

Longitudinal Data Analysis: Overview and Mixed Effects Models

Don Hedeker

University of Illinois at Chicago

hedeker@uic.edu

www.uic.edu/~hedeker/long.html

Advantages of Longitudinal Studies

- Economizes on subjects; subjects serve as own control
- Between-subject variation excluded from error
- Can provide more efficient estimators than cross-sectional designs with same number and pattern of observations
- Can separate aging effects (changes over time within individuals) from cohort effects (differences between subjects at baseline)
⇒ cross-sectional design can't do this
- Can provide information about individual change

Analysis Considerations

- Response variable
 - continuous (normal or non-normal)
 - categorical (dichotomous, ordinal, nominal, counts)
- Number of subjects N , & number of obs per subject n_i
 - $n_i = 2$ for all: change score analysis or ANCOVA
 - $n_i = n$ for all: balanced design - ANOVA or MANOVA
 - n_i varies: more general methods
- Number & type of covariates - $E(\mathbf{y}_i)$
 - one sample, multiple samples
 - regression (continuous or categorical covariates)
 - time-varying covariates
- Type of variance-covariance structure - $V(\mathbf{y}_i)$
 - homogeneous or heterogeneous variances/covariances

General Approaches

- Derived variable: not really longitudinal, per se, reduce the repeated observations into a summary variable
 - average across time, change score, linear trend across time, last observation
- Longitudinal Analysis
 - ANOVA/MANOVA for repeated measures
 - Mixed-effects regression models
 - Covariance pattern models
 - Generalized Estimating Equations (GEE) models
 - Structural Equations Models
 - Transition Models

Advantages of Mixed-effects Regression Models (MRM)

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. MRM can be extended to higher-level models \Rightarrow repeated observations within individuals within clusters
5. Generalizations for non-normal data

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

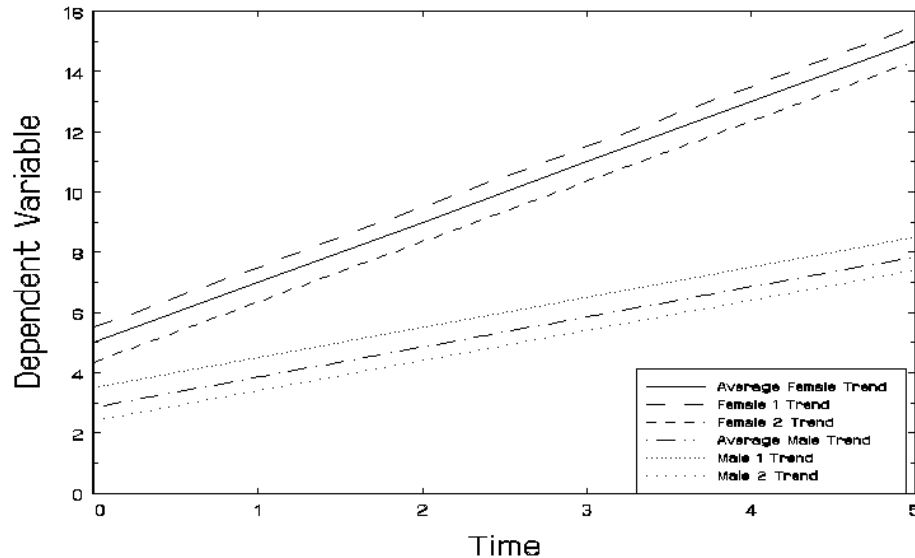
$\mathbf{Z}_i = n_i \times r$ design matrix for the random effects

$\mathbf{v}_i = r \times 1$ 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



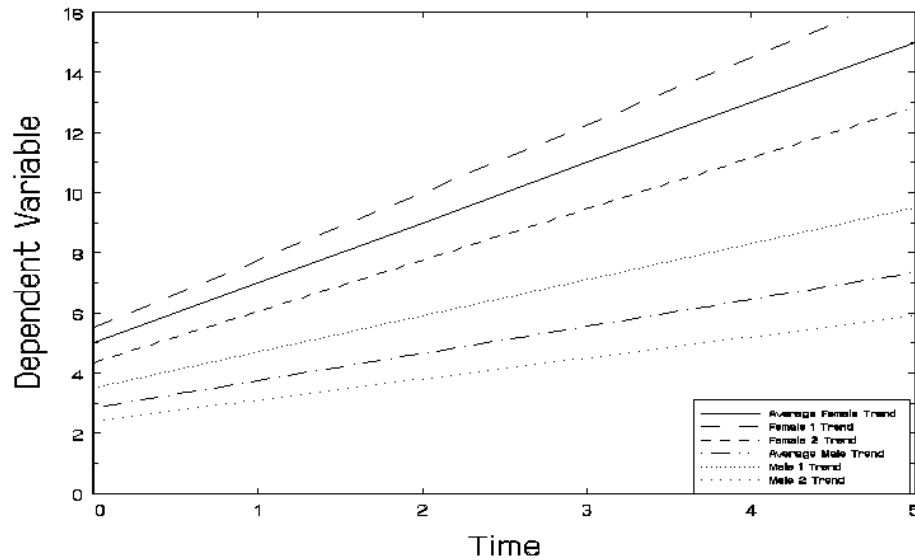
$$y = Time + Grp + (Grp \times Time) + Subj + Error$$

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

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

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 ($i = 1, \dots, N$)

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

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

$$\dots = \dots$$

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

\Rightarrow “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

$$\begin{array}{c}
 \begin{bmatrix} y_{i1} \\ y_{i2} \\ \dots \\ y_{in_i} \end{bmatrix} \\
 \mathbf{y}_i \\
 n_i \times 1
 \end{array}
 =
 \begin{array}{c}
 \begin{bmatrix} 1 & Time_{i1} & Group_i & Grp_i \times T_{i1} \\ 1 & Time_{i2} & Group_i & Grp_i \times T_{i2} \\ \dots & \dots & \dots & \dots \\ 1 & Time_{in_i} & Group_i & Grp_i \times T_{in_i} \end{bmatrix} \\
 \mathbf{X}_i \\
 n_i \times p
 \end{array}
 \begin{array}{c}
 \begin{bmatrix} \beta_0 \\ \beta_1 \\ \beta_2 \\ \beta_3 \end{bmatrix} \\
 \boldsymbol{\beta} \\
 p \times 1
 \end{array}
 \\
 \\
 +
 \begin{array}{c}
 \begin{bmatrix} 1 & Time_{i1} \\ 1 & Time_{i2} \\ \dots & \dots \\ 1 & Time_{in_i} \end{bmatrix} \\
 \mathbf{Z}_i \\
 n_i \times r
 \end{array}
 \begin{array}{c}
 \begin{bmatrix} v_{0i} \\ v_{1i} \end{bmatrix} \\
 \mathbf{v}_i \\
 r \times 1
 \end{array}
 +
 \begin{array}{c}
 \begin{bmatrix} \varepsilon_{i1} \\ \varepsilon_{i2} \\ \dots \\ \varepsilon_{in_i} \end{bmatrix} \\
 \boldsymbol{\varepsilon}_i \\
 n_i \times 1
 \end{array}
 \end{array}$$

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

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

- 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

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

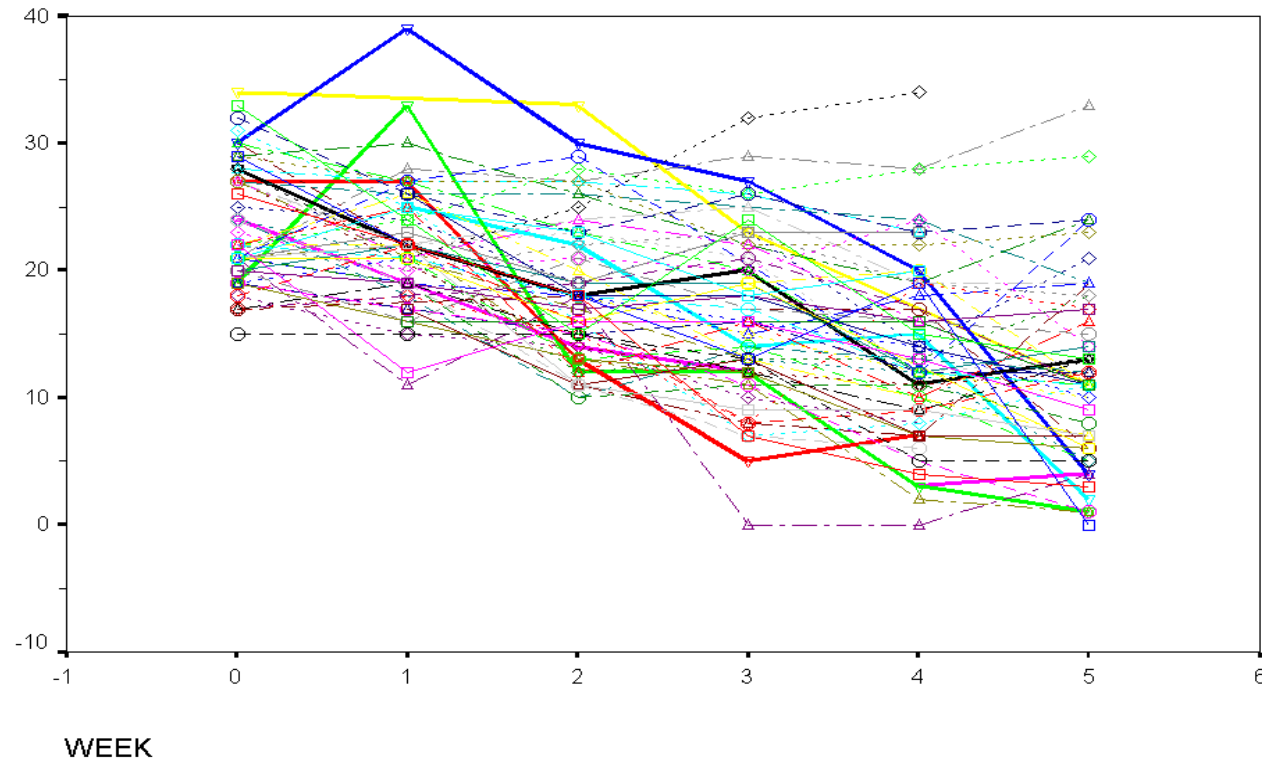
	<i>Washout</i>					
	<u>wk 0</u>	<u>wk 1</u>	<u>wk 2</u>	<u>wk 3</u>	<u>wk 4</u>	<u>wk 5</u>
Endog	24.0	23.0	19.3	17.3	14.5	12.6
n	33	34	37	36	34	31
Non-Endog	22.8	20.5	17.0	15.3	12.6	11.2
n	28	29	28	29	29	27
pooled sd	4.5	4.7	5.5	6.4	7.0	7.2

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

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

where $\max(n_i) = 6$, 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

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

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

i = 1...66 patients

j = 1... n_i 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		

$\log L = -1109.52$

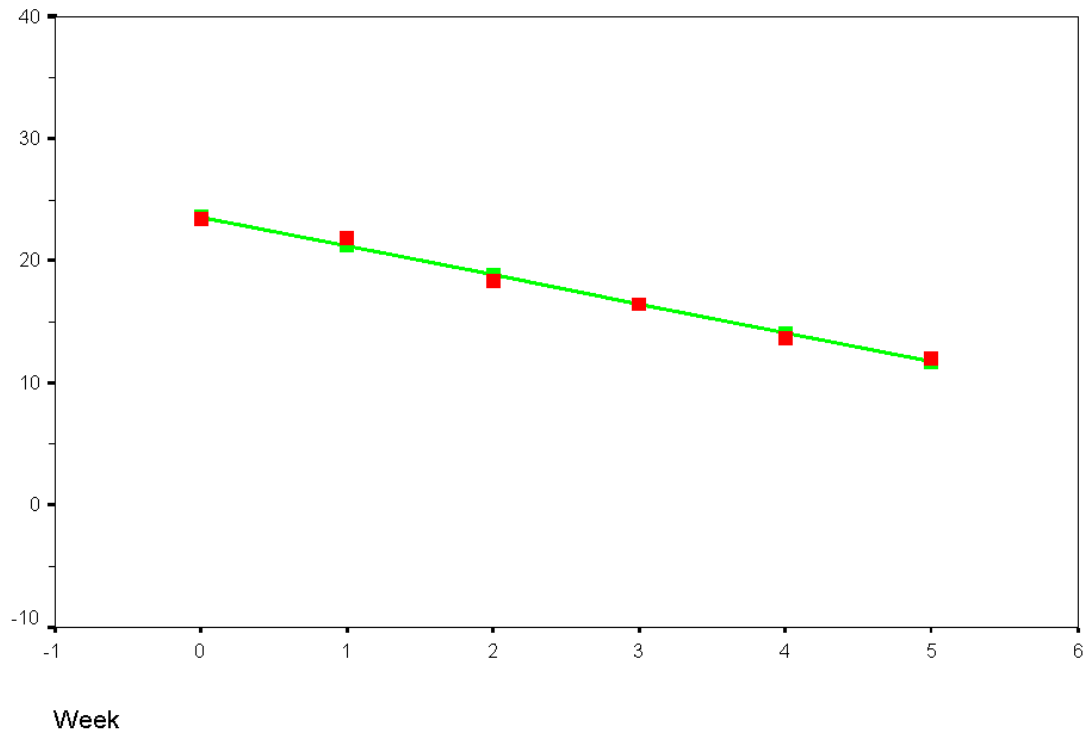
$\chi_2^2 = 66.1, p < .0001$ for $H_0: \sigma_{v_0v_1} = \sigma_{v_1}^2 = 0$

$\sigma_{v_0v_1}$ 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 ($= \mathbf{X}\hat{\beta}$)

	wk 0	wk 1	wk 2	wk 3	wk 4	wk 5
n	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

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

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

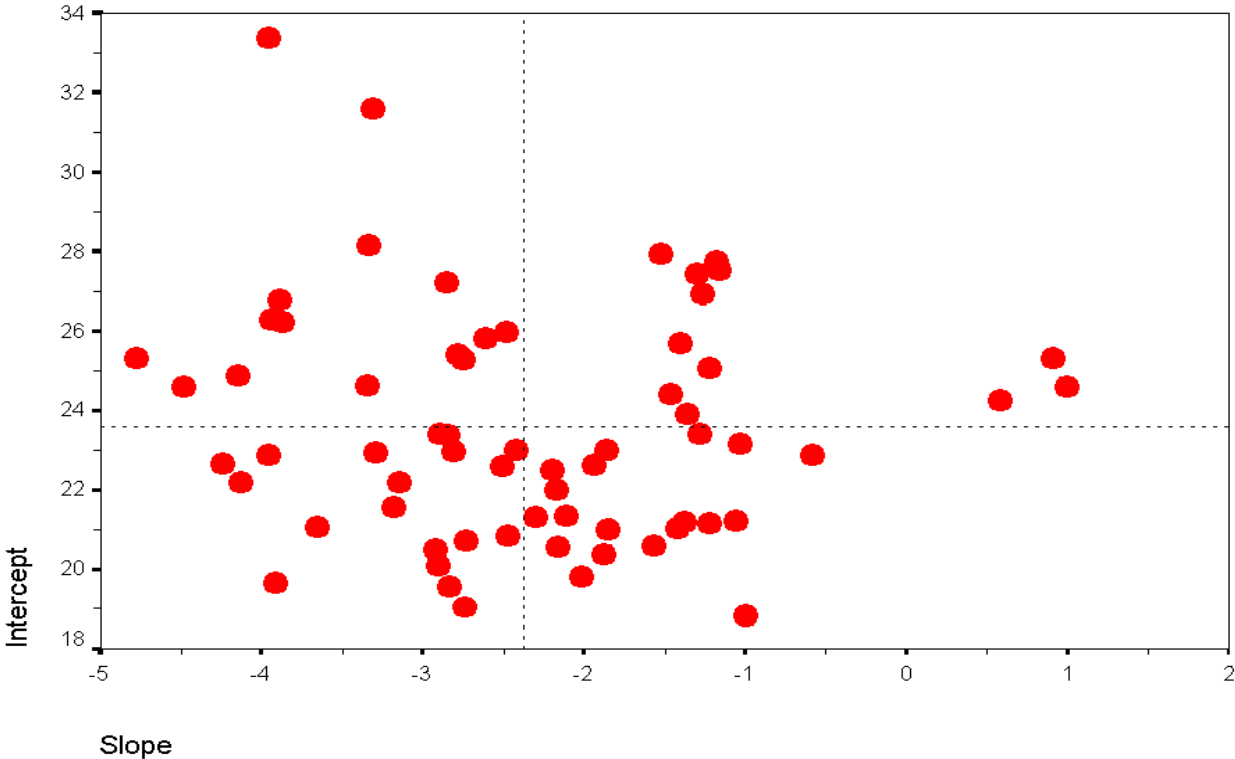
$$\mathbf{Z}' = \begin{bmatrix} 1 & 1 & 1 & 1 & 1 & 1 \\ 0 & 1 & 2 & 3 & 4 & 5 \end{bmatrix} \quad \hat{\Sigma}_v = \begin{bmatrix} 12.63 & -1.42 \\ -1.42 & 2.08 \end{bmatrix}$$

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

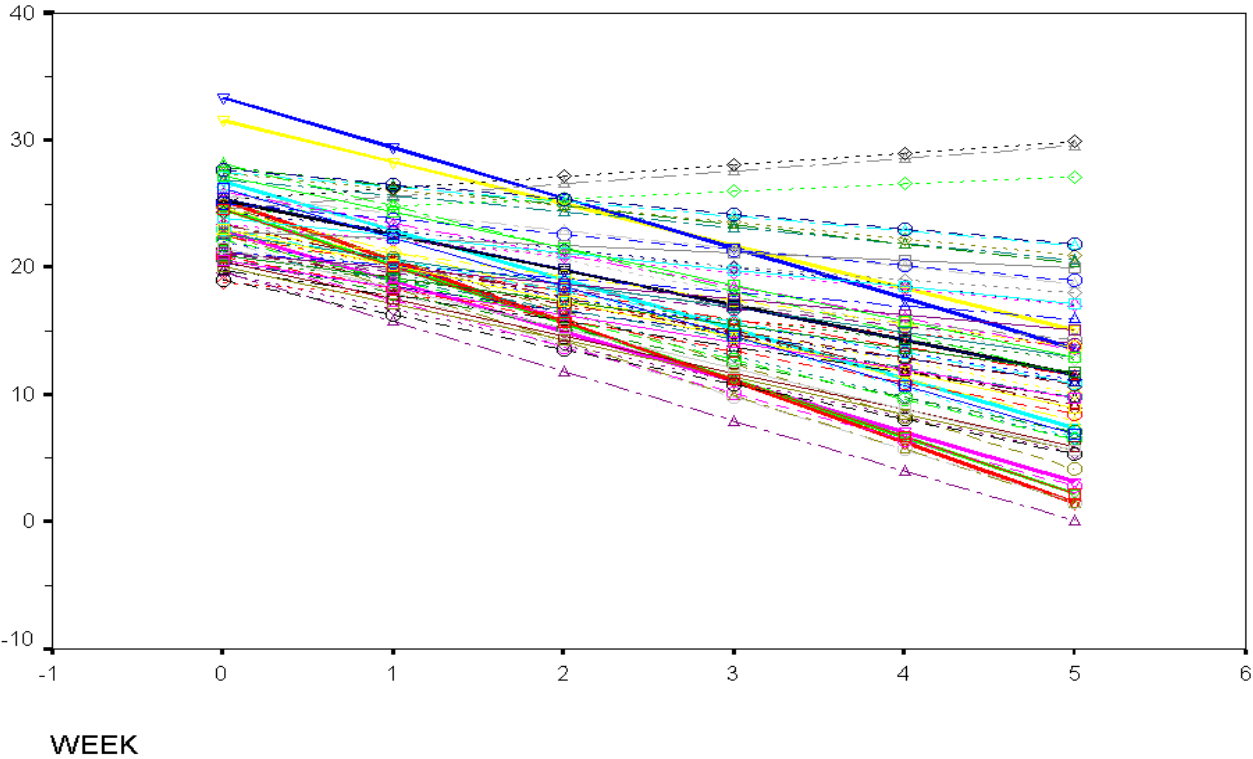
HDRS Intercepts and Slopes



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

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

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

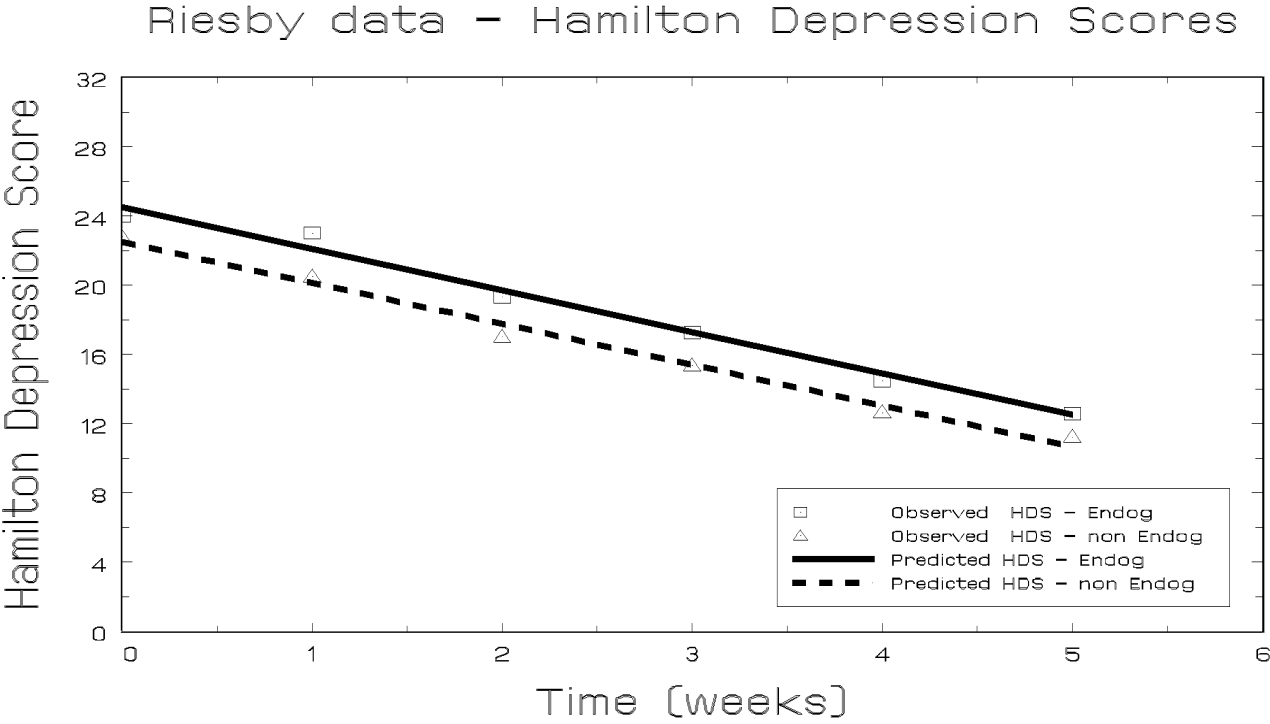
parameter	ML estimate	se	z	$p <$
NE int β_0	22.48	0.79	28.30	.0001
NE slope β_1	-2.37	0.31	-7.59	.0001
E int diff β_2	1.99	1.07	1.86	.063
E slope diff β_3	-0.03	0.42	-0.06	.95
$\sigma_{v_0}^2$	11.64	3.53		
$\sigma_{v_0v_1}$	-1.40	1.00		
$\sigma_{v_1}^2$	2.08	0.50		
σ^2	12.22	1.11		

$$\log L = -1107.47$$

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

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

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

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

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

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

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

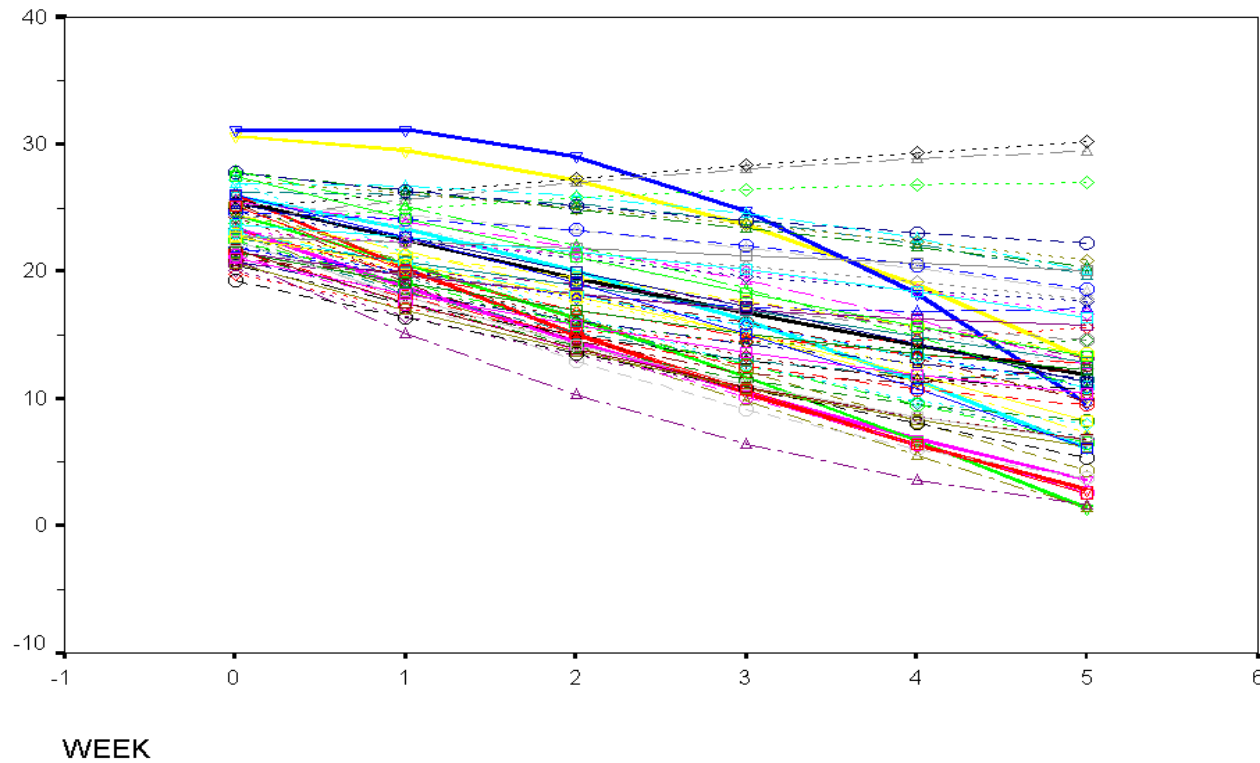
$$\begin{aligned} \hat{\Sigma}_{\mathbf{y}} &= \mathbf{Z}\hat{\Sigma}_v\mathbf{Z}' + \hat{\sigma}^2\mathbf{I} \\ &= \begin{bmatrix} 20.96 & & & & & \\ 9.41 & 23.86 & & & & \\ 8.16 & 15.57 & 31.07 & & & \\ 6.68 & 16.08 & 23.11 & 38.31 & & \\ 4.98 & 14.88 & 23.26 & 30.12 & 45.98 & \\ 3.06 & 11.97 & 20.98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

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

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

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

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

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*
* $p < 0.05$				

SAS MIXED code

```
TITLE1 'Analysis of Riesby data - HDRS scores across time';
DATA ONE; INFILE 'C:RIESBY.DAT';
INPUT ID HamD Intcpt Week Endog EndWeek ;

PROC FORMAT;
VALUE Endog 0='NonEndog' 1='Endog';
VALUE Week 0='week 0' 1='week 1' 2='week 2' 3='week 3' 4='week 4' 5='week 5';

PROC MIXED METHOD=ML COVTEST;
CLASS ID;
MODEL HAMD = WEEK /SOLUTION;
RANDOM INTERCEPT /SUB=ID TYPE=UN G;
TITLE2 'Random intercepts model:  compound symmetry structure';

PROC MIXED METHOD=ML COVTEST;
CLASS ID;
MODEL HAMD = WEEK /SOLUTION;
RANDOM INTERCEPT WEEK /SUB=ID TYPE=UN G GCORR;
TITLE2 'Random trend model';
```

```
PROC MIXED METHOD=ML COVTEST;  
CLASS ID;  
MODEL HAMD = WEEK ENDOG ENDWEEK /SOLUTION;  
RANDOM INTERCEPT WEEK /SUB=ID TYPE=UN G GCORR;  
TITLE2 'Random trend model with group effects';
```

```
PROC MIXED METHOD=ML COVTEST;  
CLASS ID;  
MODEL HAMD = WEEK WEEK*WEEK /SOLUTION;  
RANDOM INTERCEPT WEEK WEEK*WEEK /SUB=ID TYPE=UN G GCORR;  
TITLE2 'Random quadratic trend model';
```

```
RUN;
```

Missing Data and Incomplete Data Models

$$\mathbf{y} \begin{cases} \mathbf{y}^{(O)} & R = 0 \text{ (observed)} \\ \mathbf{y}^{(M)} & R = 1 \text{ (missing)} \end{cases}$$

- GEE assumes special case of “Missing Completely at Random” (MCAR)

$$P(\mathbf{R} \mid \mathbf{y}, \mathbf{X}) = P(\mathbf{R} \mid \mathbf{X}) \text{ for all } \mathbf{y}$$

conditional on covariates, \mathbf{R} is independent of both $\mathbf{y}^{(O)}$ and $\mathbf{y}^{(M)}$

\Rightarrow “covariate-dependent missingness”

- Likelihood-based MRM assumes “Missing at Random” (MAR)

$$P(\mathbf{R} \mid \mathbf{y}, \mathbf{X}) = P(\mathbf{R} \mid \mathbf{X}, \mathbf{y}^{(O)}) \text{ for all } \mathbf{y}^{(M)}$$

conditional on covariates *and* observed values of the dependent variable, \mathbf{R} is independent of $\mathbf{y}^{(M)}$

\Rightarrow “ignorable non-response”

Missing Not At Random (MNAR) Models

- When the data are nonignorable (*i.e.*, MNAR), standard statistical models can yield badly biased results
- The observed data provide no information to either confirm or refute ignorability

⇒ **cannot test MAR versus MNAR**

Two general classes of MNAR models

- Selection models - modeling of both the longitudinal and missingness processes
- Pattern mixture models - use missing data pattern information in the longitudinal modeling

⇒ will be illustrated in terms of MRMs, however they can be more broadly defined and utilized

Comments on MNAR models

- Ordinary MRM (and other full-likelihood models) assume MAR, these extended models do not
- Use of nonignorable models can be helpful in conducting a sensitivity analysis; to see how the conclusions might vary as a function of what is assumed about the missing data
- Not necessarily a good idea to rely on a single MNAR model, because the assumptions about the missing data are impossible to assess with the observed data
- One should use MNAR models sensibly, possibly examining several types of such models for a given dataset

Mixed-effects selection models

These models have also been called

- random-coefficient selection models (Little, 95)
- random-effects-dependent models (Hogan & Laird, 97)
- shared parameter models (Wu & Carroll, 88; Ten Have *et al.*, 98)

- One specifies both a model for the longitudinal outcome and a model for the dropout (or missingness)
- Both models depend on random subject effects, most or all of which are shared by both models

Longitudinal model - ordinary MRM

Let $f_y(\mathbf{y}_i | \mathbf{v})$ represent the conditional model for the longitudinal outcome \mathbf{y}_i given the random subject effects \mathbf{v}

$$\mathbf{y}_i = \mathbf{X}_i\boldsymbol{\beta} + \mathbf{Z}_i\mathbf{v}_i + \boldsymbol{\varepsilon}_i$$

Dropout model - grouped/discrete time survival analysis

Let $f_D(D_i | \mathbf{v})$ represent the conditional model for time to dropout given the same subject random effects

$$\log(-\log(1 - P(D_i = j | D_i \geq j))) = \mathbf{W}_i\boldsymbol{\alpha} + \mathbf{v}_i\boldsymbol{\alpha}^*$$

\mathbf{W}_i includes dropout predictors, some or all may be in \mathbf{X}_i

To the extent that $\boldsymbol{\alpha}^*$ are nonzero, this is a nonignorable model because missingness, here characterized simply as dropout time, is dependent on both \mathbf{y}_i^O and \mathbf{y}_i^M (via \mathbf{v}_i)

Example - Schizophrenia study

- subjects measured at baseline & weekly for up to 6 wks
- main measurement weeks were 0 (baseline), 1, 3, and 6
- some subjects also measured at weeks 2, 4, and 5
- some intermittent missingness; dropout more common missingness pattern
- 102 of 437 subjects did not complete the trial

Crosstab of treatment (denoted **Drug**) by last wave (denoted **Maxweek**)

	Maxweek						
Drug	1	2	3	4	5	6	Total
placebo	13	5	16	2	2	70	108
	(.12)	(.05)	(.15)	(.02)	(.02)	(.65)	
drug	24	5	26	3	6	265	329
	(.07)	(.02)	(.08)	(.01)	(.02)	(.81)	

⇒ dropout is more common among the placebo group

Pearson χ^2 test yields $p < .025$;

Mantel-Haenszel χ^2 test for trend yields $p < .0013$

Mixed-effects selection model - Schiz study

Longitudinal model:

$$\text{IMPS79}_{ij} = \beta_0 + \beta_1 \text{Drug}_i + \beta_2 \text{SWeek}_j + \beta_3 (\text{Drug}_i \times \text{SWeek}_j) + v_{0i} + v_{1i} \text{SWeek}_j + \varepsilon_{ij}$$

or, after orthogonalizing the random effects

$$\begin{aligned} \text{IMPS79}_{ij} = & \beta_0 + \beta_1 \text{Drug}_i + \beta_2 \text{SWeek}_j + \beta_3 (\text{Drug}_i \times \text{SWeek}_j) \\ & + (\sigma_{v_0} + (\sigma_{v_{01}}/\sigma_{v_0}) \text{SWeek}_j) \theta_{0i} + \left(\sqrt{\sigma_{v_1}^2 - \sigma_{v_{01}}^2/\sigma_{v_0}^2} \text{SWeek}_j \right) \theta_{1i} + \varepsilon_{ij} \end{aligned}$$

because $\mathbf{v}_i = \mathbf{S}\boldsymbol{\theta}_i$, where $\boldsymbol{\Sigma}_v = \mathbf{S}\mathbf{S}'$ (*i.e.*, Cholesky factorization)

$$\mathbf{S} = \begin{bmatrix} s_0 & 0 \\ s_{01} & s_1 \end{bmatrix} = \begin{bmatrix} \sigma_{v_0} & 0 \\ \sigma_{v_{01}}/\sigma_{v_0} & \sqrt{\sigma_{v_1}^2 - \sigma_{v_{01}}^2/\sigma_{v_0}^2} \end{bmatrix}$$

and so $\mathbf{Z}_i \mathbf{v}_i = \mathbf{Z}_i \mathbf{S}\boldsymbol{\theta}_i$

$$\begin{aligned} v_{0i} + v_{1i} \text{SWeek}_j &= (s_0 + s_{01} \text{SWeek}_j) \theta_{0i} + (s_1 \text{SWeek}_j) \theta_{1i} \\ &= (\sigma_{v_0} + (\sigma_{v_{01}}/\sigma_{v_0}) \text{SWeek}_j) \theta_{0i} + \left(\sqrt{\sigma_{v_1}^2 - \sigma_{v_{01}}^2/\sigma_{v_0}^2} \text{SWeek}_j \right) \theta_{1i} \end{aligned}$$

Dropout model:

$$\begin{aligned} \log(-\log(1 - P(D_i = j \mid D_i \geq j))) &= \alpha_{0j} + \alpha_1 \mathbf{Drug}_i + \alpha_2 \theta_{0i} + \alpha_3 \theta_{1i} \\ &\quad + \alpha_4 (\mathbf{Drug}_i \times \theta_{0i}) + \alpha_5 (\mathbf{Drug}_i \times \theta_{1i}) \end{aligned}$$

or as

$$\begin{aligned} P(D_i \leq j) &= 1 - \exp(-\exp(\alpha_{0j} + \alpha_1 \mathbf{Drug}_i + \alpha_2 \theta_{0i} + \alpha_3 \theta_{1i} \\ &\quad + \alpha_4 (\mathbf{Drug}_i \times \theta_{0i}) + \alpha_5 (\mathbf{Drug}_i \times \theta_{1i}))) \end{aligned}$$

- random effects are summaries of a person's observed *and unobserved* \mathbf{y} data
- this shared parameter model is a nonignorable model if $\alpha_2 = \alpha_3 = \alpha_4 = \alpha_5 = 0$ is rejected
- test of whether *a particular model of ignorability is reasonable vs a particular model of nonignorability* (i.e., not a general test of ignorability)

```

TITLE1 analysis of schizophrenic data with SAS ;
DATA one; INFILE 'c:\schizrep.dat'; INPUT id imps79 week drug sex ;

/* The coding for the variables is as follows:
id = subject id number
imps79 = overall severity (1=normal, ..., 7=most extremely ill)
week = 0,1,2,3,4,5,6 (most of the obs. are at weeks 0,1,3, and 6)
drug 0=placebo 1=drug (chlorpromazine, fluphenazine, or thioridazine)
sex 0=female 1=male    */

/* compute the square root of week to linearize relationship */
sweek = SQRT(week);

/* calculate the maximum value of WEEK for each subject
and get drug in this aggregated dataset too */
PROC MEANS NOPRINT; CLASS id; VAR week drug;
OUTPUT OUT=two MAX(week drug)=maxweek drug;
RUN;

/* setting up IMPS79 across time and MAXWEEK as one outcome vector */
DATA daty; SET one; outcome = imps79; ind = 0;
DATA datr; SET two; outcome = maxweek; ind = 1; IF id NE .;
DATA all; SET daty datr; BY id;

```

- uppercase represents specific SAS syntax; lowercase represents user-defined
- the SAS dataset **all** includes the $(n_i + 1) \times 1$ outcome vector \mathbf{y}_i^* , named **outcome**, which contains \mathbf{y}_i as its first n_i elements and D_i as its final element
- **ind** with values of 0 or 1, is also defined; this variable will be used to distinguish between the \mathbf{y}_i and D_i elements

```

PROC NLMIXED DATA=all;
PARMS b0=6 b1=0 b2=-1 b3=-1 sde=1 v0=1 v01=0 v1=.5
a1=0 a2=.5 a3=.2 a4=.1 a5=.1 i1=-1 i2=-.7 i3=-.5 i4=0 i5=.2;
IF (ind = 0) THEN
DO;
  z = (outcome - (b0 + b1*drug + b2*sweek + b3*drug*sweek
    + (v0 + v01*sweek/v0)*u1
    + SQRT(v1*v1 - (v01*v01)/(v0*v0))*sweek*u2));
  p = (1 / SQRT(2*3.14159*sde*sde)) * EXP(-.5 * (z*z) / (sde*sde));
END;
IF (ind = 1) THEN
DO;
  z = a1*drug + a2*u1 + a3*u2 + a4*u1*drug + a5*u2*drug;
  IF (outcome=1) THEN
    p = 1 - EXP(0 - EXP(i1+z));
  ELSE IF (outcome=2) THEN
    p = (1 - EXP(0 - EXP(i2+z))) - (1 - EXP(0 - EXP(i1+z)));
  ELSE IF (outcome=3) THEN
    p = (1 - EXP(0 - EXP(i3+z))) - (1 - EXP(0 - EXP(i2+z)));
  ELSE IF (outcome=4) THEN
    p = (1 - EXP(0 - EXP(i4+z))) - (1 - EXP(0 - EXP(i3+z)));
  ELSE IF (outcome=5) THEN
    p = (1 - EXP(0 - EXP(i5+z))) - (1 - EXP(0 - EXP(i4+z)));
  ELSE IF (outcome=6) THEN
    p = 1 - (1 - EXP(0 - EXP(i5+z)));
END;
IF (p > 1e-8) THEN ll = LOG(p);
else ll = -1e100;
MODEL outcome ~ GENERAL(ll);
RANDOM u1 u2 ~ NORMAL([0,0], [1,0,1]) SUBJECT=id;
RUN;

```

Separate and shared parameter models

parameter	Separate (deviance = 5380.2)			Shared (deviance = 5350.1)		
	ML est	std error	<i>p</i> -value	ML est	std error	<i>p</i> -value
<u>Outcome</u>						
intercept β_0	5.348	.088	.0001	5.320	.088	.0001
Drug β_1	.046	.101	.65	.088	.102	.87
SWeek β_2	-.336	.068	.0001	-.272	.073	.0002
Drug \times Sweek β_3	-.641	.078	.0001	-.737	.083	.0001
<u>Dropout</u>						
Drug α_1	-.693	.205	.0008	-.703	.301	.02
Random intercept α_2				.447	.333	.18
Random slope α_3				.891	.467	.06
Drug \times intercept α_4				-.592	.398	.14
Drug \times slope α_5				-1.638	.536	.003

- the separate parameter model yields identical results as running these two models, one for \mathbf{y}_i and one for D_i , separately
- shared parameter model fits better, $\chi_4^2 = 30.1, p < .0001$
- for longitudinal component, conclusions are same as MAR model
- marginally significant slope: for the placebo group there is a tendency to dropout as the slope increases
- significant negative **Drug** \times slope: the slope effect is opposite for the drug group; drug patients with more negative slopes (*i.e.*, greater improvement) are more likely to drop out

Pattern-mixture models for missing data

Little (1993, 1994, 1995); Hedeker & Gibbons (1997)

- divide subjects into groups depending on their missing data pattern
- the missing data pattern is a between-subjects variable to be used in longitudinal data analysis
- method of analysis must allow subjects to have incomplete data across time

Suppose three timepoints, there are eight (2^3) possible missing data patterns:

pattern group	time1	time2	time3
1	O	O	O
2	O	O	M
3	O	M	O
4	M	O	O
5	M	M	O
6	O	M	M
7	M	O	M
8	M	M	M

where, O=observed and M=missing. Since MMM provides no data, it is ignored in the analysis.

Representing patterns with dummy-coded variables

pattern	$D1$	$D2$	$D3$	$D4$	$D5$	$D6$
OOO	0	0	0	0	0	0
OOM	1	0	0	0	0	0
OMO	0	1	0	0	0	0
MOO	0	0	1	0	0	0
MMO	0	0	0	1	0	0
OMM	0	0	0	0	1	0
MOM	0	0	0	0	0	1

- these dummy-coded variables represent deviations from pattern OOO
- Other coding schemes can be used (“effect” or “sequential” coding)
- these variables are used as main effects and interactions

Mixed-effects pattern mixture model: Schiz data

augment the basic MRM of IMPS79 over time:

$$\text{IMPS79}_{ij} = \beta_0 + \beta_1 \text{Drug}_i + \beta_2 \text{SWeek}_j + \beta_3 (\text{Drug}_i \times \text{SWeek}_j) + \nu_{0i} + \nu_{1i} \text{SWeek}_j + \varepsilon_{ij} ,$$

with variables based on the missing data patterns

e.g., completers (N = 335) vs non-completers (N = 102)

Drop = 0 or 1 for those that did not or did dropout from the trial (*i.e.*, were not measured at the final study timepoint)

$$\begin{aligned} \text{IMPS79}_{ij} = & \beta_0 + \beta_1 \text{Drug}_i + \beta_2 \text{SWeek}_j + \beta_3 (\text{Drug}_i \times \text{SWeek}_j) \\ & + \beta_0^D \text{Drop}_i + \beta_1^D (\text{Drop}_i \times \text{Drug}_i) + \beta_2^D (\text{Drop}_i \times \text{Sweek}_j) \\ & + \beta_3^D (\text{Drop}_i \times \text{Drug}_i \times \text{Sweek}_j) + \nu_{0i} + \nu_{1i} \text{Sweek}_j + \varepsilon_{ij} \end{aligned}$$

- β_0 , β_1 , β_2 , and β_3 are for the completer subsample
- β_0^D , β_1^D , β_2^D , and β_3^D how dropouts differ from completers
- three-way interaction is of particular interest - indicates how the drug by time interaction varies with study completion

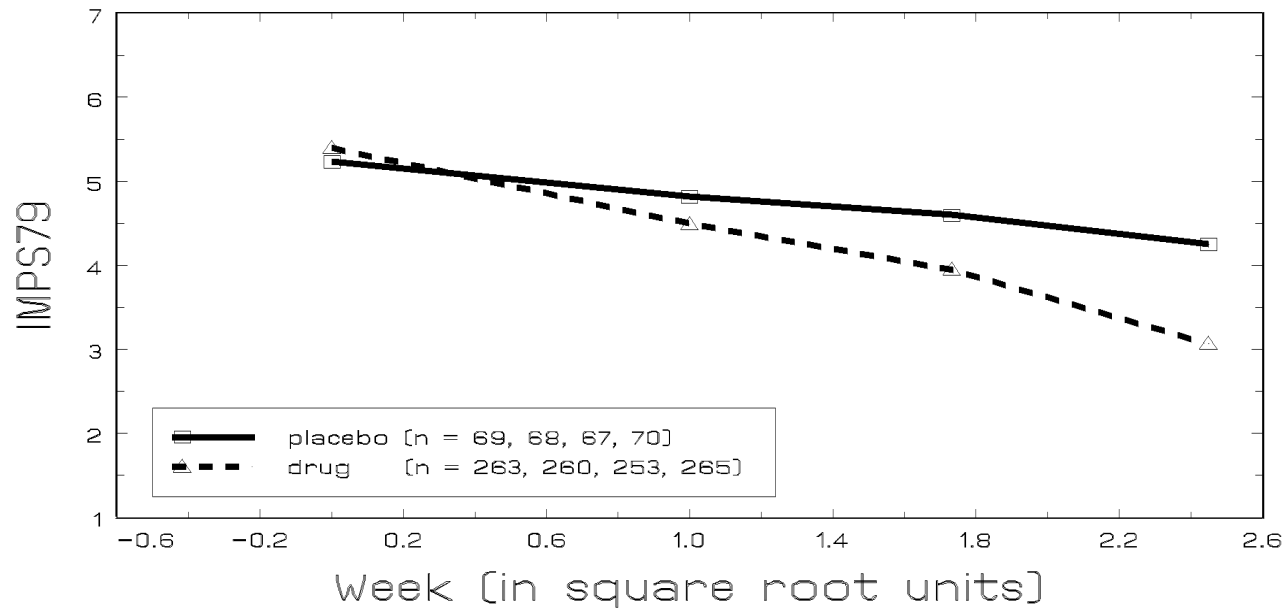
Classification of Subjects based on missing-data

$$\text{Drop}_i = \begin{cases} 0 & \text{subject measured at week 6 (last timepoint)} \\ 1 & \text{subject missing at week 6 (last timepoint)} \end{cases}$$

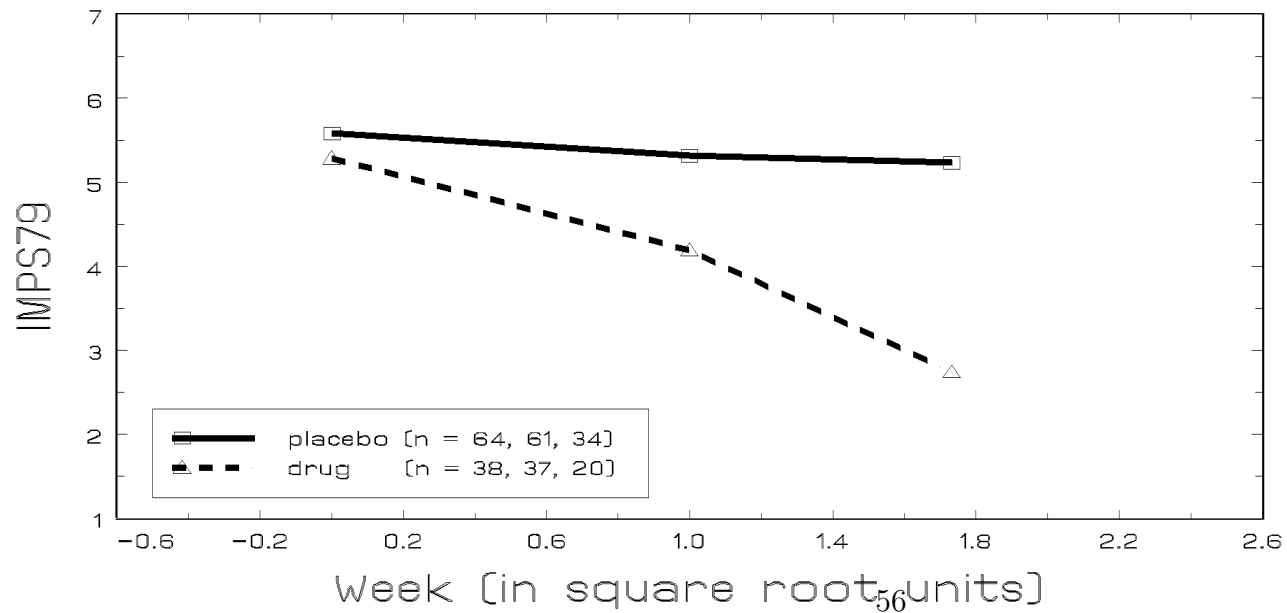
Drug group	Drop group		total
	completer	dropout	
placebo	70 (.65)	38 (.35)	108
drug	265 (.81)	64 (.19)	329
total	335	102	437

- Dropout not independent of Drug $\chi_1^2 = 11.25, p < .001$
- Is dropout related to severity of illness?
- Does dropout moderate the influence of other variables' effects on severity of illness?

Mean IMPS79 across Time by Group
Completers



Mean IMPS79 across Time by Group
Dropouts



Less Restrictive Pattern Mixture Model

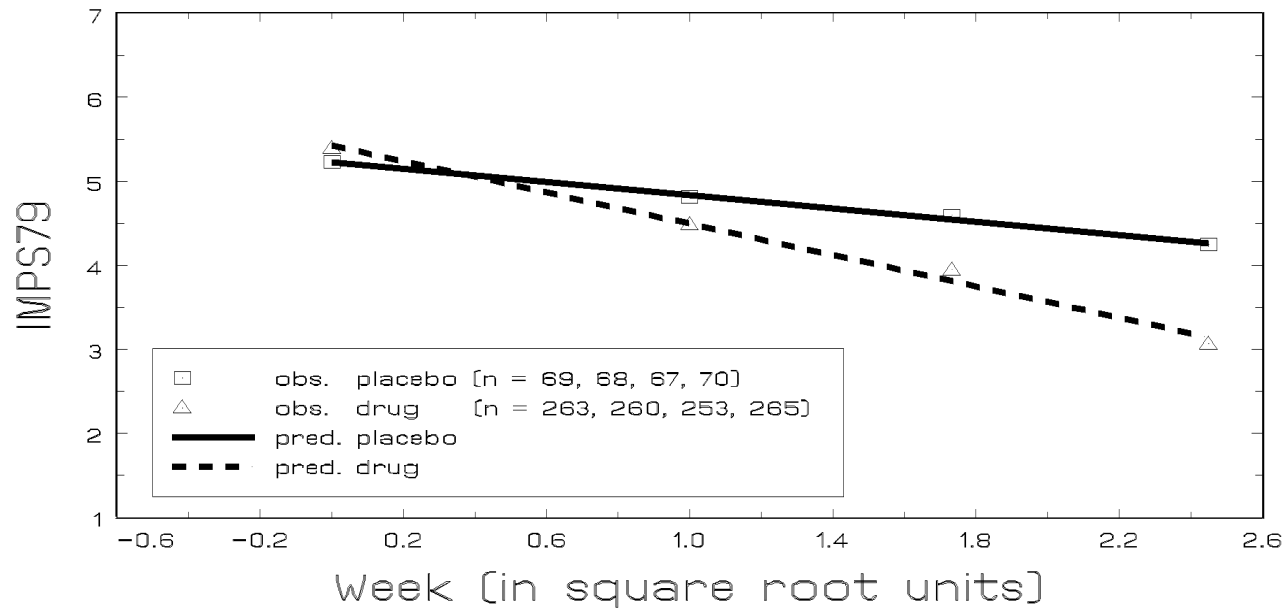
- use week of dropout variable D_i in forming missing data patterns
- six missing data patterns: five dropout weeks and completers
- Let $D_m = D_1, \dots, D_5$ denote dummy-variables which contrast each dropout pattern to the completers

$$\begin{aligned} \text{IMPS79}_{ij} = & \beta_0 + \beta_1 \text{Drug}_i + \beta_2 \text{SWeek}_j + \beta_3 (\text{Drug}_i \times \text{SWeek}_j) \\ & + \sum_{m=1}^5 \beta_0^m D_m + \beta_1^m (D_m \times \text{Drug}_i) + \beta_2^m (D_m \times \text{Sweek}_j) \\ & + \beta_3^m (D_m \times \text{Drug}_i \times \text{Sweek}_j) \\ & + v_{0i} + v_{1i} \text{SWeek}_j + \varepsilon_{ij} \end{aligned}$$

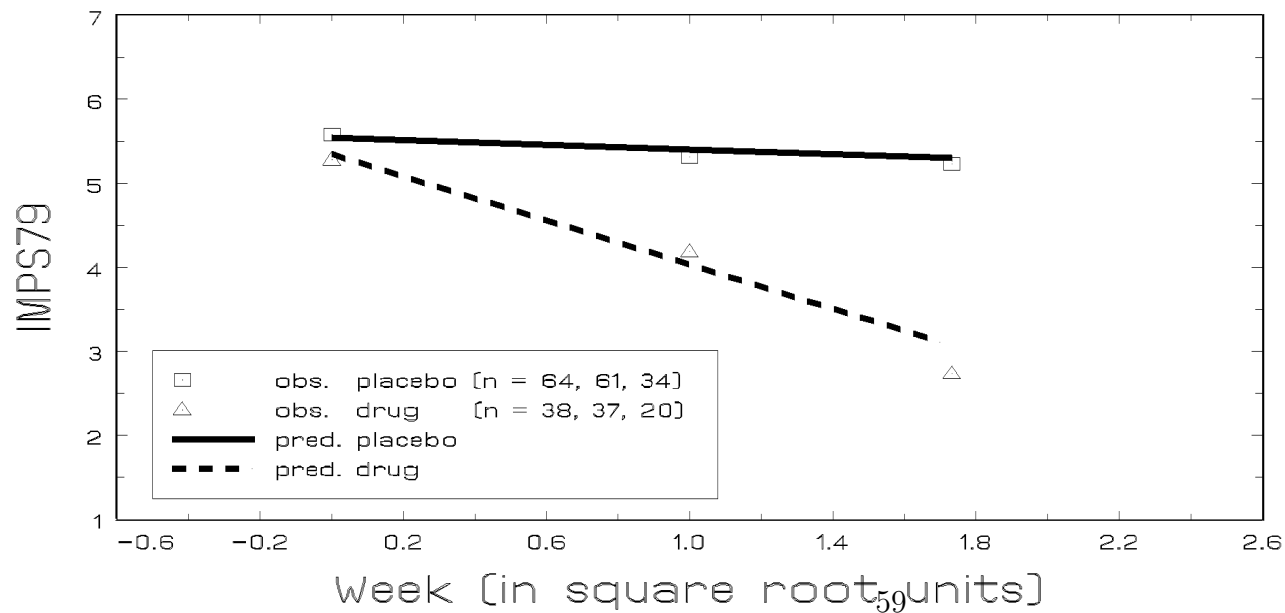
- $\beta_0, \beta_1, \beta_2,$ and β_3 are for completers
- $\beta_0^m, \beta_1^m, \beta_2^m,$ and β_3^m indicate how dropout group m differs from completers
- the β_3^m parameters are of great interest

parameter	est	se	$p <$	est	se	$p <$	est	se	$p <$
Int β_0	5.348	.088	.0001	5.221	.108	.0001	5.221	.107	.0001
Drug β_1	.046	.101	.65	.202	.121	.096	.202	.120	.094
SWeek β_2	-.336	.068	.0001	-.393	.076	.0001	-.393	.075	.0001
Drug \times SWeek β_3	-.641	.078	.0001	-.539	.086	.0001	-.539	.085	.0001
<u>Dropout</u>			<u>Drop = 1 (N = 102)</u>			<u>D = 1 (N = 37)</u>			
Int β_0^1				.320	.186	.086	.471	.288	.102
Drug β_1^1				-.399	.227	.079	-.456	.353	.20
SWeek β_2^1				.252	.159	.115	.240	.334	.47
Drug \times SWeek β_3^1				-.635	.196	.002	-.412	.412	.32
							<u>D = 2 (N = 10)</u>		
Int β_0^2							.524	.437	.23
Drug β_1^2							-.703	.613	.25
SWeek β_2^2							.338	.398	.40
Drug \times SWeek β_3^2							-.735	.562	.19
							<u>D = 3 (N = 42)</u>		
Int β_0^3							.047	.256	.85
Drug β_1^3							-.198	.318	.53
SWeek β_2^3							.377	.208	.07
Drug \times SWeek β_3^3							-.835	.261	.002
							<u>D = 4 (N = 5)</u>		
Int β_0^4							.801	.653	.22
Drug β_1^4							-.237	.841	.78
SWeek β_2^4							-.101	.485	.84
Drug \times SWeek β_3^4							-1.210	.625	.054
							<u>D = 5 (N = 8)</u>		
Int β_0^5							.337	.645	.60
Drug β_1^5							-.842	.746	.26
SWeek β_2^5							-.157	.466	.74
Drug \times SWeek β_3^5							.231	.538	.67
Deviance	4649.0			4623.3			4607.8		

Mean IMPS79 across Time by Group
Completers



Mean IMPS79 across Time by Group
Dropouts



Pattern-mixture averaged results (Little, 1995)

- Obtained averaging over missing-data patterns
 - *e.g.*, completers and dropouts
- Uses sample proportions as estimates of missing-data pattern proportions
- Can use Delta Method to obtain standard errors
 - uncertainty in model estimates
 - uncertainty in using sample proportions as estimates
- Depends on “model” for missing-data patterns
 - *e.g.*, completer versus dropout status varies by tx

Completer

placebo *70/108*

drug *265/329*

335/437

Dropout

placebo *38/108*

drug *64/329*

102/437

Pattern-mixture averaged results

$$\hat{\beta} = \hat{\pi}_c \hat{\beta}_c + \hat{\pi}_d \hat{\beta}_d$$

or

$$\hat{\beta} = (1 - \hat{\pi}_d) \hat{\beta}_c + \hat{\pi}_d \hat{\beta}_d = \hat{\beta}_c + \hat{\pi}_d (\hat{\beta}_d - \hat{\beta}_c) = \hat{\beta}_c + \hat{\pi}_d \hat{\beta}_\Delta$$

note:

- $\hat{\beta}_c$ correspond to the coefficients in the current model formulation not involving dropout (*i.e.*, intercept, drug, time, drug by time)
- $(\hat{\beta}_d - \hat{\beta}_c) = \hat{\beta}_\Delta$ correspond to the dropout-related coefficients in the current model formulation (*i.e.*, dropout, dropout by drug, dropout by time, dropout by drug by time)
- $\hat{\pi}_d$ is the sample proportion of dropouts

\Rightarrow averaged estimates are simple linear combinations of model estimates

Placebo Intercept

$$\frac{335}{437}(5.22) + \frac{102}{437}(5.22 + 0.32) = 5.22 + (.233)(0.32) = 5.30$$

Completers *Dropouts*

Placebo Time effect

$$\frac{335}{437}(-0.39) + \frac{102}{437}(-0.39 + 0.25) = -0.39 + (.233)(0.25) = -0.33$$

Completers *Dropouts*

Drug Intercept difference

$$\frac{335}{437}(0.20) + \frac{102}{437}(0.20 - 0.40) = 0.20 + (.233)(-0.40) = 0.11$$

Completers *Dropouts*

Drug Time difference

$$\frac{335}{437}(-0.54) + \frac{102}{437}(-0.54 - 0.64) = -0.54 + (.233)(-0.64) = -0.69$$

Completers *Dropouts*

Delta Method for estimating asymptotic variance of averaged estimates

$$\hat{\beta} = \hat{\beta}_c + \hat{\pi}_d \hat{\beta}_\Delta$$

$$\hat{V}(\hat{\beta}) = \begin{bmatrix} \frac{\partial \hat{\beta}}{\partial \hat{\beta}_c} & \frac{\partial \hat{\beta}}{\partial \hat{\beta}_\Delta} & \frac{\partial \hat{\beta}}{\partial \hat{\pi}_d} \end{bmatrix} \begin{bmatrix} \hat{V}(\hat{\beta}_c) & \hat{C}(\hat{\beta}_c, \hat{\beta}_\Delta) & 0 \\ \hat{C}(\hat{\beta}_c, \hat{\beta}_\Delta) & \hat{V}(\hat{\beta}_\Delta) & 0 \\ 0 & 0 & \hat{V}(\hat{\pi}_d) \end{bmatrix} \begin{bmatrix} \frac{\partial \hat{\beta}}{\partial \hat{\beta}_c} \\ \frac{\partial \hat{\beta}}{\partial \hat{\beta}_\Delta} \\ \frac{\partial \hat{\beta}}{\partial \hat{\pi}_d} \end{bmatrix}$$

where

$$\frac{\partial \hat{\beta}}{\partial \hat{\beta}_c} = 1 \quad \frac{\partial \hat{\beta}}{\partial \hat{\beta}_\Delta} = \hat{\pi}_d \quad \frac{\partial \hat{\beta}}{\partial \hat{\pi}_d} = \hat{\beta}_\Delta$$

Thus,

$$\hat{V}(\hat{\beta}) = \hat{V}(\hat{\beta}_c) + \hat{\pi}_d^2 \hat{V}(\hat{\beta}_\Delta) + 2\hat{\pi}_d \hat{C}(\hat{\beta}_c, \hat{\beta}_\Delta) + \hat{\beta}_\Delta^2 \hat{V}(\hat{\pi}_d)$$

Note, under marginal model for completion (*i.e.*, = binomial(π_c))

$$\begin{aligned} \hat{V}(\hat{\pi}_d) &= \hat{\pi}_d(1 - \hat{\pi}_d)/N \\ &= (n_d/N)(n_c/N)/N = \frac{n_d n_c}{N^3} \end{aligned}$$

Standard Errors for Averaged Estimates

$$\begin{aligned}\hat{V}(\hat{\boldsymbol{\beta}}) &= \hat{V}(\hat{\boldsymbol{\beta}}_c) + \hat{\pi}_d^2 \hat{V}(\hat{\boldsymbol{\beta}}_\Delta) + 2\hat{\pi}_d \hat{C}(\hat{\boldsymbol{\beta}}_c, \hat{\boldsymbol{\beta}}_\Delta) + \hat{\boldsymbol{\beta}}_\Delta^2 \hat{V}(\hat{\pi}_d) \\ &= \hat{V}(\hat{\boldsymbol{\beta}})_F + \frac{n_d n_c}{N^3} \hat{\boldsymbol{\beta}}_\Delta^2\end{aligned}$$

where, $\hat{V}(\hat{\boldsymbol{\beta}})_F$ is the variance treating the sample proportions as known, *i.e.*, the square of the standard error one gets using

$$\hat{\boldsymbol{\beta}} = \hat{\boldsymbol{\beta}}_c + \hat{\pi}_d \hat{\boldsymbol{\beta}}_\Delta$$

and not taking into account the fact that π_d is estimated (*i.e.*, this is obtained using methods that yield linear combinations of estimates and their associated standard errors)

\Rightarrow simple augmentation of $\hat{V}(\hat{\boldsymbol{\beta}})_F$ to get correct standard errors

Calculation of $\hat{V}(\hat{\beta}) = \hat{V}(\hat{\beta})_F + \frac{n_d n_c}{N^3} \hat{\beta}_\Delta^2$

parameter	$\hat{\beta}$	$\hat{V}(\hat{\beta})_F$	$\hat{\beta}_\Delta$	Augment	$\hat{V}(\hat{\beta})$	SE
intercept	5.2958	$(.0898)^2 = .00806$.3203	.000042	.00810	.0900
time	-.3346	$(.0670)^2 = .00449$.2517	.000026	.00452	.0672
drug	.1086	$(.1029)^2 = .01059$	-.3987	.000065	.01066	.1032
drug \times time	-.6868	$(.0776)^2 = .00602$	-.6348	.000165	.00619	.0786

Augment = $\frac{n_d n_c}{N^3} \hat{\beta}_\Delta^2$, here, $\frac{n_d n_c}{N^3} = \frac{102 \times 335}{(437)^3} = .00040945$

Pattern-mixture averaged results - using drug-stratified proportions

Placebo Intercept

$$\frac{70}{108}(5.22) + \frac{38}{108}(5.22 + 0.32) = 5.22 + (.352)(0.32) = 5.33$$

Completers *Dropouts*

Placebo Time effect

$$\frac{70}{108}(-0.39) + \frac{38}{108}(-0.39 + 0.25) = -0.39 + (.352)(0.25) = -0.30$$

Completers *Dropouts*

Drug Intercept difference

$$\frac{265}{329}(0.20) + \frac{64}{329}(0.20 - 0.40) = 0.20 + (.195)(-0.40) = 0.12$$

Completers *Dropouts*

Drug Time difference

$$\frac{265}{329}(-0.54) + \frac{64}{329}(-0.54 - 0.64) = -0.54 + (.195)(-0.64) = -0.66$$

Completers *Dropouts*

Calculation of $\hat{V}(\hat{\beta}) = \hat{V}(\hat{\beta})_F + \frac{n_d n_c}{N^3} \hat{\beta}_\Delta^2$

parameter	$\hat{\beta}$	$\hat{V}(\hat{\beta})_F$	$\hat{\beta}_\Delta$	Augment	$\hat{V}(\hat{\beta})$	SE
intercept	5.3337	$(.0879)^2 = .00773$.3203	.000217	.00795	.0891
time	-.3048	$(.0698)^2 = .00487$.2517	.000134	.00500	.0707
drug	.1241	$(.1043)^2 = .01088$	-.3987	.000076	.01096	.1047
drug \times time	-.6621	$(.0772)^2 = .00596$	-.6348	.000192	.00615	.0784

Augment = $\frac{n_d n_c}{N^3} \hat{\beta}_\Delta^2$, where

$$\frac{n_d n_c}{N^3} = \frac{38 \times 70}{(108)^3} = .00211159 \text{ for placebo}$$

$$\frac{n_d n_c}{N^3} = \frac{64 \times 265}{(329)^3} = .00047625 \text{ for drug}$$

SAS MIXED code

```
FILENAME IN1 'C:SCHIZREP.DAT';
DATA ONE; INFILE IN1 ;
INPUT ID IMPS79 WEEK DRUG SEX ;

/* The coding for the variables is as follows:
ID = subject ID number
IMPS79 = overall severity (1=normal, ..., 7=among the most extremely ill)
WEEK = 0,1,2,3,4,5,6 (most of the obs. are at weeks 0,1,3, and 6)
DRUG 0=placebo 1=drug (chlorpromazine, fluphenazine, or thioridazine)
SEX 0=female 1=male */

/* compute the square root of week to linearize relationship */
SWEEK = SQRT(WEEK);

/* calculate the maximum value of WEEK for each subject
(suppress the printing of the output for this procedure) */
PROC MEANS NOPRINT; CLASS ID; VAR WEEK; OUTPUT OUT=TWO MAX=MAXWEEK;
RUN;

/* determine if a subject has data at WEEK 6
DROPOUT = 0 (for completers) or = 1 (for dropouts) */
DATA THREE; SET TWO;
DROPOUT=0;
IF MAXWEEK LT 6 THEN DROPOUT=1;
```

```

/* dataset with all subjects (adding the DROPOUT variable) */
DATA FOUR; MERGE ONE THREE; BY ID;

/* Random intercept and trend model */
PROC MIXED DATA=FOUR METHOD=ML COVTEST;
CLASS ID;
MODEL IMPS79 = SWEET DRUG SWEET*DRUG / SOLUTION;
RANDOM INTERCEPT SWEET /SUB=ID TYPE=UN G GCORR;
RUN;

/* Pattern-mixture random intercept and trend model
/* using marginal dropout proportion to estimate averaged results */
PROC MIXED DATA=FOUR METHOD=ML COVTEST;
CLASS ID;
MODEL IMPS79 = SWEET DRUG SWEET*DRUG DROPOUT DROPOUT*SWEET
                DROPOUT*DRUG DROPOUT*DRUG*SWEET / SOLUTION COVB;
RANDOM INTERCEPT SWEET /SUB=ID TYPE=UN G GCORR;
ESTIMATE 'AVG INT' intercept 1 sweek 0 drug 0 sweek*drug 0 dropout .2334
                dropout*sweek 0 dropout*drug 0 dropout*drug*sweek 0;
ESTIMATE 'AVG SWEET' intercept 0 sweek 1 drug 0 sweek*drug 0 dropout 0
                dropout*sweek .2334 dropout*drug 0 dropout*drug*sweek 0;
ESTIMATE 'AVG DRUG' intercept 0 sweek 0 drug 1 sweek*drug 0 dropout 0
                dropout*sweek 0 dropout*drug .2334 dropout*drug*sweek 0;
ESTIMATE 'AVG SWEET*DRUG' intercept 0 sweek 0 drug 0 sweek*drug 1 dropout 0
                dropout*sweek 0 dropout*drug 0 dropout*drug*sweek .2334;
RUN;

```

```

/* Pattern-mixture random intercept and trend model
/* using drug-specific dropout proportions to estimate averaged results */

PROC MIXED DATA=FOUR METHOD=ML COVTEST;
CLASS ID;
MODEL IMPS79 = SWEET DRUG SWEET*DRUG DROPOUT DROPOUT*SWEET
              DROPOUT*DRUG DROPOUT*DRUG*SWEET / SOLUTION COVB;
RANDOM INTERCEPT SWEET /SUB=ID TYPE=UN G GCORR;
ESTIMATE 'AVG INT' intercept 1 sweek 0 drug 0 sweek*drug 0 dropout .35185
        dropout*sweek 0 dropout*drug 0 dropout*drug*sweek 0;
ESTIMATE 'AVG SWEET' intercept 0 sweek 1 drug 0 sweek*drug 0 dropout 0
        dropout*sweek .35185 dropout*drug 0 dropout*drug*sweek 0;
ESTIMATE 'AVG DRUG' intercept 0 sweek 0 drug 1 sweek*drug 0 dropout 0
        dropout*sweek 0 dropout*drug .19453 dropout*drug*sweek 0;
ESTIMATE 'AVG SWEET*DRUG' intercept 0 sweek 0 drug 0 sweek*drug 1 dropout 0
        dropout*sweek 0 dropout*drug 0 dropout*drug*sweek .19453;

RUN;

```

NIMH Schizophrenia Study: Severity across Time
 ML Estimates (se) *random intercept and slope models*

	<i>Completers</i> <i>N = 335</i>	<i>All cases</i> <i>N = 437</i>	<i>Shared</i> <i>Parameter</i> <i>N = 437</i>	<i>Pattern</i> <i>Mixture</i> <i>N = 437</i>
intercept	5.221 (.109)	5.348 (.088)	5.320 (.088)	5.334 (.089)
Drug (0=P; 1=D)	0.202 (.123)	0.046 (.101)	0.088 (.102)	0.124 (.105)
Time (sqrt wk)	-0.393 (.073)	-0.336 (.068)	-0.272 (.073)	-0.305 (.071)
Drug by Time	-0.539 (.083)	-0.641 (.078)	-0.737 (.083)	-0.662 (.078)

Conclusions

- Mixed-effects regression models (MRMs) useful for incomplete longitudinal data
 - can handle subjects measured incompletely or at different timepoints
 - missing data assumed MAR
 - * dependent on covariates *and*
 - * available data on dependent variable

- Mixed-effects selection (*i.e.*, shared parameter) and pattern-mixture models augment MRM

Selection

- missingness in terms of important covariates
- missingness in terms of (shared) random subject effects

Pattern-mixture

- adds missing-data pattern as between-subjects factor
- assesses degree to which “missingness” influences outcomes
- assesses degree to which “missingness” interacts with model terms (*i.e.*, intervention group, intervention group by time)

⇒ Does not invent data, maximizes information obtained from *available* data