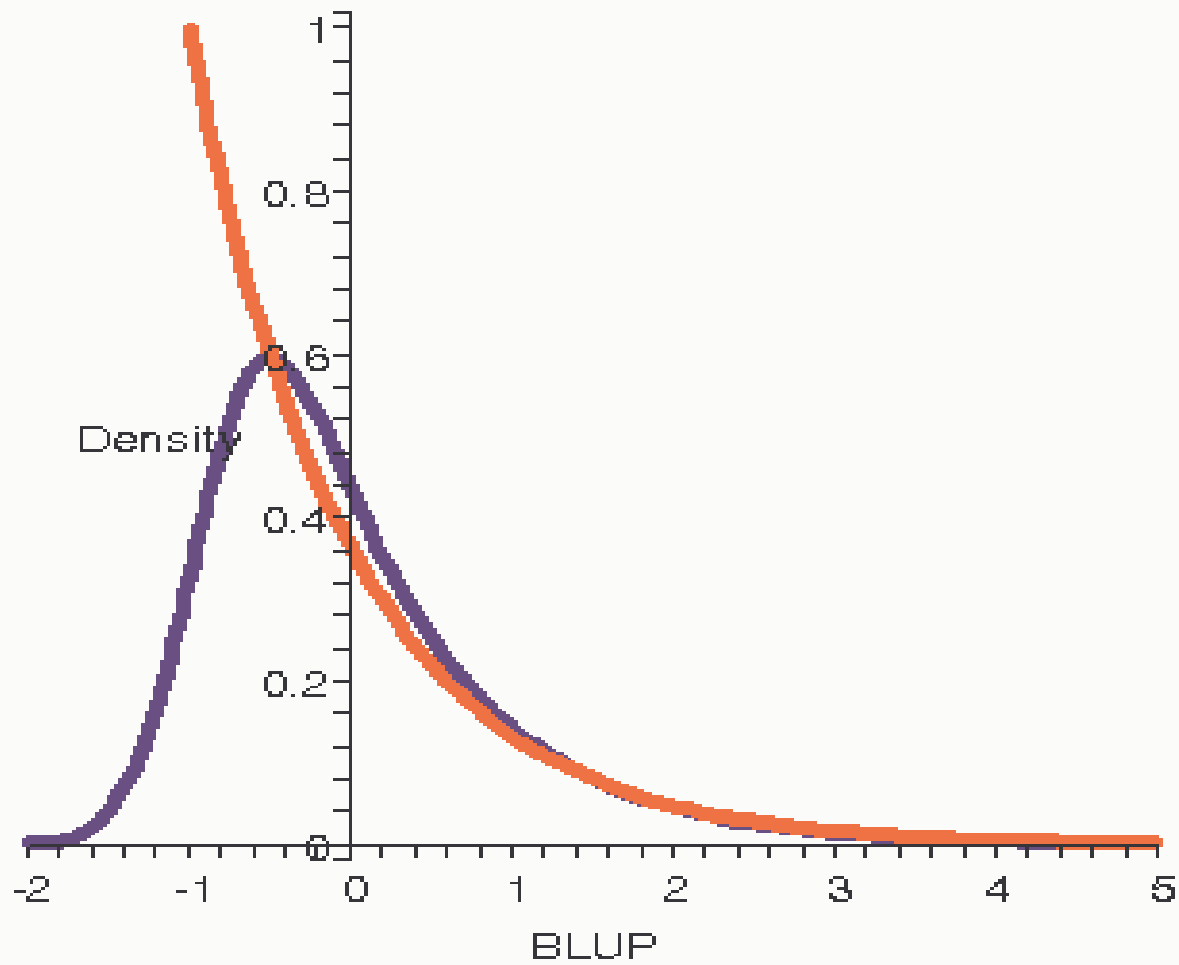
Plot of BLUP Densities for Cluster Size 4  
Assumed Normal - Blue, Exponential - Red



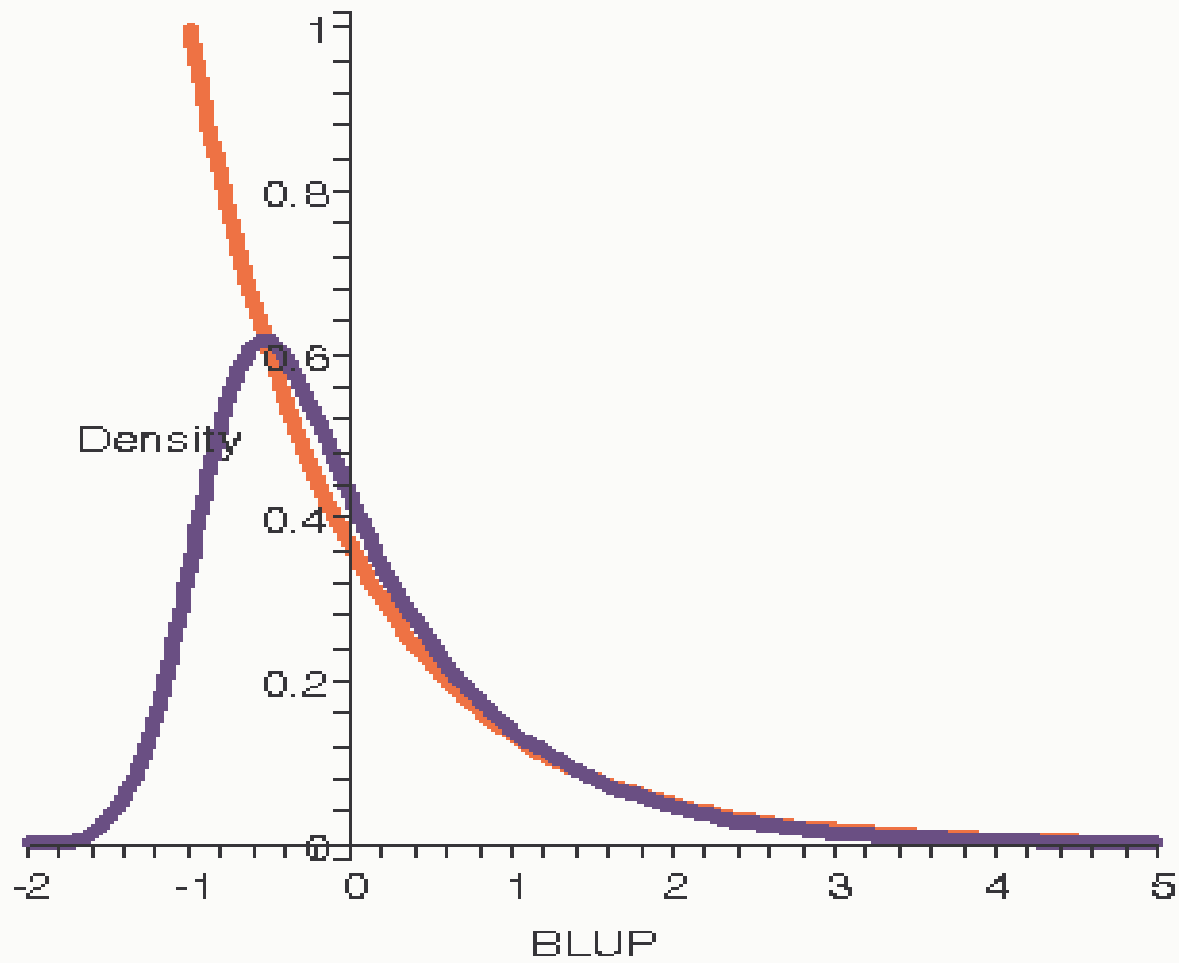
Plot of BLUP Densities for Cluster Size 6  
Assumed Normal - Blue, Exponential - Red



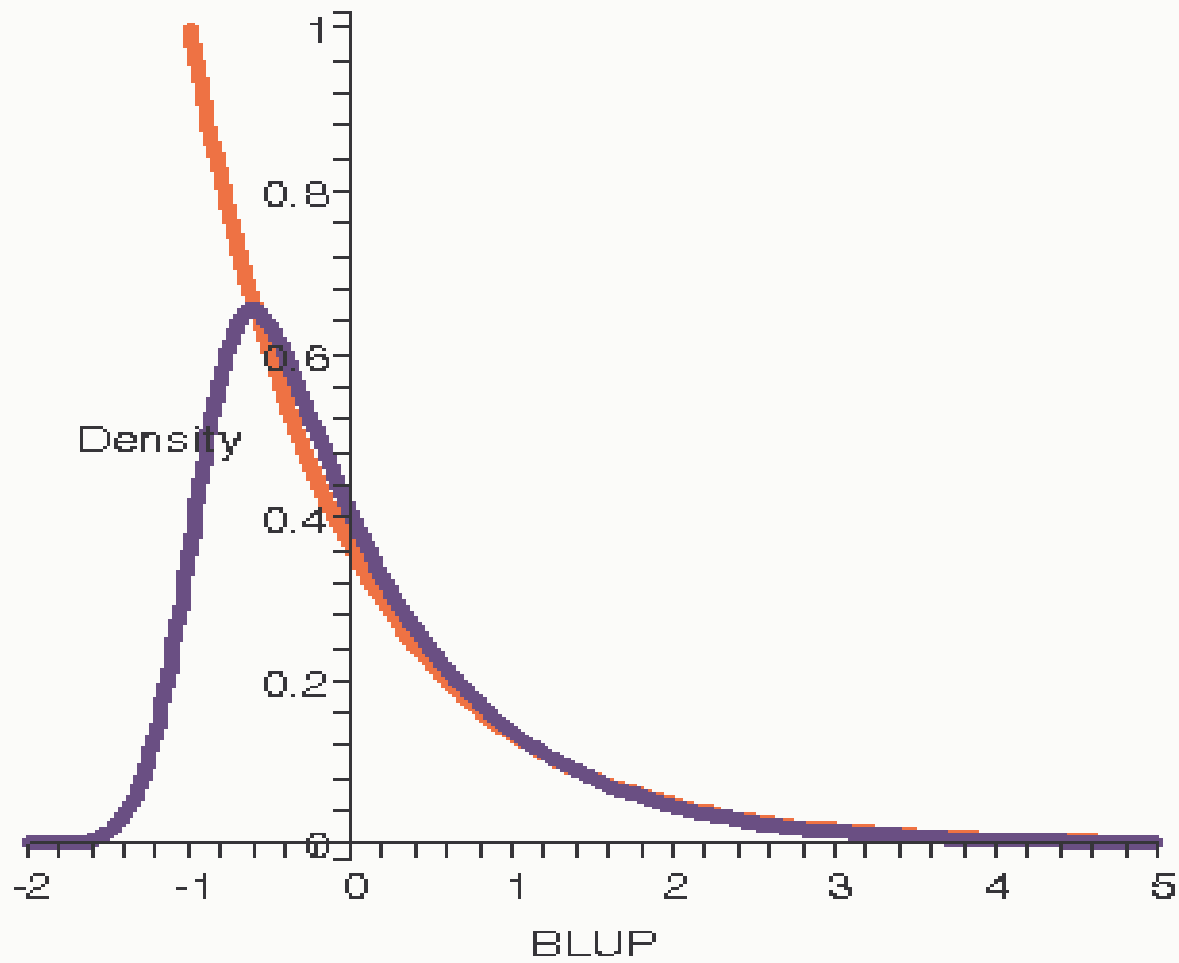
Plot of BLUP Densities for Cluster Size 8  
Assumed Normal - Blue, Exponential - Red



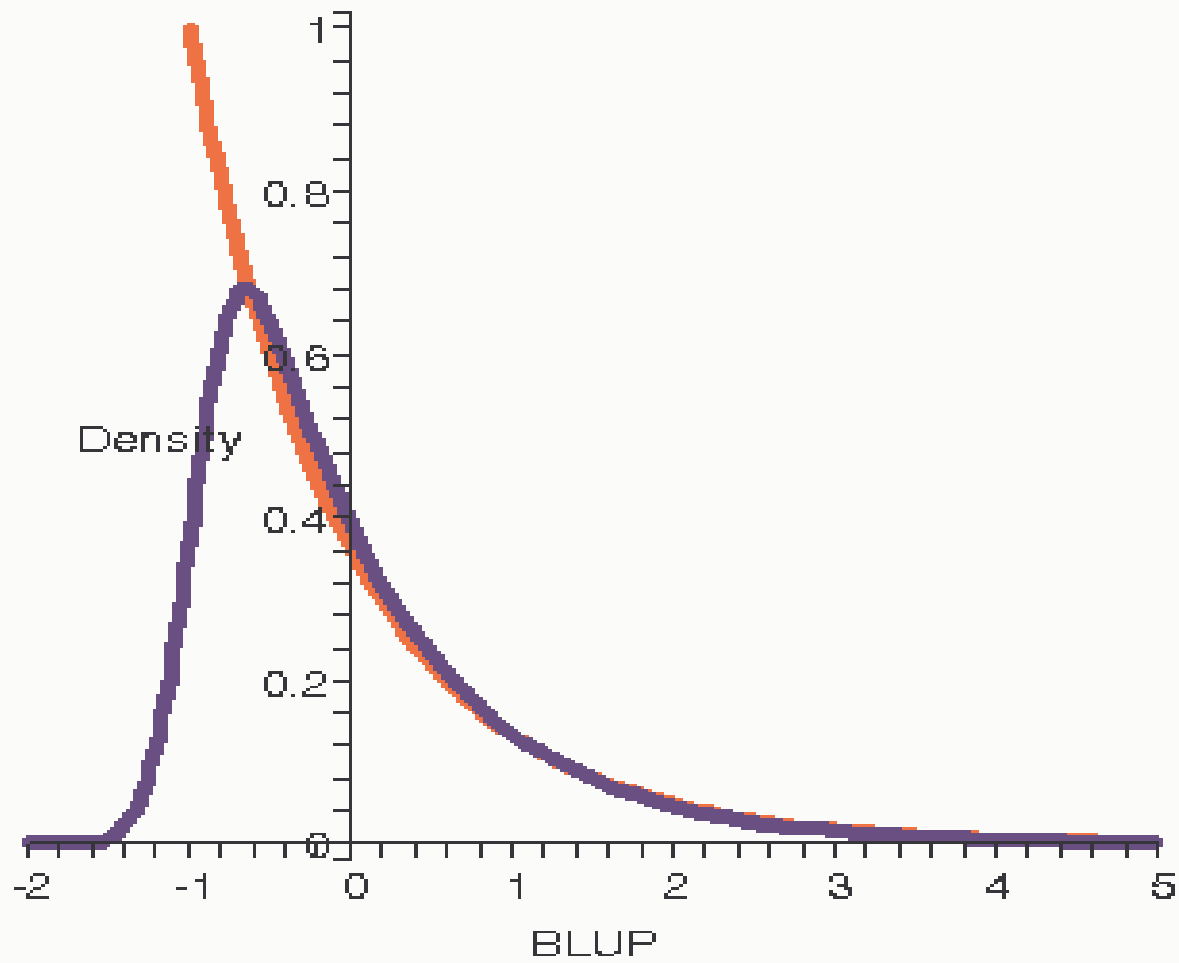
Plot of BLUP Densities for Cluster Size 10  
Assumed Normal - Blue, Exponential - Red



Plot of BLUP Densities for Cluster Size 15  
Assumed Normal - Blue, Exponential - Red



Plot of BLUP Densities for Cluster Size 20  
Assumed Normal - Blue, Exponential - Red



What is the BLUP under the exponential assumption?

Model:

$$Y_{it} = \mu + b_i + \varepsilon_{it}, t = 1, \dots, n_i; i = 1, \dots, q$$

$$b_i \sim \text{i.i.d. } \sigma_b (\mathcal{E}(1) - 1)$$

$$\varepsilon_{it} \sim \text{i.i.d. } N(0, \sigma_\varepsilon^2)$$

$$\varepsilon_{it} \perp b_i; \mu, \sigma_\varepsilon^2, \text{ and } \sigma_b^2 \text{ known}$$

$$\text{Define } \Delta = \frac{\sqrt{n}}{\sigma_\varepsilon} (\bar{Y} - \mu + \sigma_b) - \frac{\sigma_\varepsilon}{\sqrt{n}\sigma_b} .$$

Then

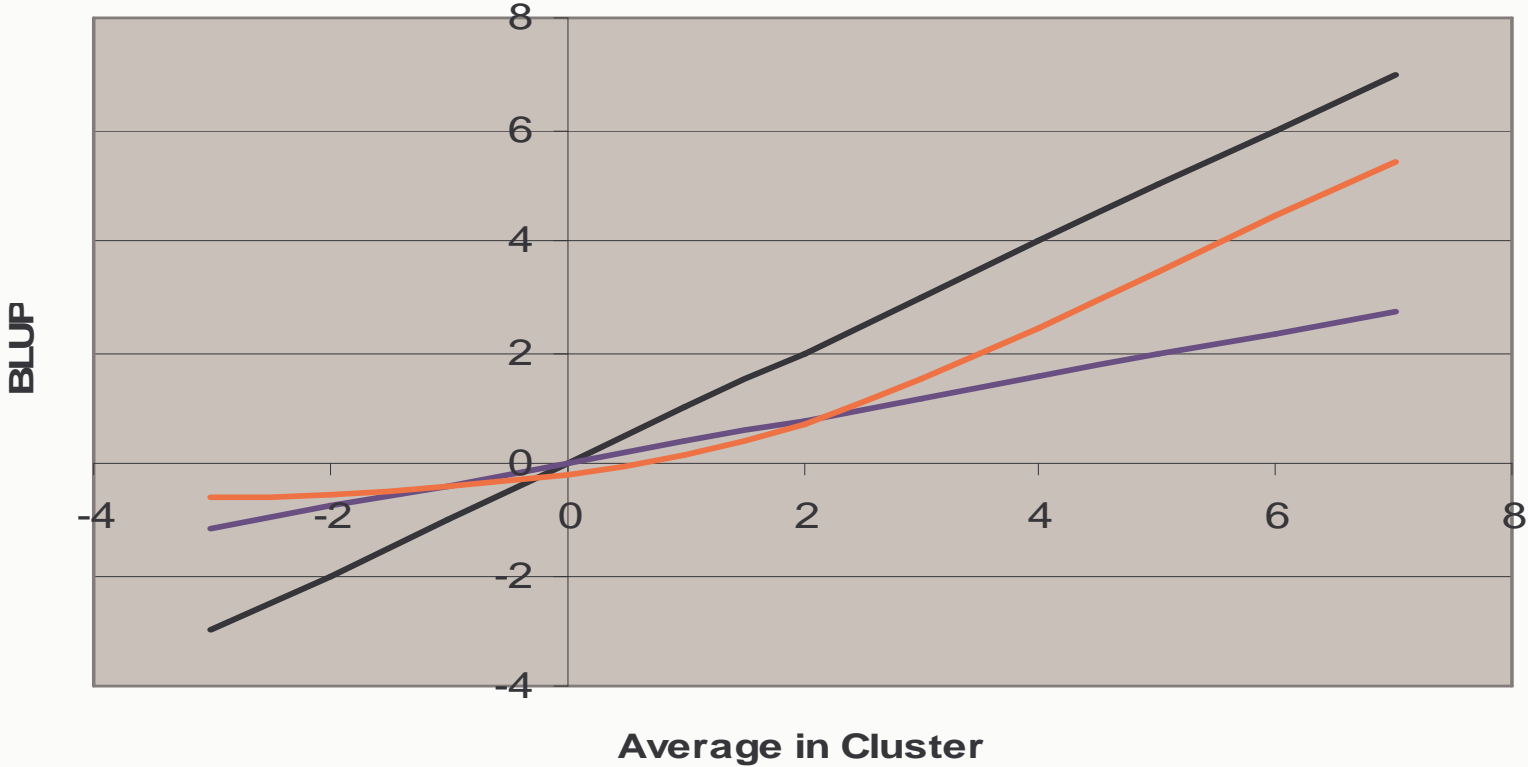
$$\tilde{b}_i = \bar{Y} - \mu - \frac{\sigma_\varepsilon^2}{n\sigma_b} + \frac{\phi(\Delta)\sigma_\varepsilon}{\Phi(\Delta)\sqrt{n}},$$

where  $\phi(t)$  and  $\Phi(t)$  are the standard normal p.d.f. and c.d.f.

How do the assumed normal and assumed exponential BLUPs compare?



# BLUPS Under Different Distributional Assumptions



— Raw deviation — Normal — Exponential

## 5. Theoretical calculations (Binary matched pairs)

Assumed model

$$Y_{it} | b_i \sim \text{Binomial}(p_{it}), i = 1, \dots, q; t = 1, 2$$

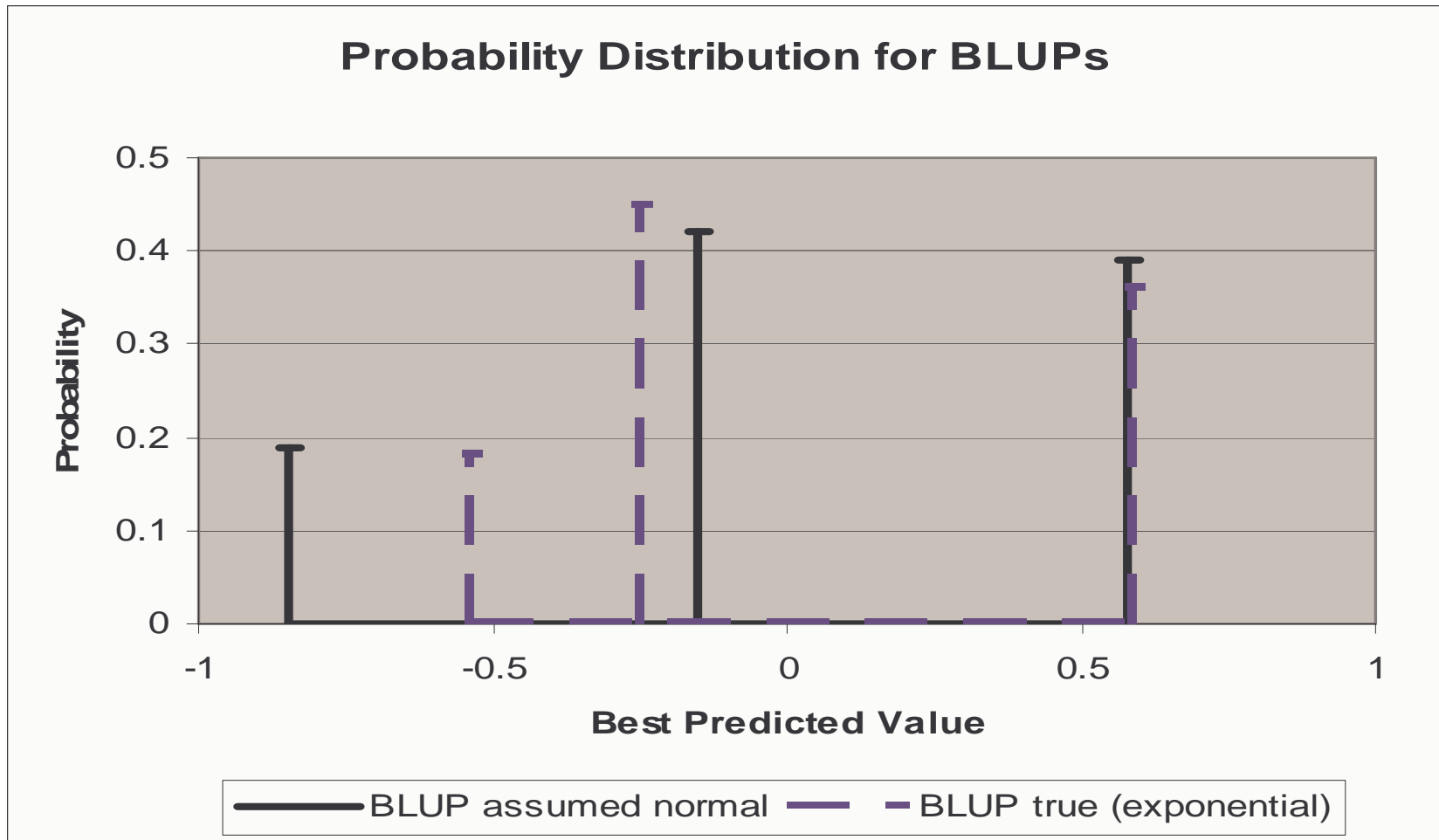
$$\text{logit}(p_{it}) = \mu + b_i + \beta I_{\{t=2\}}$$

$$b_i \sim \text{i.i.d. } N(\beta_0, \sigma_b^2)$$

Since there are only 4 data configurations per cluster there are only four possible values for  $\tilde{b}_i$ , for a given set of parameter values. For example, when  $y_{i1} = y_{i2} = 1$ ,  $\tilde{b}_i$  is given by (with  $p(t) = 1/(1 + e^{-t})$ )

$$\tilde{b}_i = \frac{\int_{-\infty}^{\infty} b \phi(b) p(\mu + \sigma_b b) p(\mu + \sigma_b b + \beta) db}{\int_{-\infty}^{\infty} \phi(b) p(\mu + \sigma_b b) p(\mu + \sigma_b b + \beta) db}$$

These depend on the assumed distribution. The probabilities of the four (actually three) values depends on the true distribution.



It is also straightforward to calculate the mean square error of prediction using the assumed and true models under the true model. For example, if the assumed model is normal, but the true is exponential here are some values of the mean square error of prediction:

Mean squared error of prediction  $MSEP = E[(\tilde{b}_i - b_i)^2]$  with  $\mu = 0, \sigma = 1$ :

$\beta$	Normal (assumed)	Exponential (true)	Percent increase
0	0.77	0.75	3.5%
1	0.82	0.79	3.0%
2	0.85	0.83	2.1%
3	0.87	0.85	1.4%

## 6. Simulation

We simulated data from the one-way random model:

$$Y_{it} = \mu + b_i + \varepsilon_{it}, t = 1, \dots, n_i; i = 1, \dots, q$$

$$b_i \sim \text{i.i.d. } N(0, \sigma_b^2) \text{ or } b_i \sim \text{i.i.d. } \sigma_b \{\mathcal{E}(1) - 1\}$$

$$\varepsilon_{it} \sim \text{i.i.d. } N(0, \sigma_\varepsilon^2), \varepsilon_{it} \perp b_i,$$

with  $q = 10 = n_i$  and using the same random numbers for both the normal and exponential random effects (and the same error terms). 10,000 replications. An assumed normal model was fit.

## Simulation results

### Estimates of the parameters

<u>Normal</u>	True	Ave	SD	Ave SE
$\mu$	1	1.00	0.33	0.32
$\ln(\sigma_{\varepsilon}^2)$	0	-0.01	0.075	0.075
$\ln(\sigma_b^2)$	0	-0.07	0.29*	0.27*
<u>Exponential</u>				
$\mu$	1	1.00	0.33	0.31
$\ln(\sigma_{\varepsilon}^2)$	0	-0.01	0.075	0.075
$\ln(\sigma_b^2)$	0	-0.18*	0.47	0.29

\*Excludes one outlier

## Estimates of fixed effects parameters are little affected.

Estimates of the parameters

<i>Normal</i>	True	Ave	SD	Ave SE
$\mu$	1	1.00	0.33	0.32
$\ln(\sigma_{\varepsilon}^2)$	0	-0.01	0.075	0.075
$\ln(\sigma_b^2)$	0	-0.07	0.29*	0.27*
<i>Exponential</i>				
$\mu$	1	1.00	0.33	0.31
$\ln(\sigma_{\varepsilon}^2)$	0	-0.01	0.075	0.075
$\ln(\sigma_b^2)$	0	-0.18*	0.47*	0.29*

\*Excludes one outlier

As is the estimate of log of the residual variance.

Estimates of the parameters

<i>Normal</i>	True	Ave	SD	Ave SE
$\mu$	1	1.00	0.33	0.32
$\ln(\sigma_{\varepsilon}^2)$	0	-0.01	0.075	0.075
$\ln(\sigma_b^2)$	0	-0.07	0.29*	0.27*
<i>Exponential</i>				
$\mu$	1	1.00	0.33	0.31
$\ln(\sigma_{\varepsilon}^2)$	0	-0.01	0.075	0.075
$\ln(\sigma_b^2)$	0	-0.18*	0.47*	0.29*

\*Excludes one outlier



**But the estimate of the random effects variance is off.**

Estimates of the parameters

<i>Normal</i>	True	Ave	SD	Ave SE
$\mu$	1	1.00	0.33	0.32
$\ln(\sigma_{\varepsilon}^2)$	0	-0.01	0.075	0.075
$\ln(\sigma_b^2)$	0	-0.07	0.29*	0.27*
<i>Exponential</i>				
$\mu$	1	1.00	0.33	0.31
$\ln(\sigma_{\varepsilon}^2)$	0	-0.01	0.075	0.075
$\ln(\sigma_b^2)$	0	-0.18*	0.47*	0.29*

\*Excludes one outlier

Confidence interval coverage for  $\mu$  was slightly lower than nominal for the normal (92%), and low for the exponential (88%).

Mean square error of prediction for the BLUPs was 1.87 for the normal model and 1.84 for the exponential.

Do the BLUPs calculated under the assumption of normality reflect the true underlying shape (exponential)?

For data simulated with normally distributed random effects the average skewness was -0.01 and the average kurtosis was 2.50 (with a normal having values 0 and 3).

For data simulated with exponentially distributed random effects the average skewness was 0.85 and the average kurtosis was 3.14 (with an exponential(1) having values 2 and 9).

## 7. Example (HERS)

Recall the HERS example: We will consider the 1,378 women who did not have high blood pressure and were not diabetic at the baseline visit. We will use the baseline and visits 1 through 3 to predict the blood pressure at visits 4 and 5 and whether or not the woman had developed high blood pressure on either visit 4 or 5.

Brief descriptive statistics:

<u>Variable</u>	<u>Mean/Percentage</u>	<u>SD</u>
Age	66.3	6.9
BMI	27.3	4.9
Weight	70.3 kg	13.4 kg
On BP meds	79%	

Predictive model (for baseline and visits 1, 2 and 3):

$$BP_{it} = \beta_0 + b_{0i} + \beta_1 BMI + \beta_2 EXER + \beta_3 AGE \\ + \beta_4 MEDS + \beta_5 DM + \varepsilon_{it},$$

$$b_{0i} \sim \text{i.i.d. } N(0, \sigma_b^2) \text{ or } b_{0i} \sim \text{i.i.d. } \sigma_b \{\mathcal{E}(1) - 1\}$$

calculate  $\tilde{BP}_{it} = \hat{\beta}_0 + \tilde{b}_{0i} + \hat{\beta}_1 BMI + \hat{\beta}_2 EXER + \hat{\beta}_3 AGE \\ + \hat{\beta}_4 MEDS + \hat{\beta}_5 DM$  (mixed model pred)

or  $\hat{BP}_{it} = \hat{\beta}_0 + \hat{\beta}_1 BMI + \hat{\beta}_2 EXER + \hat{\beta}_3 AGE \\ + \hat{\beta}_4 MEDS + \hat{\beta}_5 DM$  (fixed effects only)

How well do the predictions work?

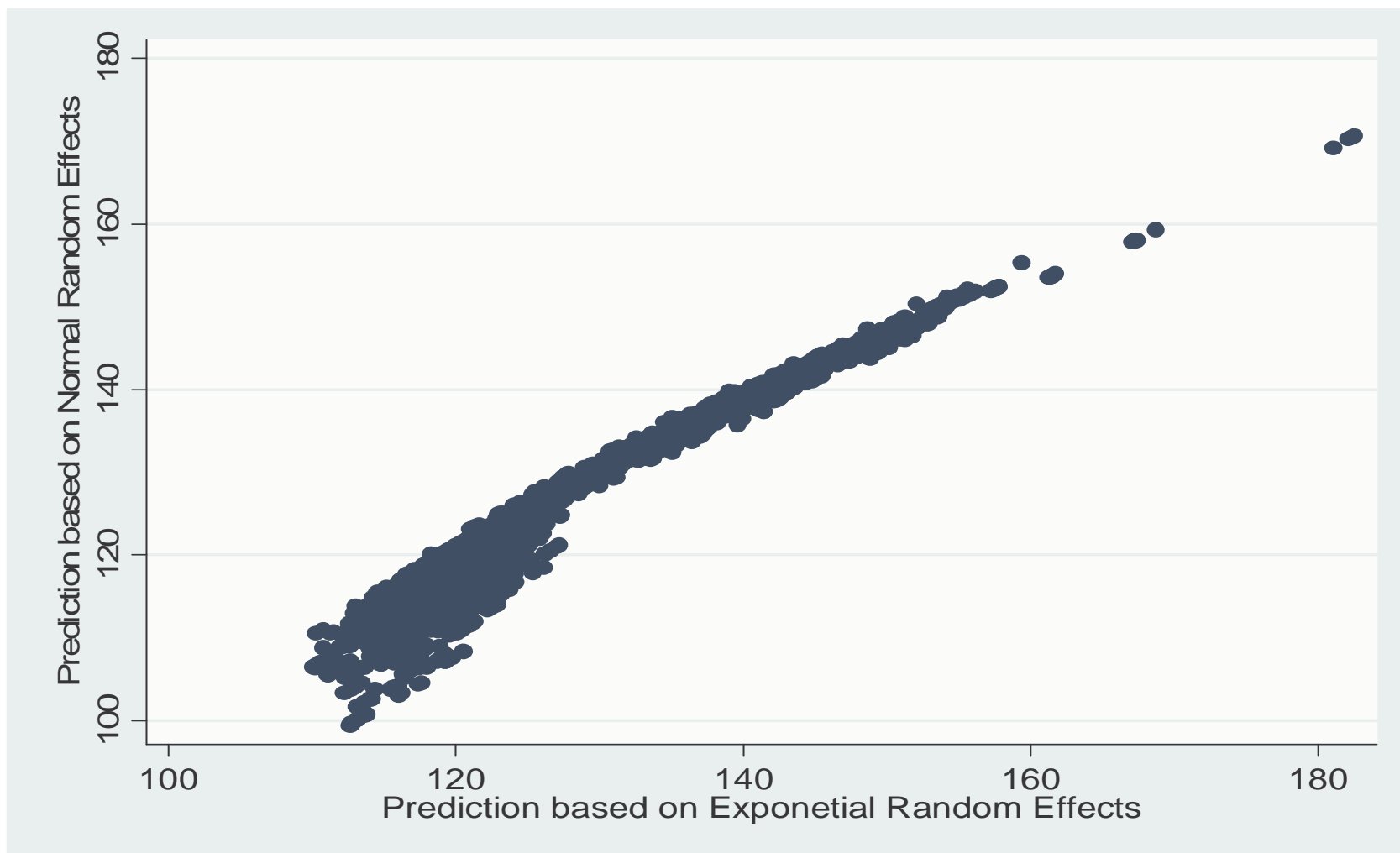
*For predicting the actual systolic blood pressure:*

<u>Method</u>	Prediction Errors		
	Ave	Ave abs	RMSE
Fixed effects only	3.4	13.8	18.1
Mixed model (normal)	3.9	11.0	14.9
Mixed model (exponential)	3.1	11.1	14.9

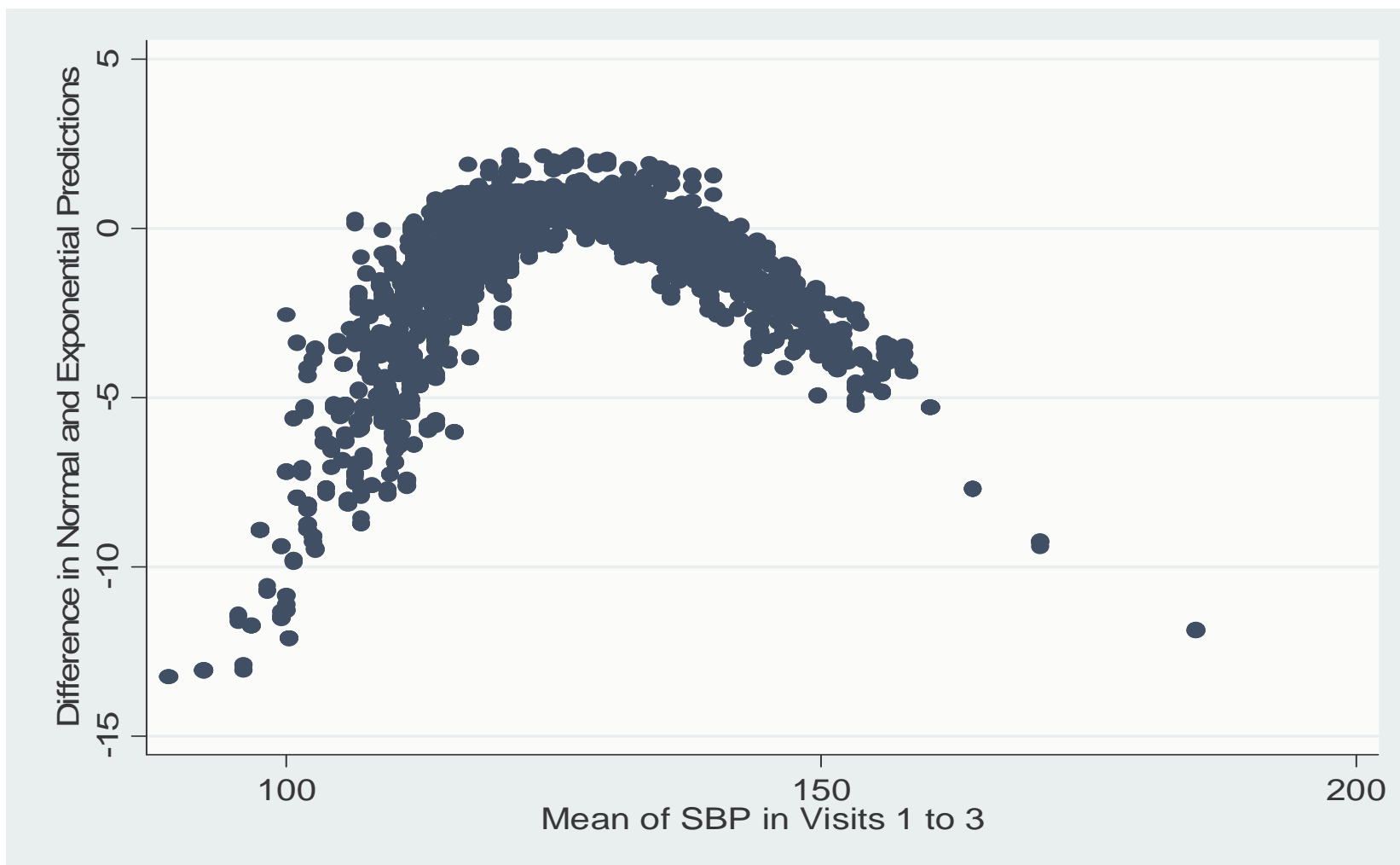
*For predicting high BP or not:*

Area under the ROC curve: Fixed effects – 0.55, Normal – 0.80, Exponential – 0.80.

Do they give the same predicted values? No, but close:



Here is a plot of the difference between the predicted values:





## 8. Summary

- Predicted values of random effects show modest sensitivity to the assumed distributional shape.
- Distribution shape of BLUPs often not reflective of true random effects distribution.
- The ranking of predicted values is little affected.
- Fitting flexible distributional shapes is an easy way to check sensitivity of the results to the assumed shape.

I can be contacted at: [chuck@biostat.ucsf.edu](mailto:chuck@biostat.ucsf.edu)

Talk can be downloaded from my website, which can be found by starting at: [www.biostat.ucsf.edu](http://www.biostat.ucsf.edu)