

Mixed Models for Longitudinal Data: An Applied Introduction

Don Hedeker
University of Illinois at Chicago

This work was supported by National Institute of Mental Health Contract N44MH32056.

Pros of Mixed-effects Regression Models (MRM) over Repeated Measures ANOVA/MANOVA

1. MRM explicitly models individual change across time
2. MRM more flexible in terms of repeated measures
 - (a) need not have same number of obs per subject
 - (b) time can be continuous, rather than a fixed set of points
3. Flexible specification of the covariance structure among repeated measures \Rightarrow methods for testing specific determinants of this structure
4. If data requirements and assumptions of ANOVA/MANOVA apply, MRM yields same estimates for fixed effects
5. MRM can be extended to higher-level models \Rightarrow repeated observations within individuals within clusters
6. Generalizations for non-normal outcomes

2-level model for longitudinal data

$$\begin{array}{ccccccc} \mathbf{y}_i & = & \mathbf{X}_i & \boldsymbol{\beta} & + & \mathbf{Z}_i & \mathbf{v}_i & + & \boldsymbol{\varepsilon}_i \\ n_i \times 1 & & n_i \times p & p \times 1 & & n_i \times r & r \times 1 & & n_i \times 1 \end{array}$$

$i = 1 \dots N$ individuals

$j = 1 \dots n_i$ observations for individual i

$\mathbf{y}_i = n_i \times 1$ response vector for individual i

$\mathbf{X}_i = n_i \times p$ design matrix for the fixed effects

$\boldsymbol{\beta} = p \times 1$ vector of unknown fixed parameters

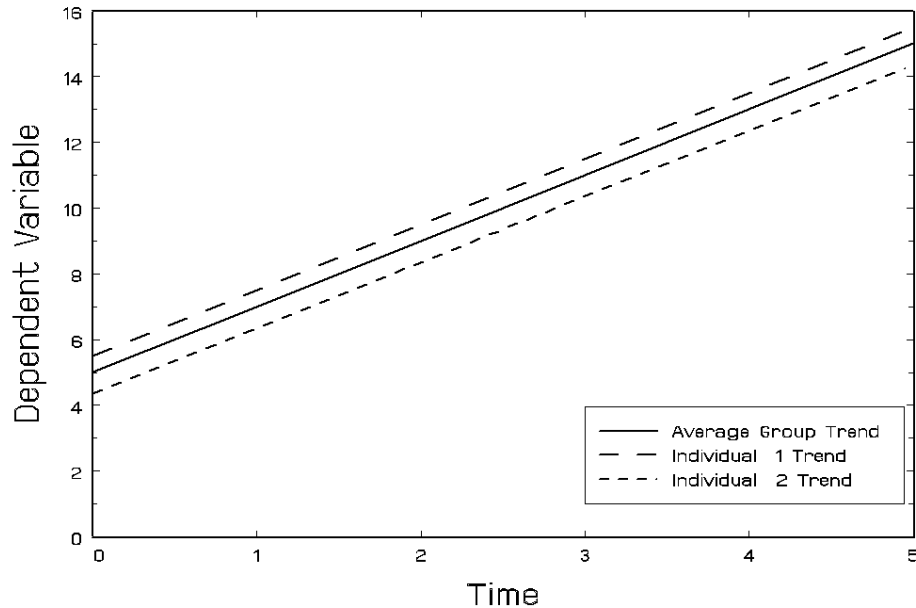
$\mathbf{Z}_i = n_i \times r$ design matrix for the random effects

$\mathbf{v}_i = r \times 1$ vector of unknown random effects $\sim \mathcal{N}(0, \boldsymbol{\Sigma}_v)$

$\boldsymbol{\varepsilon}_i = n_i \times 1$ residual vector $\sim \mathcal{N}(0, \sigma^2 \mathbf{I}_{n_i})$

Random-intercepts Model

each subject is parallel to their group trend



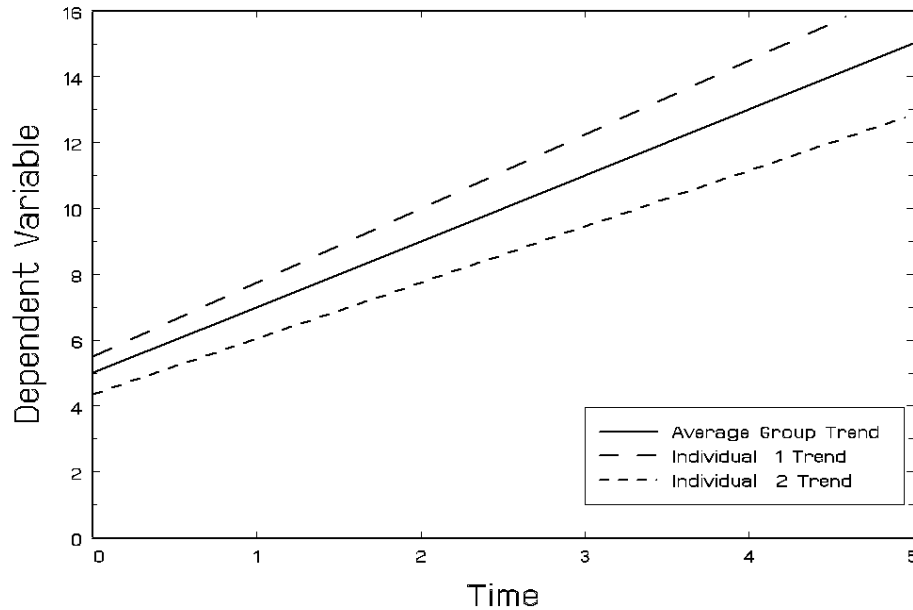
$$y = Time + Grp + (Grp \times Time) + Subj + Error$$

$$y_{ij} = \beta_0 + \beta_1 T_{ij} + \beta_2 G_i + \beta_3 (G_i \times T_{ij}) + v_{0i} + \varepsilon_{ij}$$

$$v_{0i} \sim \mathcal{N}(0, \sigma_v^2) \quad \varepsilon_{ij} \sim \mathcal{N}(0, \sigma^2)$$

Random Intercepts and Trend Model

subjects deviate in terms of both intercept & slope



$$y = Time + Grp + (G \times T) + Subj + (S \times T) + Error$$

$$y_{ij} = \beta_0 + \beta_1 T_{ij} + \beta_2 G_i + \beta_3 (G_i \times T_{ij}) + v_{0i} + v_{1i} T_{ij} + \varepsilon_{ij}$$

$$\begin{bmatrix} v_{0i} \\ v_{1i} \end{bmatrix} \sim \mathcal{N} \left\{ \begin{bmatrix} 0 \\ 0 \end{bmatrix}, \begin{bmatrix} \sigma_{v_0}^2 & \sigma_{v_0 v_1} \\ \sigma_{v_0 v_1} & \sigma_{v_1}^2 \end{bmatrix} \right\} \quad \varepsilon_{ij} \sim \mathcal{N}(0, \sigma^2)$$

Within-Unit / Between-Unit representation

Within-subjects model - level 1 ($j = 1, \dots, n_i$)

$$y_{ij} = b_{0i} + b_{1i}X_{1ij} + \dots + b_{p1i}X_{p1ij} + \varepsilon_{ij}$$

Between-subjects model - level 2 ($i = 1, \dots, N$)

$$b_{0i} = \beta_0 + \boldsymbol{\beta}'_{0(2)}\mathbf{x}_i + v_{0i}$$

$$b_{1i} = \beta_1 + \boldsymbol{\beta}'_{1(2)}\mathbf{x}_i + v_{1i}$$

$$\dots = \dots$$

$$b_{p1i} = \beta_{p1} + \boldsymbol{\beta}'_{p1(2)}\mathbf{x}_i$$

⇒ “slopes as outcomes” model

$$\boldsymbol{\beta}' = \left[\begin{array}{c|c|c|c} \beta_0 & \beta_1 \dots \beta_{p1} & \boldsymbol{\beta}'_{0(2)} & \boldsymbol{\beta}'_{1(2)} \dots \boldsymbol{\beta}'_{p1(2)} \\ \text{intercept} & \text{level-1} & \text{level-2} & \text{cross-level} \end{array} \right]$$

Matrix form of model for individual i

$$\begin{array}{c}
 \begin{bmatrix} y_{i1} \\ y_{i2} \\ \dots \\ y_{in_i} \end{bmatrix} \\
 \mathbf{y}_i \\
 n_i \times 1
 \end{array}
 =
 \begin{array}{c}
 \begin{bmatrix} 1 & Time_{i1} & Group_i & Grp_i \times T_{i1} \\ 1 & Time_{i2} & Group_i & Grp_i \times T_{i2} \\ \dots & \dots & \dots & \dots \\ 1 & Time_{in_i} & Group_i & Grp_i \times T_{in_i} \end{bmatrix} \\
 \mathbf{X}_i \\
 n_i \times p
 \end{array}
 \begin{array}{c}
 \begin{bmatrix} \beta_0 \\ \beta_1 \\ \beta_2 \\ \beta_3 \end{bmatrix} \\
 \boldsymbol{\beta} \\
 p \times 1
 \end{array}$$

$$+
 \begin{array}{c}
 \begin{bmatrix} 1 & Time_{i1} \\ 1 & Time_{i2} \\ \dots & \dots \\ 1 & Time_{in_i} \end{bmatrix} \\
 \mathbf{Z}_i \\
 n_i \times r
 \end{array}
 \begin{array}{c}
 \begin{bmatrix} v_{0i} \\ v_{1i} \end{bmatrix} \\
 \mathbf{v}_i \\
 r \times 1
 \end{array}
 +
 \begin{array}{c}
 \begin{bmatrix} \varepsilon_{i1} \\ \varepsilon_{i2} \\ \dots \\ \varepsilon_{in_i} \end{bmatrix} \\
 \boldsymbol{\varepsilon}_i \\
 n_i \times 1
 \end{array}$$

Time might be years or months, and could differ for each subject

The conditional variance-covariance matrix is now of the form:

- $\Sigma \mathbf{y}_i = \mathbf{Z}_i \Sigma_v \mathbf{Z}_i' + \sigma^2 \mathbf{I}_{n_i}$

For example, with $r = 2$, $n = 3$, and $\mathbf{Z}_i' = \begin{bmatrix} 1 & 1 & 1 \\ 0 & 1 & 2 \end{bmatrix}$

the conditional variance-covariance $\Sigma \mathbf{y}_i = \sigma^2 \mathbf{I}_{n_i} +$

$$\begin{bmatrix} \sigma_{v_0}^2 & \sigma_{v_0}^2 + \sigma_{v_0 v_1} & \sigma_{v_0}^2 + 2\sigma_{v_0 v_1} \\ \sigma_{v_0}^2 + \sigma_{v_0 v_1} & \sigma_{v_0}^2 + 2\sigma_{v_0 v_1} + \sigma_{v_1}^2 & \sigma_{v_0}^2 + 3\sigma_{v_0 v_1} + 2\sigma_{v_1}^2 \\ \sigma_{v_0}^2 + 2\sigma_{v_0 v_1} & \sigma_{v_0}^2 + 3\sigma_{v_0 v_1} + 2\sigma_{v_1}^2 & \sigma_{v_0}^2 + 4\sigma_{v_0 v_1} + 4\sigma_{v_1}^2 \end{bmatrix}$$

- variances and covariances change across time

More general models allow autocorrelated errors,

$\boldsymbol{\varepsilon}_i \sim \mathcal{N}(0, \sigma^2 \boldsymbol{\Omega}_i)$, where $\boldsymbol{\Omega}$ might represent AR or MA process

Estimation - EM algorithm

opposite process of “I do cocaine so I can work more, so I can do more cocaine, so I can work more, *etc.*, ”

Effect of increasing cocaine use

<u>Cocaine</u>		<u>Work</u>	<u>Health</u>
do cocaine	→	work more	declines
do more cocaine	→	work more	declines more
do even more cocaine	→	work even more	declines even more
...
do a ton of cocaine	→	always working	death

Effect of EM estimation of parameters

<u>M-Step (ML)</u>		<u>E-Step (EB)</u>	<u>Estimation</u>
starting values $\beta, \sigma^2, \Sigma_v$	→	estimate $\tilde{v}_i \Sigma_{v y_i}$	improves
re-estimate $\beta, \sigma^2, \Sigma_v$	→	re-estimate $\tilde{v}_i \Sigma_{v y_i}$	improves more
re-re-estimate $\beta, \sigma^2, \Sigma_v$	→	re-re-estimate $\tilde{v}_i \Sigma_{v y_i}$	improves even more
...
RE-estimate $\beta, \sigma^2, \Sigma_v$	→	RE-estimate $\tilde{v}_i \Sigma_{v y_i}$	convergence

→ EM is better than cocaine since EM leads to convergence and not death

EM solution - random intercepts model

- E-step (expectation - “Expected A Posteriori” or Empirical Bayes)

$$\tilde{v}_i = \rho_{n_i n_i} \left[\frac{1}{n_i} \sum_{j=1}^{n_i} y_{ij} - \mathbf{x}'_{ij} \boldsymbol{\beta} \right]$$

$$\sigma_{v|y_i}^2 = \sigma_v^2 (1 - \rho_{n_i n_i}) \quad \text{where } \rho_{n_i n_i} = \frac{n_i r}{1 + (n_i - 1)r} \quad \text{and} \quad r = \frac{\sigma_v^2}{\sigma_v^2 + \sigma^2}$$

- M-step (maximization - “Maximum Likelihood”)

$$\hat{\boldsymbol{\beta}} = \left(\sum_i^N \mathbf{X}'_i \mathbf{X}_i \right)^{-1} \sum_i^N \mathbf{X}'_i (\mathbf{y}_i - \mathbf{1}_i \tilde{v}_i)$$

$$\hat{\sigma}_v^2 = \frac{1}{N} \sum_i^N \tilde{v}_i^2 + \sigma_{v|y_i}^2$$

$$\hat{\sigma}^2 = \left(\sum_i^N n_i \right)^{-1} \sum_i^N (\mathbf{y}_i - \mathbf{X}_i \hat{\boldsymbol{\beta}} - \mathbf{1}_i \tilde{v}_i)' (\mathbf{y}_i - \mathbf{X}_i \hat{\boldsymbol{\beta}} - \mathbf{1}_i \tilde{v}_i) + n_i \sigma_{v|y_i}^2$$

- provide starting values for $\boldsymbol{\beta}$, σ_v^2 , and σ^2
- perform E-step, perform M-step, repeat early and often (until convergence)

Example: Drug Plasma Levels and Clinical Response

Riesby and associates (Riesby *et al.*, 1977) examined the relationship between Imipramine (IMI) and Desipramine (DMI) plasma levels and clinical response in 66 depressed inpatients (37 endogenous and 29 non-endogenous)

		<i>Drug-Washout</i>					
		day0	day7	day14	day21	day28	day35
		wk 0	wk 1	wk 2	wk 3	wk 4	wk 5
Hamilton							
Depression		HD_1	HD_2	HD_3	HD_4	HD_5	HD_6
Diagnosis		Dx					
IMI				IMI_3	IMI_4	IMI_5	IMI_6
DMI				DMI_3	DMI_4	DMI_5	DMI_6
	n	61	63	65	65	63	58

outcome variable - Hamilton Depression Scores (HD)

independent variables - Dx , IMI and DMI

- Dx - endogenous (=1) or non-endogenous (=0)
- IMI (imipramine) drug-plasma levels ($\mu\text{g/l}$)
 - antidepressant given 225 mg/day, weeks 3-6
- DMI (desipramine) drug-plasma levels ($\mu\text{g/l}$)
 - metabolite of imipramine

Descriptive Statistics

Observed HDRS Means, n , and sd

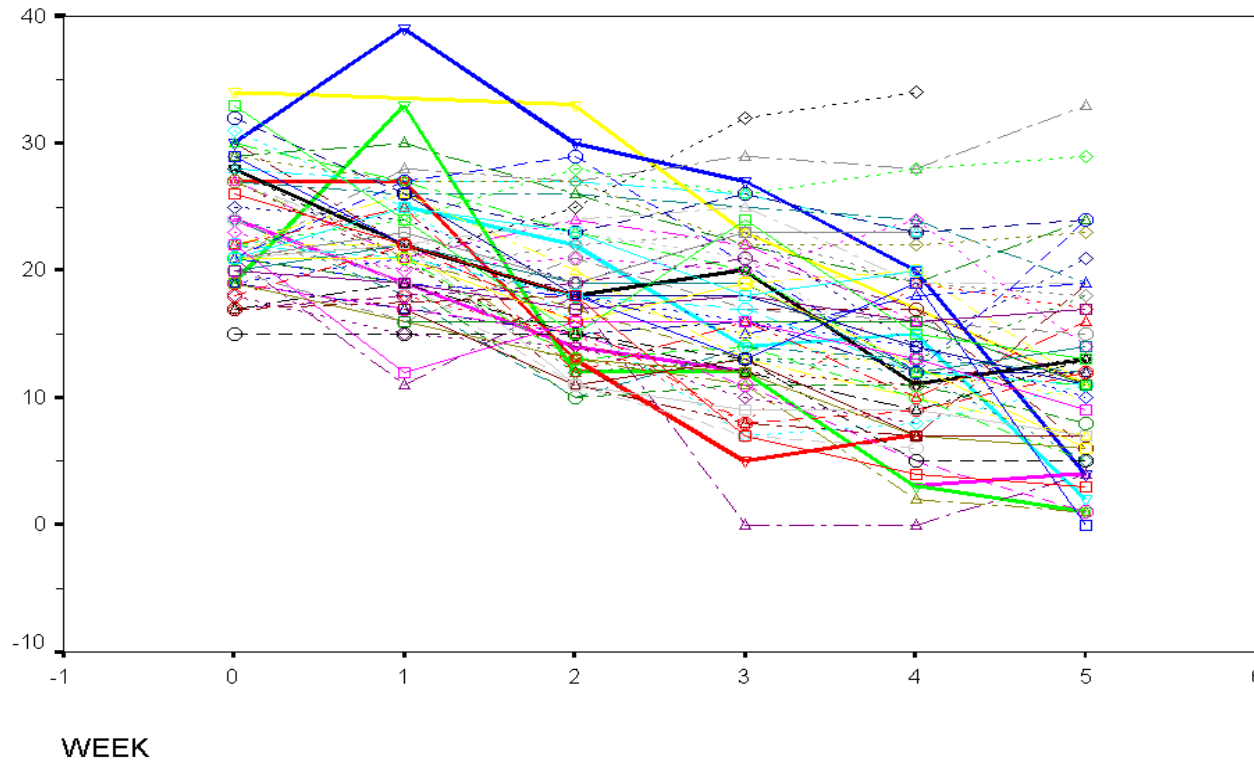
	<i>Washout</i>					
	<u>wk 0</u>	<u>wk 1</u>	<u>wk 2</u>	<u>wk 3</u>	<u>wk 4</u>	<u>wk 5</u>
Endog	24.0	23.0	19.3	17.3	14.5	12.6
n	33	34	37	36	34	31
Non-Endog	22.8	20.5	17.0	15.3	12.6	11.2
n	28	29	28	29	29	27
pooled sd	4.5	4.7	5.5	6.4	7.0	7.2

Correlations: $n = 46$ and $46 \leq n \leq 66$

	<u>wk 0</u>	<u>wk 1</u>	<u>wk 2</u>	<u>wk 3</u>	<u>wk 4</u>	<u>wk 5</u>
week 0	1.0	.49	.41	.33	.23	.18
week 1	.49	1.0	.49	.41	.31	.22
week 2	.42	.49	1.0	.74	.67	.46
week 3	.44	.51	.73	1.0	.82	.57
week 4	.30	.35	.68	.78	1.0	.65
week 5	.22	.23	.53	.62	.72	1.0

Riesby Data - Spaghetti plot (n=66)

Hamilton Depression Scores across Time



- increasing variance across time
- general linear decline over time

Examination of HD across all weeks

$$\begin{array}{c}
 \begin{bmatrix} HD_{i1} \\ HD_{i2} \\ \dots \\ HD_{in_i} \end{bmatrix} \\
 \mathbf{y}_i \\
 n_i \times 1
 \end{array}
 =
 \begin{array}{c}
 \begin{bmatrix} 1 & WEEK_{i1} \\ 1 & WEEK_{i2} \\ \dots & \dots \\ 1 & WEEK_{in_i} \end{bmatrix} \\
 \mathbf{X}_i \\
 n_i \times p
 \end{array}
 \begin{array}{c}
 \begin{bmatrix} \beta_0 \\ \beta_1 \end{bmatrix} \\
 \boldsymbol{\beta} \\
 p \times 1
 \end{array}
 \\
 \\
 +
 \begin{array}{c}
 \begin{bmatrix} 1 & WEEK_{i1} \\ 1 & WEEK_{i2} \\ \dots & \dots \\ 1 & WEEK_{in_i} \end{bmatrix} \\
 \mathbf{Z}_i \\
 n_i \times r
 \end{array}
 \begin{array}{c}
 \begin{bmatrix} \nu_{0i} \\ \nu_{1i} \end{bmatrix} \\
 \boldsymbol{\nu}_i \\
 r \times 1
 \end{array}
 +
 \begin{array}{c}
 \begin{bmatrix} \varepsilon_{i1} \\ \varepsilon_{i2} \\ \dots \\ \varepsilon_{in_i} \end{bmatrix} \\
 \boldsymbol{\varepsilon}_i \\
 n_i \times 1
 \end{array}
 \end{array}$$

where $\max(n_i) = 6$, and $\mathbf{X}'_i = \mathbf{Z}'_i = \begin{bmatrix} 1 & 1 & 1 & 1 & 1 & 1 \\ 0 & 1 & 2 & 3 & 4 & 5 \end{bmatrix}$

Within-subjects and between-subjects components

Within-subjects model

$$HD_{ij} = b_{0i} + b_{1i}Time_{ij} + ERR_{ij}$$

$$y_{ij} = b_{0i} + b_{1i}x_{ij} + \varepsilon_{ij}$$

$$i = 1 \dots 66 \text{ patients}$$

$$j = 1 \dots n_i \text{ observations (max = 6) for patient } i$$

b_{0i} = week 0 HD level for patient i

b_{1i} = weekly change in HD for patient i

Between-subjects models

$$b_{0i} = \beta_0 + v_{0i}$$

$$b_{1i} = \beta_1 + v_{1i}$$

β_0 = average week 0 *HD* level

β_1 = average *HD* weekly improvement

v_{0i} = individual deviation from average intercept

v_{1i} = individual deviation from average improvement

parameter	ML estimate	se	z	$p <$
β_0	23.58	0.55	43.22	.0001
β_1	-2.38	0.21	-11.39	.0001
$\sigma_{v_0}^2$	12.63	3.47		
$\sigma_{v_0v_1}$	-1.42	1.03		
$\sigma_{v_1}^2$	2.08	0.50		
σ^2	12.22	1.11		

$\log L = -1109.52$

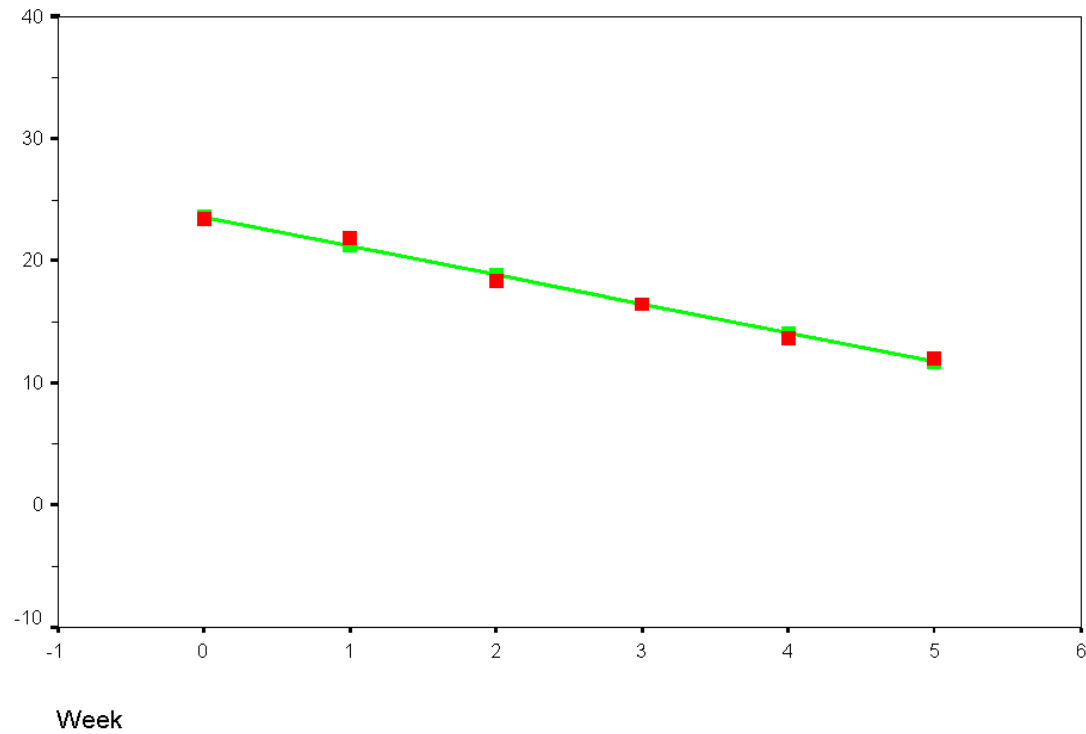
$\chi_2^2 = 66.1, p < .0001$ for $H_0 : \sigma_{v_0v_1} = \sigma_{v_1}^2 = 0$

$\sigma_{v_0v_1}$ as corr between intercept and slope = -0.28

- Wald tests are dubious for variance parameters, likelihood-ratio tests are preferred (though divide p-value by 2)
- Wald z -statistics sometimes expressed as χ_1^2 (by squaring z -value)

Riesby Data - Estimated Average Trend

Hamilton Depression Scores across Time



Observed and estimated means ($= \mathbf{X}\hat{\boldsymbol{\beta}}$)

	wk 0	wk 1	wk 2	wk 3	wk 4	wk 5
<i>n</i>	61	63	65	65	63	58
obs	23.44	21.84	18.31	16.42	13.62	11.95
est	23.58	21.21	18.82	16.45	14.07	11.69

Obs. (pairwise) and est. variance-covariance matrix

$$\Sigma_{\mathbf{y}} = \begin{bmatrix} 20.55 & & & & & & \\ 10.50 & 22.07 & & & & & \\ 10.20 & 12.74 & 30.09 & & & & \\ 9.69 & 12.43 & 25.96 & 41.15 & & & \\ 7.17 & 10.10 & 25.56 & 36.54 & 48.59 & & \\ 6.02 & 7.39 & 18.25 & 26.31 & 32.93 & 52.12 & \end{bmatrix}$$

$$\begin{aligned} \hat{\Sigma}_{\mathbf{y}} &= \mathbf{Z}\hat{\Sigma}_v\mathbf{Z}' + \hat{\sigma}^2\mathbf{I} \\ &= \begin{bmatrix} 24.85 & & & & & & \\ 11.21 & 24.08 & & & & & \\ 9.79 & 12.52 & 27.48 & & & & \\ 8.37 & 13.18 & 18.00 & 35.03 & & & \\ 6.95 & 13.84 & 20.73 & 27.63 & 46.74 & & \\ 5.53 & 14.50 & 23.47 & 32.44 & 41.41 & 62.60 & \end{bmatrix} \end{aligned}$$

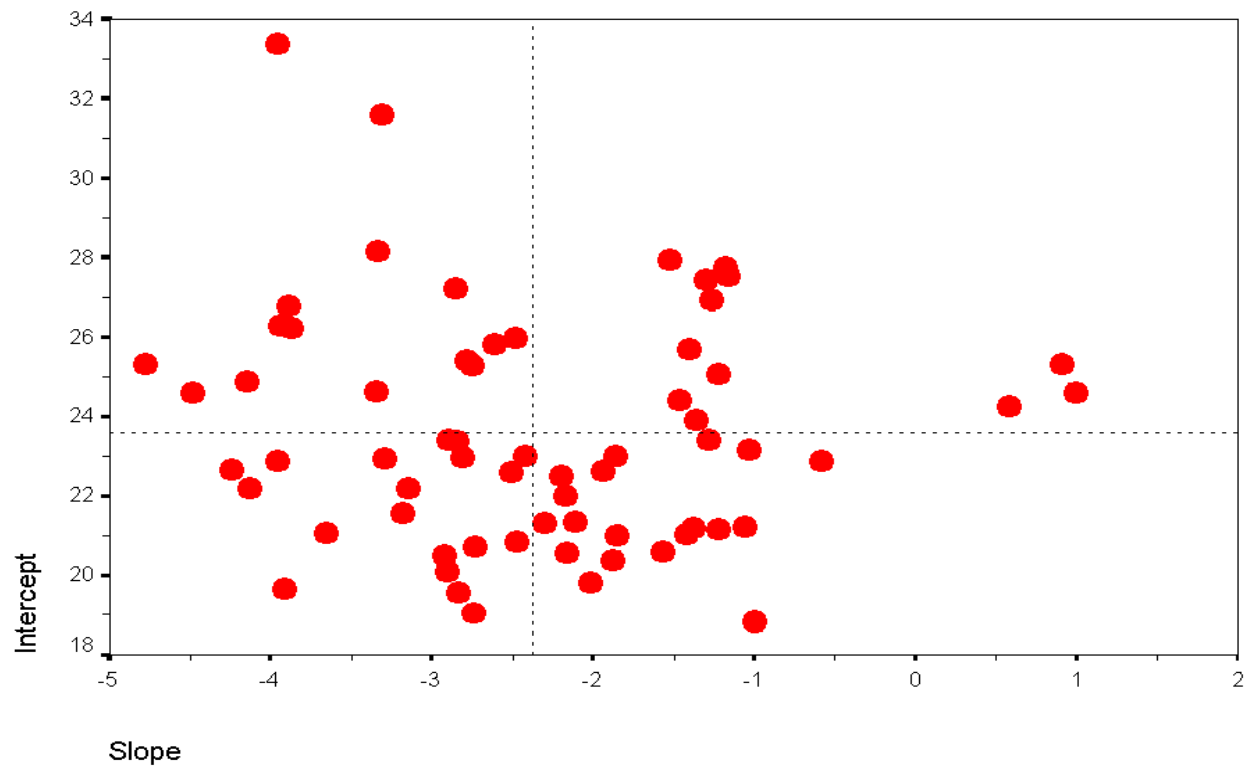
$$\mathbf{Z}' = \begin{bmatrix} 1 & 1 & 1 & 1 & 1 & 1 \\ 0 & 1 & 2 & 3 & 4 & 5 \end{bmatrix} \quad \hat{\Sigma}_v = \begin{bmatrix} 12.63 & -1.42 \\ -1.42 & 2.08 \end{bmatrix}$$

note: from random-int model: $\hat{\sigma}_v^2 = 16.16$ and $\hat{\sigma}^2 = 19.04$

Empirical Bayes estimates of Subject Trends

Riesby Data - Estimated Random Effects

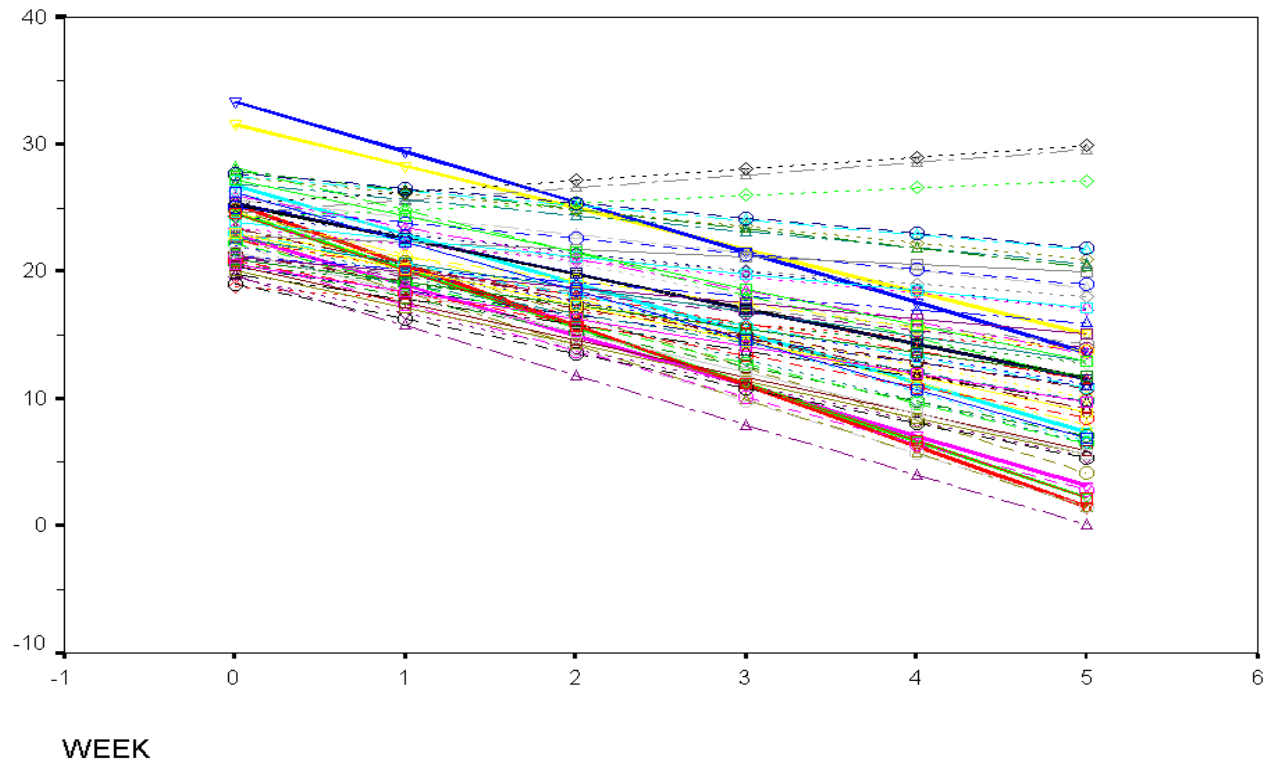
HDRS Intercepts and Slopes



Empirical Bayes estimates of Subject Trends

Riesby Data - Estimated Trends (n=66)

Hamilton Depression Scores across Time



Examination of HD across all weeks by diagnosis

$$\begin{array}{c}
 \begin{bmatrix} HD_{i1} \\ HD_{i2} \\ \dots \\ HD_{in_i} \end{bmatrix} \\
 \mathbf{y}_i \\
 n_i \times 1
 \end{array}
 =
 \begin{array}{c}
 \begin{bmatrix} 1 & WEEK_{i1} & Dx_i & Dx_i * Wk_{i1} \\ 1 & WEEK_{i2} & Dx_i & Dx_i * Wk_{i2} \\ \dots & \dots & \dots & \dots \\ 1 & WEEK_{in_i} & Dx_i & Dx_i * Wk_{in_i} \end{bmatrix} \\
 \mathbf{X}_i \\
 n_i \times p
 \end{array}
 \begin{array}{c}
 \begin{bmatrix} \beta_0 \\ \beta_1 \\ \beta_2 \\ \beta_3 \end{bmatrix} \\
 \boldsymbol{\beta} \\
 p \times 1
 \end{array}$$

$$+
 \begin{array}{c}
 \begin{bmatrix} 1 & WEEK_{i1} \\ 1 & WEEK_{i2} \\ \dots & \dots \\ 1 & WEEK_{in_i} \end{bmatrix} \\
 \mathbf{Z}_i \\
 n_i \times r
 \end{array}
 \begin{array}{c}
 \begin{bmatrix} v_{0i} \\ v_{1i} \end{bmatrix} \\
 \mathbf{v}_i \\
 r \times 1
 \end{array}
 +
 \begin{array}{c}
 \begin{bmatrix} \varepsilon_{i1} \\ \varepsilon_{i2} \\ \dots \\ \varepsilon_{in_i} \end{bmatrix} \\
 \boldsymbol{\varepsilon}_i \\
 n_i \times 1
 \end{array}$$

where $\max(n_i) = 6$, $\mathbf{Z}'_i = \begin{bmatrix} 1 & 1 & 1 & 1 & 1 & 1 \\ 0 & 1 & 2 & 3 & 4 & 5 \end{bmatrix}$, $Dx_i = \begin{cases} 0 & \text{for NE} \\ 1 & \text{for E} \end{cases}$

Within-subjects and between-subjects components

Within-subjects model

$$HD_{ij} = b_{0i} + b_{1i}Time_{ij} + ERR_{ij}$$

b_{0i} = week 0 HD level for patient i

b_{1i} = weekly change in HD for patient i

Between-subjects models

$$b_{0i} = \beta_0 + \beta_2 Dx_i + v_{0i}$$

$$b_{1i} = \beta_1 + \beta_3 Dx_i + v_{1i}$$

β_0 = average week 0 HD level for NE patients ($Dx_i = 0$)

β_1 = average HD weekly improvement for NE patients ($Dx_i = 0$)

β_2 = average week 0 HD difference for E patients

β_3 = average HD weekly improvement difference for endogenous patients

v_{0i} = individual deviation from average intercept

v_{1i} = individual deviation from average improvement

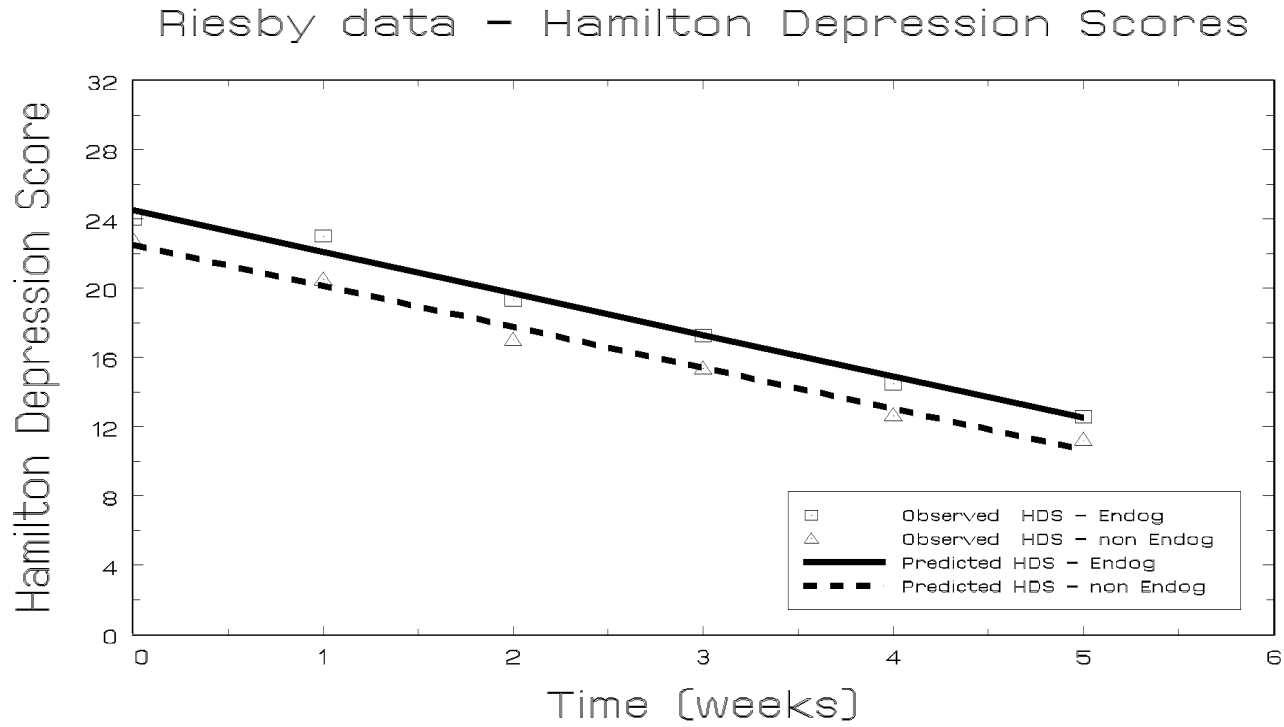
parameter	ML estimate	se	z	$p <$
NE int β_0	22.48	0.79	28.30	.0001
NE slope β_1	-2.37	0.31	-7.59	.0001
E int diff β_2	1.99	1.07	1.86	.063
E slope diff β_3	-0.03	0.42	-0.06	.95
$\sigma_{v_0}^2$	11.64	3.30		
$\sigma_{v_0v_1}$	-1.40	1.00		
$\sigma_{v_1}^2$	2.08	0.50		
σ^2	12.22	1.11		

$$\log L = -1107.47$$

$$\chi_2^2 = 4.1, p \text{ ns, compared to model with } \beta_2 = \beta_3 = 0$$

$$\sigma_{\beta_0\beta_1} \text{ as corr between intercept and slope} = -0.29$$

Riesby data - model fit by diagnosis



Examination of HD across all weeks - quadratic trend

$$\begin{array}{c}
 \begin{bmatrix} HD_{i1} \\ HD_{i2} \\ \dots \\ HD_{in_i} \end{bmatrix} \\
 \mathbf{y}_i \\
 n_i \times 1
 \end{array}
 =
 \begin{array}{c}
 \begin{bmatrix} 1 & WEEK_{i1} & WEEK_{i1}^2 \\ 1 & WEEK_{i2} & WEEK_{i2}^2 \\ \dots & \dots & \dots \\ 1 & WEEK_{in_i} & WEEK_{in_i}^2 \end{bmatrix} \\
 \mathbf{X}_i \\
 n_i \times p
 \end{array}
 \begin{array}{c}
 \begin{bmatrix} \beta_0 \\ \beta_1 \\ \beta_2 \end{bmatrix} \\
 \boldsymbol{\beta} \\
 p \times 1
 \end{array}
 \\
 \\
 +
 \begin{array}{c}
 \begin{bmatrix} 1 & WEEK_{i1} & WEEK_{i1}^2 \\ 1 & WEEK_{i2} & WEEK_{i2}^2 \\ \dots & \dots & \dots \\ 1 & WEEK_{in_i} & WEEK_{in_i}^2 \end{bmatrix} \\
 \mathbf{Z}_i \\
 n_i \times r
 \end{array}
 \begin{array}{c}
 \begin{bmatrix} \nu_{0i} \\ \nu_{1i} \\ \nu_{2i} \end{bmatrix} \\
 \boldsymbol{\nu}_i \\
 r \times 1
 \end{array}
 +
 \begin{array}{c}
 \begin{bmatrix} \varepsilon_{i1} \\ \varepsilon_{i2} \\ \dots \\ \varepsilon_{in_i} \end{bmatrix} \\
 \boldsymbol{\varepsilon}_i \\
 n_i \times 1
 \end{array}
 \end{array}$$

where $\max(n_i) = 6$, and $\mathbf{X}'_i = \mathbf{Z}'_i = \begin{bmatrix} 1 & 1 & 1 & 1 & 1 & 1 \\ 0 & 1 & 2 & 3 & 4 & 5 \\ 0 & 1 & 4 & 9 & 16 & 25 \end{bmatrix}$

Within-subjects and between-subjects components

Within-subjects model

$$HD_{ij} = b_{0i} + b_{1i}Time_{ij} + b_{2i}Time_{ij}^2 + ERR_{ij}$$

$$y_{ij} = b_{0i} + b_{1i}x_{ij} + b_{2i}x_{ij}^2 + \varepsilon_{ij}$$

b_{0i} = week 0 HD level for patient i

b_{1i} = weekly linear change in HD for patient i

b_{2i} = weekly quadratic change in HD for patient i

Between-subjects models

$$b_{0i} = \beta_0 + v_{0i}$$

$$b_{1i} = \beta_1 + v_{1i}$$

$$b_{2i} = \beta_2 + v_{2i}$$

β_0 = average week 0 *HD* level

β_1 = average *HD* weekly linear change

β_2 = average *HD* weekly quadratic change

v_{0i} = individual deviation from average intercept

v_{1i} = individual deviation from average linear change

v_{2i} = individual deviation from average quadratic change

parameter	ML estimate	se	z	$p <$
β_0	23.76	0.55	43.04	.0001
β_1	-2.63	0.48	-5.50	.0001
β_2	0.05	0.09	0.58	.56
$\sigma_{v_0}^2$	10.44	3.58		
$\sigma_{v_0v_1}$	-0.92	2.42		
$\sigma_{v_1}^2$	6.64	2.75		
$\sigma_{v_0v_2}$	-0.11	0.42		
$\sigma_{v_1v_2}$	-0.94	0.48		
$\sigma_{v_2}^2$	0.19	0.09		
σ^2	10.52	1.10		

$$\log L = -1103.82$$

$\chi_4^2 = 11.4, p < 0.025$, compared to model with $\beta_2 = \sigma_{v_2}^2 = \sigma_{v_0v_2} = \sigma_{v_1v_2} = 0$

$\chi_3^2 = 11.0, p < 0.02$, compared to model with $\sigma_{v_2}^2 = \sigma_{v_0v_2} = \sigma_{v_1v_2} = 0$

$\sigma_{v_1v_2}$ as corr between linear and quadratic terms = -0.83

Observed (pairwise) and estimated variance-covariance matrix

$$\Sigma_{\mathbf{y}} = \begin{bmatrix} 20.55 & & & & & \\ 10.50 & 22.07 & & & & \\ 10.20 & 12.74 & 30.09 & & & \\ 9.69 & 12.43 & 25.96 & 41.15 & & \\ 7.17 & 10.10 & 25.56 & 36.54 & 48.59 & \\ 6.02 & 7.39 & 18.25 & 26.31 & 32.93 & 52.12 \end{bmatrix}$$

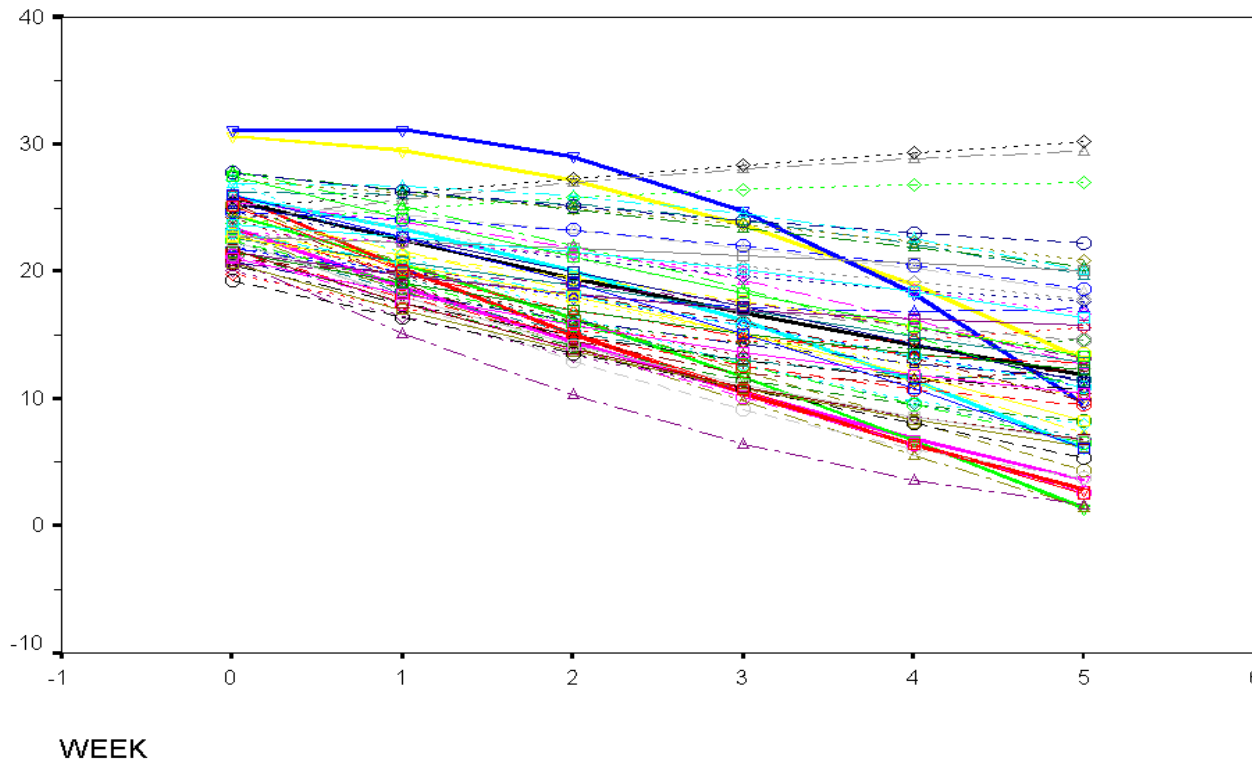
$$\begin{aligned} \hat{\Sigma}_{\mathbf{y}} &= \mathbf{Z}\hat{\Sigma}_v\mathbf{Z}' + \hat{\sigma}^2\mathbf{I} \\ &= \begin{bmatrix} 20.96 & & & & & \\ 9.41 & 23.86 & & & & \\ 8.16 & 15.57 & 31.07 & & & \\ 6.68 & 16.08 & 23.11 & 38.31 & & \\ 4.98 & 14.88 & 23.26 & 30.12 & 45.98 & \\ 3.06 & 11.97 & 20.98 & 30.09 & 39.29 & 59.11 \end{bmatrix} \end{aligned}$$

$$\text{where } \mathbf{Z}' = \begin{bmatrix} 1 & 1 & 1 & 1 & 1 & 1 \\ 0 & 1 & 2 & 3 & 4 & 5 \\ 0 & 1 & 4 & 9 & 16 & 25 \end{bmatrix} \quad \hat{\Sigma}_v = \begin{bmatrix} 10.44 & -0.92 & -0.11 \\ -0.92 & 6.64 & -0.94 \\ -0.11 & -0.94 & 0.19 \end{bmatrix}$$

Empirical Bayes estimates of Subject Trends

Riesby Data - Estimated Curvilinear Trends (n=66)

Hamilton Depression Scores across Time



Examination of HD across 4 weeks by plasma drug-levels

$$\begin{array}{c}
 \begin{bmatrix} HD_{i1} \\ HD_{i2} \\ \dots \\ HD_{in_i} \end{bmatrix} \\
 \mathbf{y}_i \\
 n_i \times 1
 \end{array}
 =
 \begin{array}{c}
 \begin{bmatrix} 1 & WEEK_{i1} & \ln IMI_{i1} & \ln DMI_{i1} \\ 1 & WEEK_{i2} & \ln IMI_{i2} & \ln DMI_{i2} \\ \dots & \dots & \dots & \dots \\ 1 & WEEK_{in_i} & \ln IMI_{in_i} & \ln DMI_{in_i} \end{bmatrix} \\
 \mathbf{X}_i \\
 n_i \times p
 \end{array}
 \begin{array}{c}
 \begin{bmatrix} \beta_0 \\ \beta_1 \\ \beta_2 \\ \beta_3 \end{bmatrix} \\
 \boldsymbol{\beta} \\
 p \times 1
 \end{array}
 \\
 \\
 +
 \begin{array}{c}
 \begin{bmatrix} 1 & WEEK_{i1} \\ 1 & WEEK_{i2} \\ \dots & \dots \\ 1 & WEEK_{in_i} \end{bmatrix} \\
 \mathbf{Z}_i \\
 n_i \times r
 \end{array}
 \begin{array}{c}
 \begin{bmatrix} v_{0i} \\ v_{1i} \end{bmatrix} \\
 \mathbf{v}_i \\
 r \times 1
 \end{array}
 +
 \begin{array}{c}
 \begin{bmatrix} \varepsilon_{i1} \\ \varepsilon_{i2} \\ \dots \\ \varepsilon_{in_i} \end{bmatrix} \\
 \boldsymbol{\varepsilon}_i \\
 n_i \times 1
 \end{array}
 \end{array}$$

where $\max(n_i) = 4$, and $\mathbf{Z}'_i = \begin{bmatrix} 1 & 1 & 1 & 1 \\ 0 & 1 & 2 & 3 \end{bmatrix}$

Within-subjects and between-subjects components

Within-subjects model

$$HD_{ij} = b_{0i} + b_{1i}T_{ij} + b_{2i} \ln IMI_{ij} + b_{3i} \ln DMI_{ij} + Res_{ij}$$

b_{0i} = week 2 HD level for patient i with both $\ln IMI$ and $\ln DMI = 0$

b_{1i} = weekly change in HD for patient i

b_{2i} = change in HD due to $\ln IMI$

b_{3i} = change in HD due to $\ln DMI$

Between-subjects models

$$b_{0i} = \beta_0 + v_{0i}$$

$$b_{1i} = \beta_1 + v_{1i}$$

$$b_{2i} = \beta_2$$

$$b_{3i} = \beta_3$$

- β_0 = average week 2 *HD* level for drug-free patients
- β_1 = average *HD* weekly improvement
- β_2 = average *HD* difference for unit change in $\ln IMI$
- β_3 = average *HD* difference for unit change in $\ln DMI$
- v_{0i} = individual intercept deviation from model
- v_{1i} = individual slope deviation from model

Here, week 2 is the actual study week (*i.e.*, one week after the drug washout period), which is coded as 0 in this analysis of the last four study timepoints

parameter	ML estimate	se	z	$p <$
int β_0	21.37	3.89	5.49	.0001
slope β_1	-2.03	0.28	-7.15	.0001
$\ln IMI$ β_2	0.60	0.85	0.71	.48
$\ln DMI$ β_3	-1.20	0.63	-1.90	.06
$\sigma_{v_0}^2$	24.83	5.75		
$\sigma_{v_0v_1}$	-0.72	1.72		
$\sigma_{v_1}^2$	2.73	0.93		
σ^2	10.46	1.35		

$$\log L = -751.23$$

$\sigma_{v_0v_1}$ as corr between intercept and slope = -0.09

parameter	estimate	se	$p <$
<i>HD total score</i>			
intercept β_0	10.97	4.44	.013
slope β_1	-1.99	0.28	.0001
Baseline HD β_2	0.54	0.14	.0001
ln IMI β_3	0.54	0.78	.49
ln DMI β_4	-1.63	0.59	.006
$\sigma_{v_0}^2$	17.82	4.55	
$\sigma_{v_0v_1}$	0.08	1.53	
$\sigma_{v_1}^2$	2.74	0.94	
σ^2	10.50	1.36	
<i>HD change from baseline</i>			
intercept β_0	1.52	3.74	ns
slope β_1	-1.97	0.28	.0001
ln IMI β_3	0.63	0.82	ns
ln DMI β_4	-1.97	0.60	.001
$\sigma_{v_0}^2$	20.50	5.01	
$\sigma_{v_0v_1}$	0.84	1.58	
$\sigma_{v_1}^2$	2.78	0.94	
σ^2	10.53	1.36	

Correlation between HD scores
and plasma levels (ln units)

	HD total score			
	week 2	week 3	week 4	week 5
IMI	-0.034	-0.034	-0.003	-0.189
DMI	-0.178	-0.075	-0.250*	-0.293*
	HD change from baseline			
	week 2	week 3	week 4	week 5
IMI	-0.025	-0.100	-0.034	-0.250
DMI	-0.350*	-0.274*	-0.348*	-0.401*
* $p < 0.05$				

Practical Issues - # of random effects depends on sample sizes

- number of repeated observations n_i
 - simpler choices if n is small
 - * random-int with CS assumption more reasonable ($q = 2$, number of var-covar parameters)
 - * need to check degree of within-subject variation of outcome variable (esp. for categorical outcomes)
 - more choices as n gets large
- number of subjects N
 - more limited choices as N is small
 - * not much information for estimation of var-covar matrix
 - more possibilities as N is large

Practical Issues - estimation problems usually occur for variance-covariance parameters, not fixed effects

- near-zero variance components
 - rescale time to larger units (*e.g.*, months instead of days)
- extrapolation of intercept variance
 - for models with multiple random effects, ensure that 0 value (of variables with random effects) is meaningful; centering variables helps
- near-unity correlation of variance terms
 - orthogonal polynomials instead of raw metric for time effects
 - may suggest simpler model - not enough information from data to simultaneously estimate all model parameters

Practical Issues - ML or REML estimation

- ML estimates of variance parameters biased downward when N is small and p (number of covariates) is large
 - *e.g.*, ML estimate of error variance = SSE / N in ordinary regression, instead of $\text{SSE} / (N - p - 1)$
- use of likelihood-ratio tests for comparison of nested models is tricky under REML
 - ok only if the covariates in the two models are identical;
i.e., ok only for comparisons of different models of $V(\mathbf{y})$

Practical Issues - Statistical tests of fixed effects

- likelihood-ratio tests for nested models
 - *e.g.*, model with group and time versus model with group, time, and group by time
 - sample size must be identical (careful with inclusion of covariates)
 - only valid under ML, not REML
- Wald tests (z-statistics) for specific model parameters
 - estimate / standard error $\sim N(0,1)$ under null hypothesis

Practical Issues - Statistical tests of var-covar parameters

- Wald tests generally not recommended
 - normality for the sampling distribution of variance parameters is not reasonable
- LR tests under REML or ML
 - REML tests slightly better, but both accept null hypothesis (of variance parameters = 0) too often
 - simulation studies suggest halved p -values for null hypothesis tests of var-covar parameters

Practical Issues - Model selection

- tests of fixed effects depends on structure fitted for $V(\mathbf{y})$
- many possible models for $V(\mathbf{y})$: random intercepts, random intercepts and trends, random intercepts trends and AR(1) errors . . .
- 2-step procedure for model selection
 1. fit model with all fixed effects of potential interest and perform model selection of $V(\mathbf{y})$ structure
 - Likelihood ratio tests for nested models; simulation studies suggest halved p -values for null hypothesis tests of var-covar parameters
 - AIC and BIC for non-nested models (penalized versions of log-likelihood value for number of parameters)
 2. test fixed effects using model selected for $V(\mathbf{y})$