Modeling Effectiveness Trial Data with Missing Observations

Don Hedeker
University of Illinois at Chicago

hedeker@uic.edu
www.uic.edu/∼hedeker/long.html

Treatment Effectiveness Trials

- Two or more groups compared, usually longitudinally
- Follow-up time can be long (*e.g.*, two years)
- Missing data and subject dropout are inevitable
- Statistical methods for incomplete data across time
  - Mixed-effects regression models (MRM)
- Sensitivity analysis regarding missing data assumptions
  - Mixed-effects selection models
  - Mixed-effects pattern mixture models

⇒ Examine how results vary with missing data assumptions

Example: Treatment-Related Change Across Time

Data from the NIMH Schizophrenia collaborative study on treatment related changes in overall severity. IMPS item 79, *Severity of Illness*, was scored as:

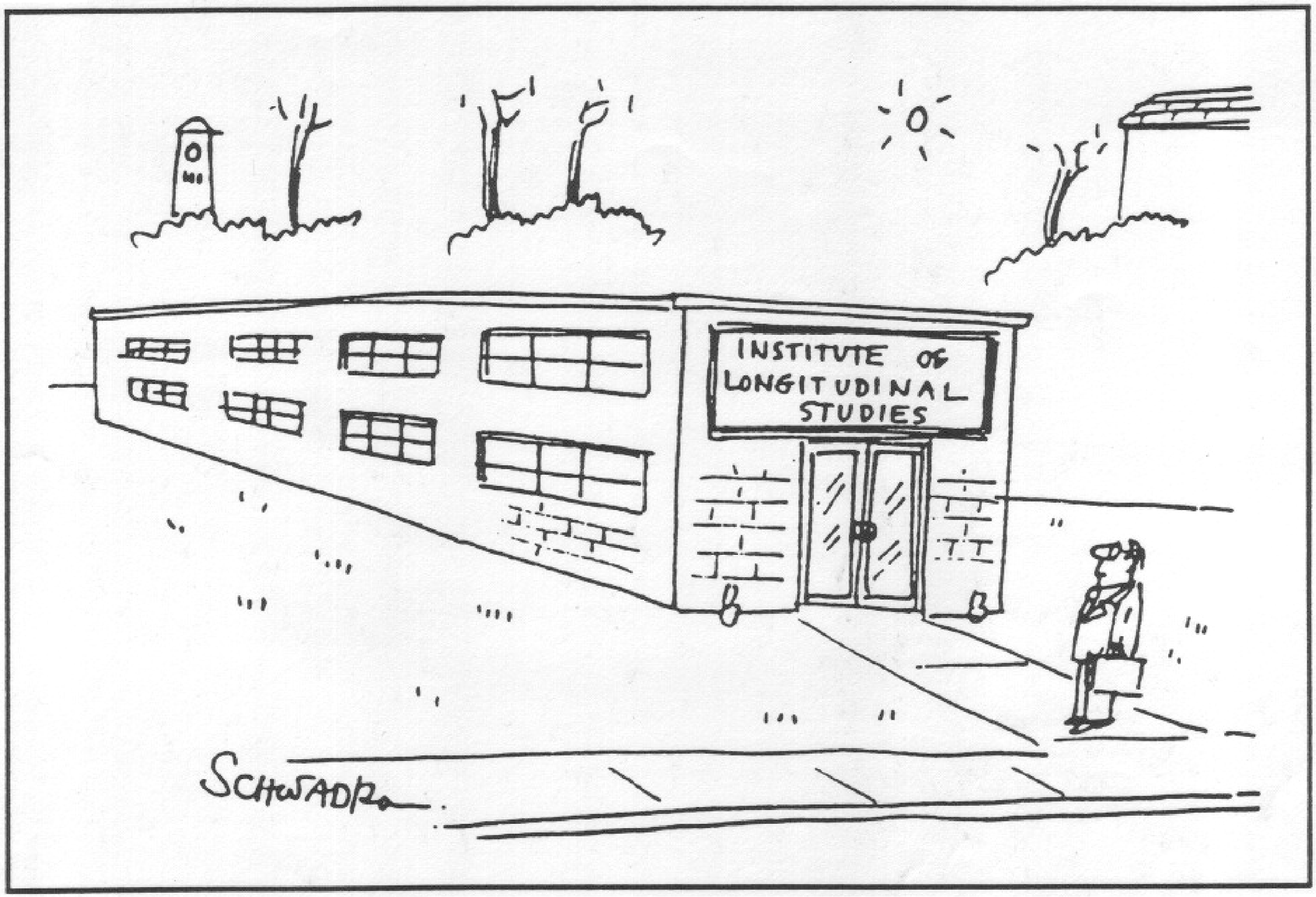
1	=	normal
2	=	borderline mentally ill
3	=	mildly ill
4	=	moderately ill
5	=	markedly ill
6	=	severely ill
7	=	among the most extremely ill

	Sample size at Week							
Group	0	1	2	3	4	5	6	*completers*
PLC (n=108)	107	105	5	87	2	2	70	65%
DRUG (n=329)	327	321	9	287	9	7	265	81%

Drug = Chlorpromazine, Fluphenazine, or Thioridazine

Mixed-effects Regression Models (MRM) For Longitudinal Clinical Trials Data

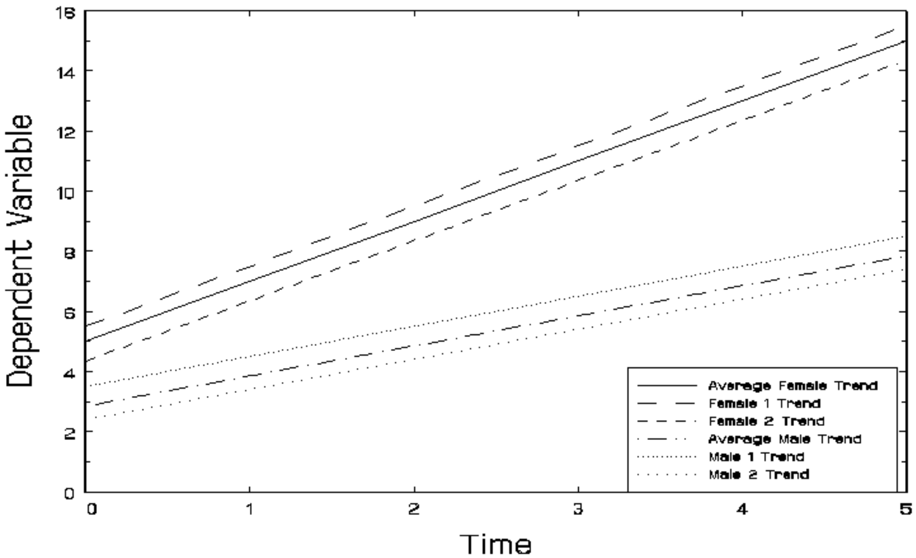
1. MRM explicitly models individual change across time
2. MRM more flexible in terms of repeated measures
   (a) need not have same number of obs per subject
   (b) time can be continuous, rather than a fixed set of points
3. Flexible specification of the covariance structure among repeated measures $\Rightarrow$ 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


INSTITUTE OF
LONGITUDINAL
STUDIES
SCHWADRON


INSTITUTE of
SOLID GEOMETRY
Institute of
Plane
Geometry

Random-intercepts Model

each subject is parallel to their group trend



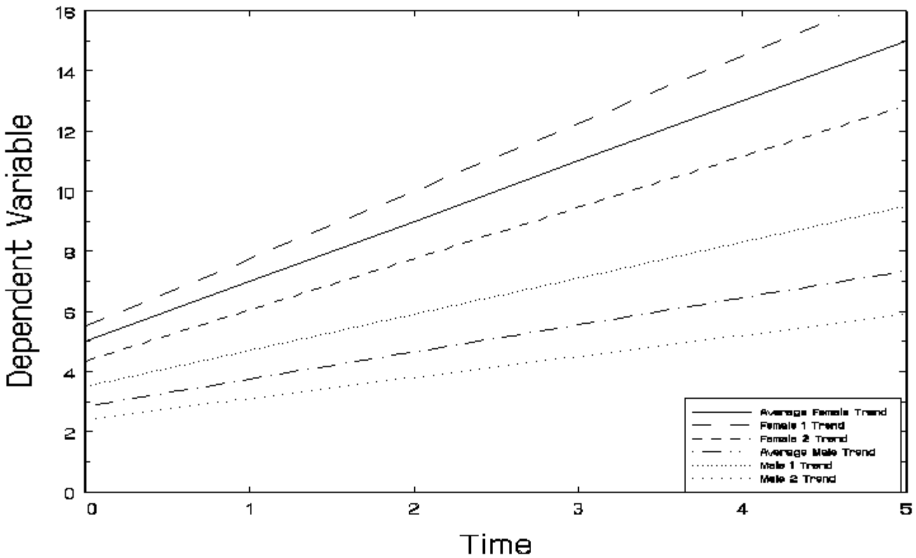
$$y = Time + Grp + (Grp \times Time) + Subj + Error$$

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

$$\upsilon_{0i} \sim \mathcal{N}(0, \sigma_\upsilon^2) \quad \varepsilon_{ij} \sim \mathcal{N}(0, \sigma^2)$$

Random Intercepts and Trend Model

subjects deviate in terms of both intercept & slope



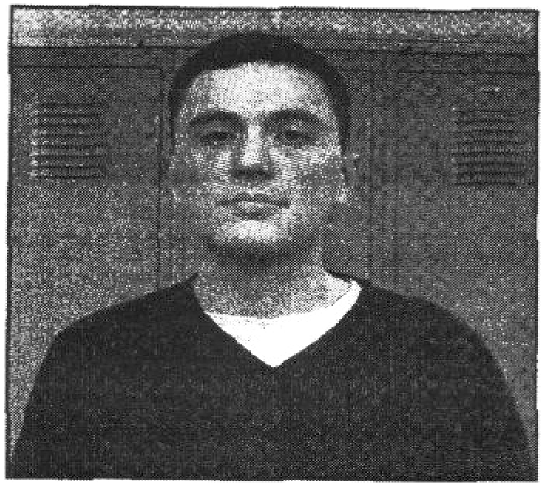
$$y \; = \; Time + Grp + (G \times T) + Subj + (S \times T) + Error$$

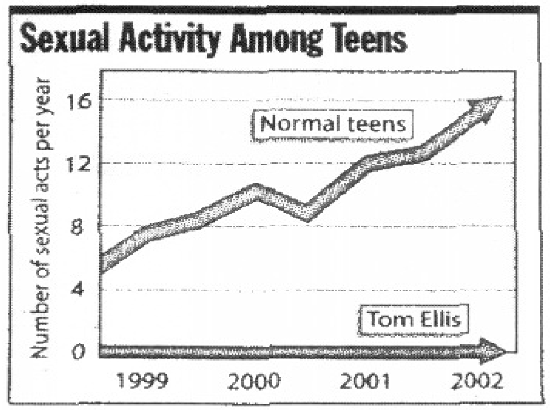
$$y_{ij} \; = \; \beta_0 + \beta_1 T_{ij} + \beta_2 G_i + \beta_3 (G_i \times T_{ij}) + \upsilon_{0i} + \upsilon_{1i} T_{ij} + \varepsilon_{ij}$$

$$\begin{bmatrix} \upsilon_{0i} \\ \upsilon_{1i} \end{bmatrix} \sim \mathcal{N} \left\{ \begin{bmatrix} 0 \\ 0 \end{bmatrix}, \begin{bmatrix} \sigma^2_{\upsilon_0} & \sigma_{\upsilon_0 \upsilon_1} \\ \sigma_{\upsilon_0 \upsilon_1} & \sigma^2_{\upsilon_1} \end{bmatrix} \right\} \qquad \varepsilon_{ij} \sim \mathcal{N}(0, \sigma^2)$$

Example of random intercept and trend model

Rise In Teen Sexual Activity Comes As Surprise To Area Teen





Left: Tom Ellis, who was surprised by statistics showing that other teens are having sex (above).

SALEM, OR—The Alan Guttmacher Institute released a report Friday that showed a dramatic increase in teen sexual activity, a finding that surprised policy-makers, public-health professionals, and 17-year-old Tom Ellis.

"So, more teens are having sex, are they?" Ellis asked Monday. "Well, I'm not sure where those guys got all their data, but it sure wasn't from me."

Ellis, a senior this fall at Sprague High School in Salem, learned of the trend while watching television at home Saturday, as he does most weekend nights. A *20/20* story titled "The Teen Sex Epidemic"

see TEEN page 8

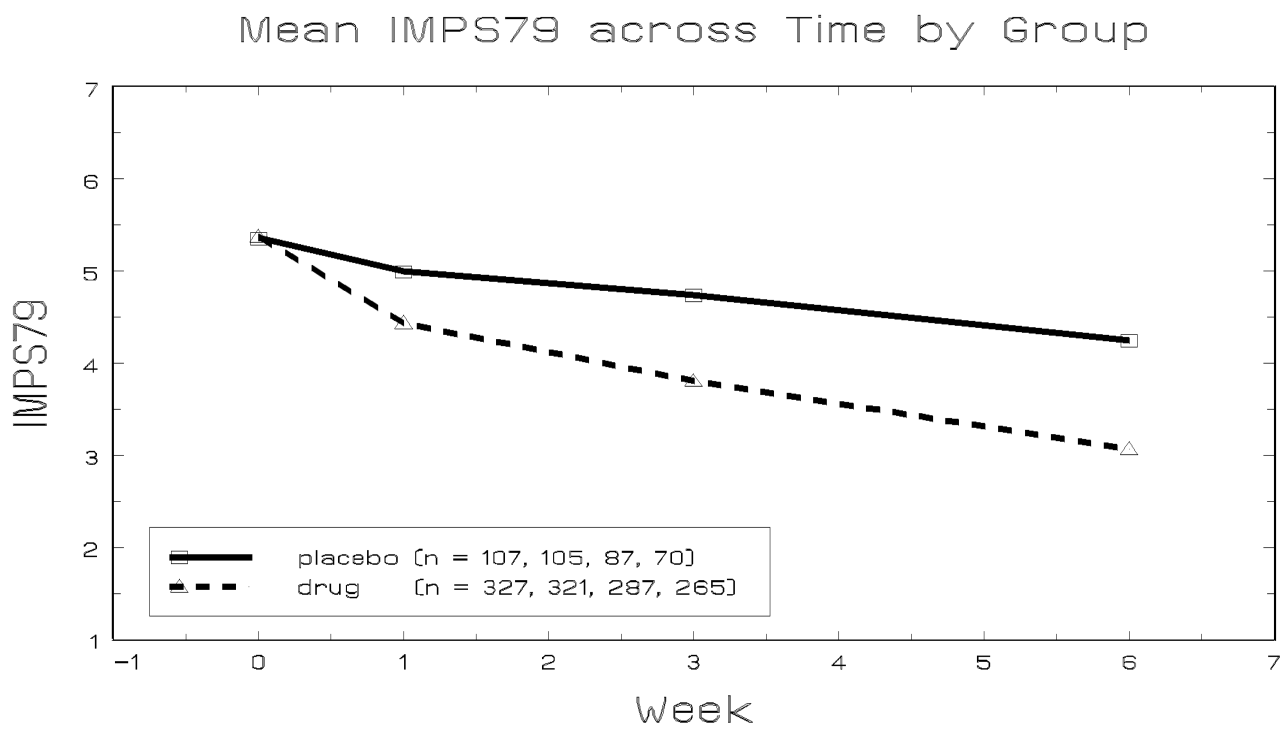
Descriptive Statistics - NIMH Schizophrenia study

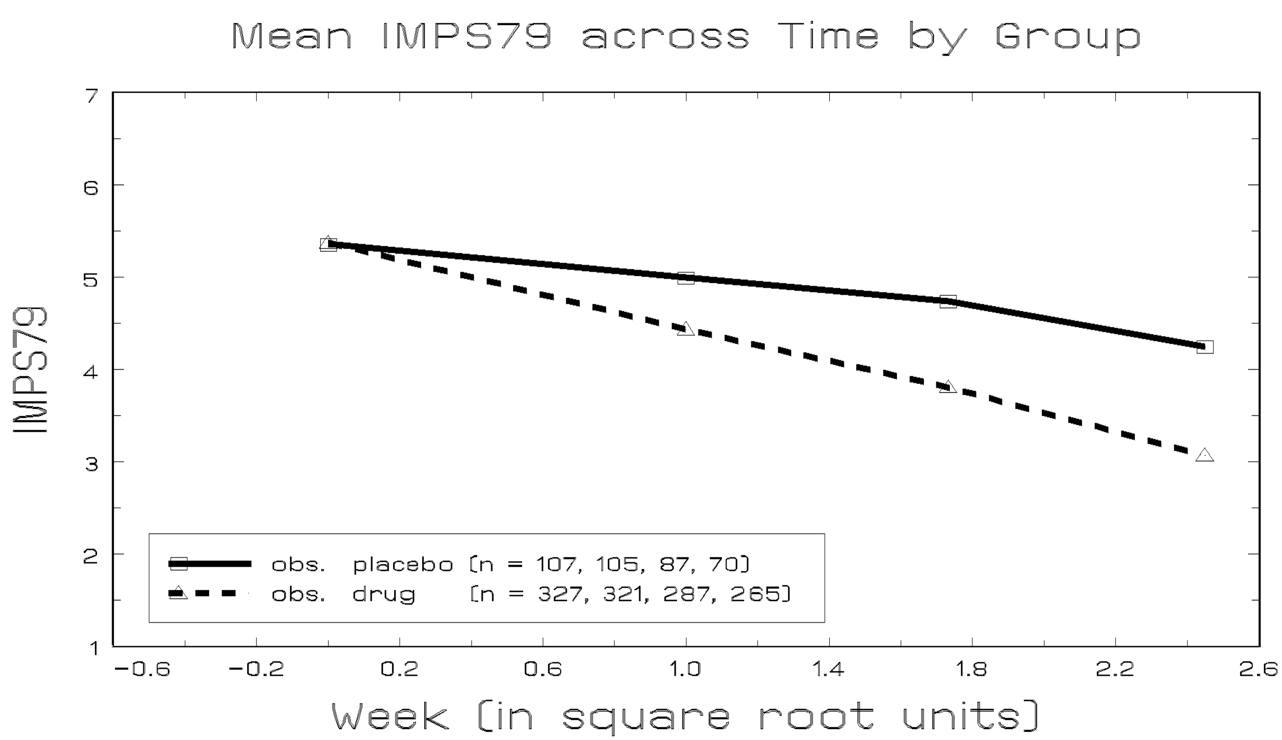
Observed IMPS79 Means, n, and sd

	week 0	week 1	week 3	week 6
placebo	5.35	4.99	4.74	4.25
n	107	105	87	70
drug	5.37	4.43	3.80	3.06
n	327	321	287	265
pooled sd	.87	1.23	1.44	1.48

Correlations: $\mathbf{n = 313}$ and *321* $\leq n \leq$ *424*

	week 0	week 1	week 3	week 6
week 0	1.0	*.41*	*.25*	*.15*
week 1	**.43**	1.0	*.67*	*.47*
week 3	**.29**	**.67**	1.0	*.67*
week 6	**.14**	**.47**	**.68**	1.0


Mean IMPS79 across Time by Group
IMPS79
Week
placebo (n = 107, 105, 87, 70)
drug (n = 327, 321, 287, 265)
7
6
5
4
3
2
1
-1
0
1
2
3
4
5
6
7


Mean IMPS79 across Time by Group
IMPS79
Week (in square root units)
obs. placebo (n = 107, 105, 87, 70)
obs. drug (n = 327, 321, 287, 265)
7
6
5
4
3
2
1
-0.6
-0.2
0.2
0.6
1.0
1.4
1.8
2.2
2.6

Mixed-effects regression model (MRM) - Schizophrenia study

$$IMPS79 = Drug + Time + (Drug \times Time) + Subj + (Subj \times Time) + Error$$

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

$i = 1, \ldots, N$ subjects $\quad j = 1, \ldots, n_i$ obs

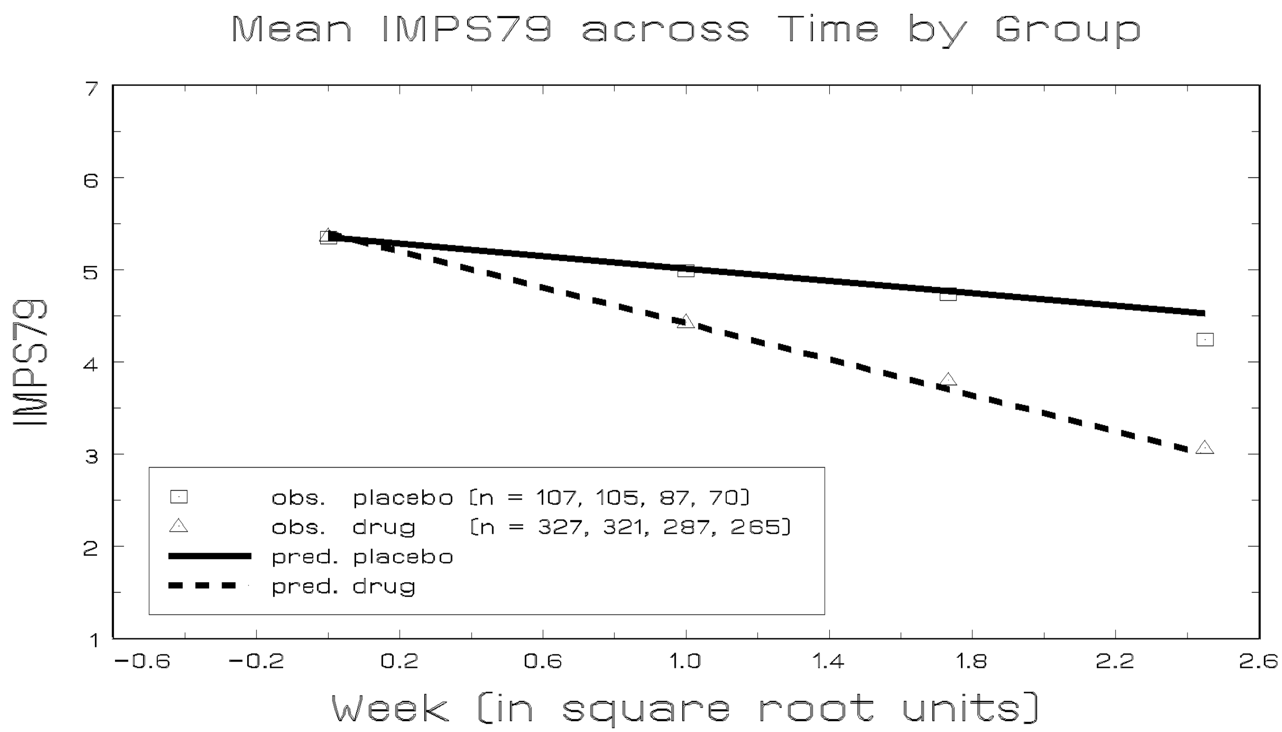
$\texttt{Drug} = 0$ for placebo, 1 for drug

$\texttt{SWeek} = 0, \sqrt{1} = 1, \sqrt{2} = 1.41, \sqrt{3} = 1.73, \sqrt{4} = 2, \sqrt{5} = 2.24, \sqrt{6} = 2.45$

NIMH Schizophrenia Study - IMPS79 across Time: ML Estimates (se)

	Completers $N = 335$			*All Subjects* $N = 437$		
	est.	se	$p <$	est.	se	$p <$
intercept	5.221	0.109	.001	5.348	0.088	.001
Drug (0 = plc; 1 = drug)	0.202	0.123	.101	0.046	0.101	.65
Time (sqrt week)	-0.393	0.073	.001	-0.336	0.068	.001
Drug by Time	-0.539	0.083	.001	-0.641	0.078	.001
Intercept variance	0.398	0.068		0.369	0.060	
	sd = .63			*sd =.61*		
Int-Time covariance	-0.011	0.035		0.021	0.034	
	r = -.04			*r =.07*		
Time variance	0.205	0.031		0.242	0.032	
	sd = .45			*sd =.49*		

Fitted and Obs. Means across Time by Condition



$$I\hat{MPS}_{ij} = 5.35 + .05\, Drug_i - .34\, Time_j - .64(D_i \times T_j)$$

Missing Data and Incomplete Data Models

$$\boldsymbol{y} \;\; \Big\langle \;\; \begin{array}{ll} \boldsymbol{y}^{(O)} & \text{observed} \\ \boldsymbol{y}^{(M)} & \text{missing} \end{array}$$

- GEE assumes "Missing Completely at Random" (MCAR)

  Conditional on covariates, missingness is independent of both $\boldsymbol{y}^{(O)}$ and $\boldsymbol{y}^{(M)}$

  $\Rightarrow$ "covariate-dependent missingness"

- Likelihood-based MRM assumes "Missing at Random" (MAR)

  Conditional on covariates *and* observed values of the dependent variable, missingness is independent of $\boldsymbol{y}^{(M)}$

  $\Rightarrow$ "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

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

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

Comments on MNAR models

- 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

Pattern-mixture models for missing data

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

- divide subjects into groups based on missing data pattern
- missing data pattern is a between-subjects variable used in longitudinal data analysis

With three timepoints, 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

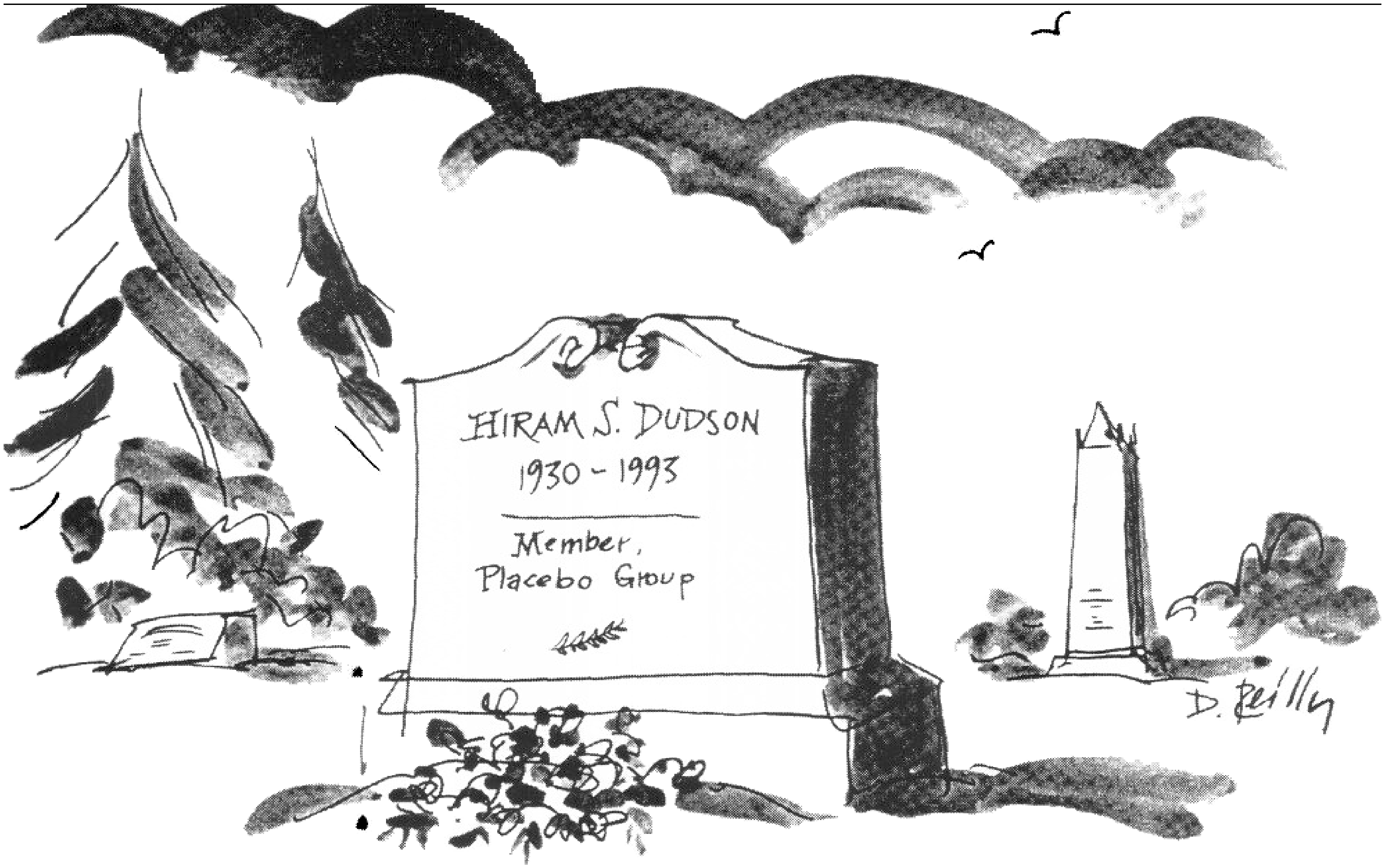
Classification based on missing data at week 6

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

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

- Dropout not independent of Drug $\chi^2_1 = 11.25, p < .001$
- Is dropout related to severity of illness?
- Does dropout moderate the influence of other variables' effects on severity of illness?

Where did the placebo patients go?



Mixed-effects pattern mixture model: Schiz data

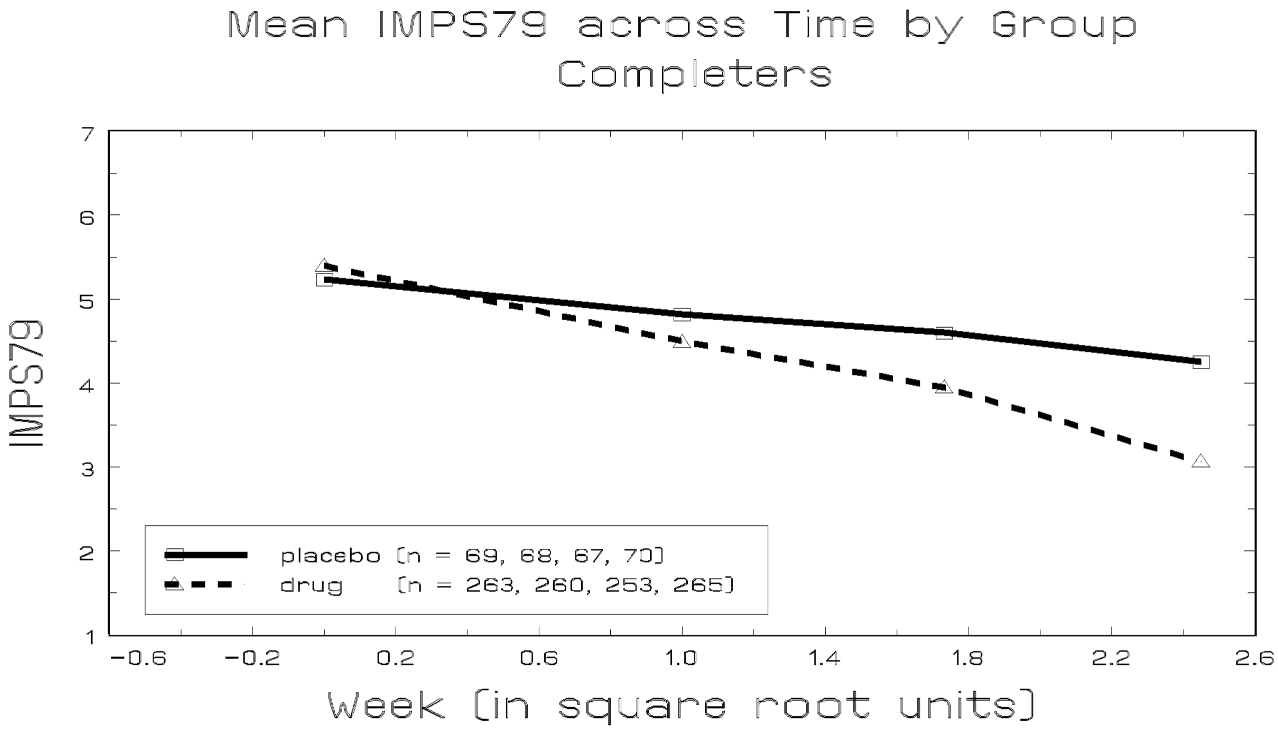
augment the basic MRM of `IMPS79` over time:

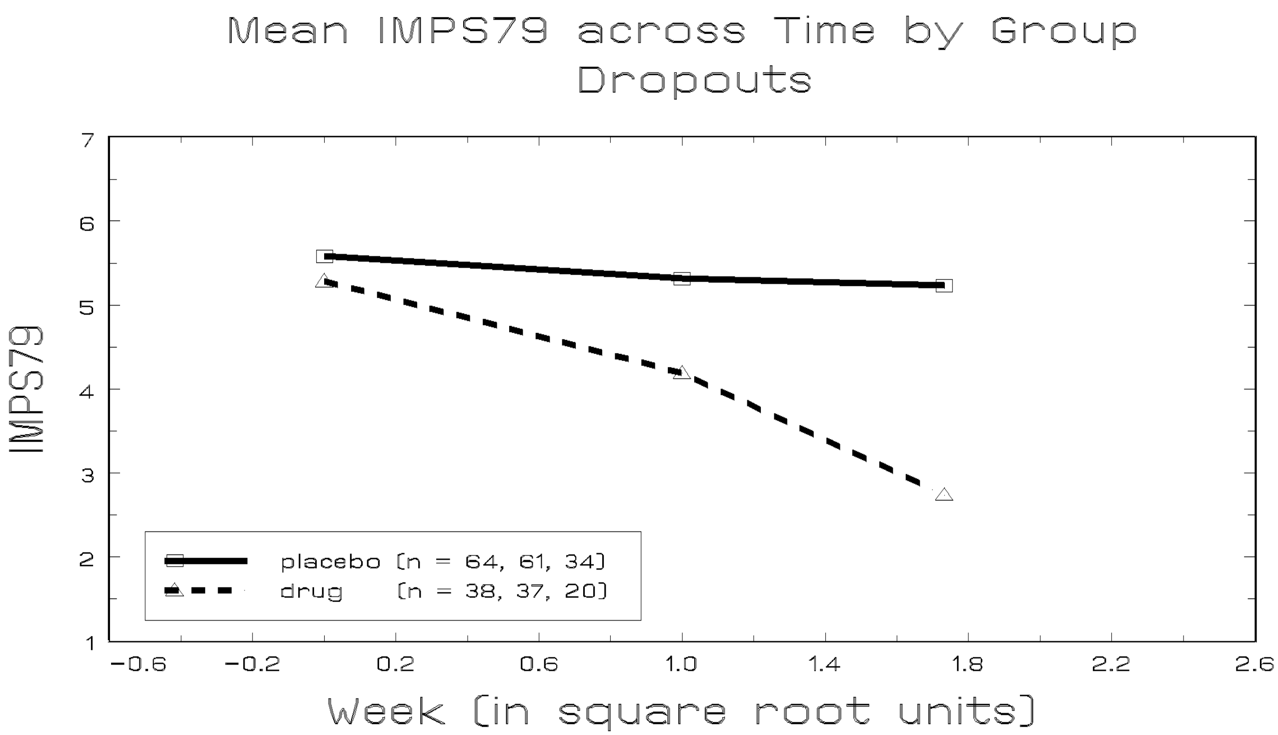
$$\texttt{IMPS79}_{ij} = \beta_0 + \beta_1 \texttt{Drug}_i + \beta_2 \texttt{SWeek}_j + \beta_3 (\texttt{Drug}_i \times \texttt{SWeek}_j) + \upsilon_{0i} + \upsilon_{1i} \texttt{SWeek}_j + \varepsilon_{ij}$$

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

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

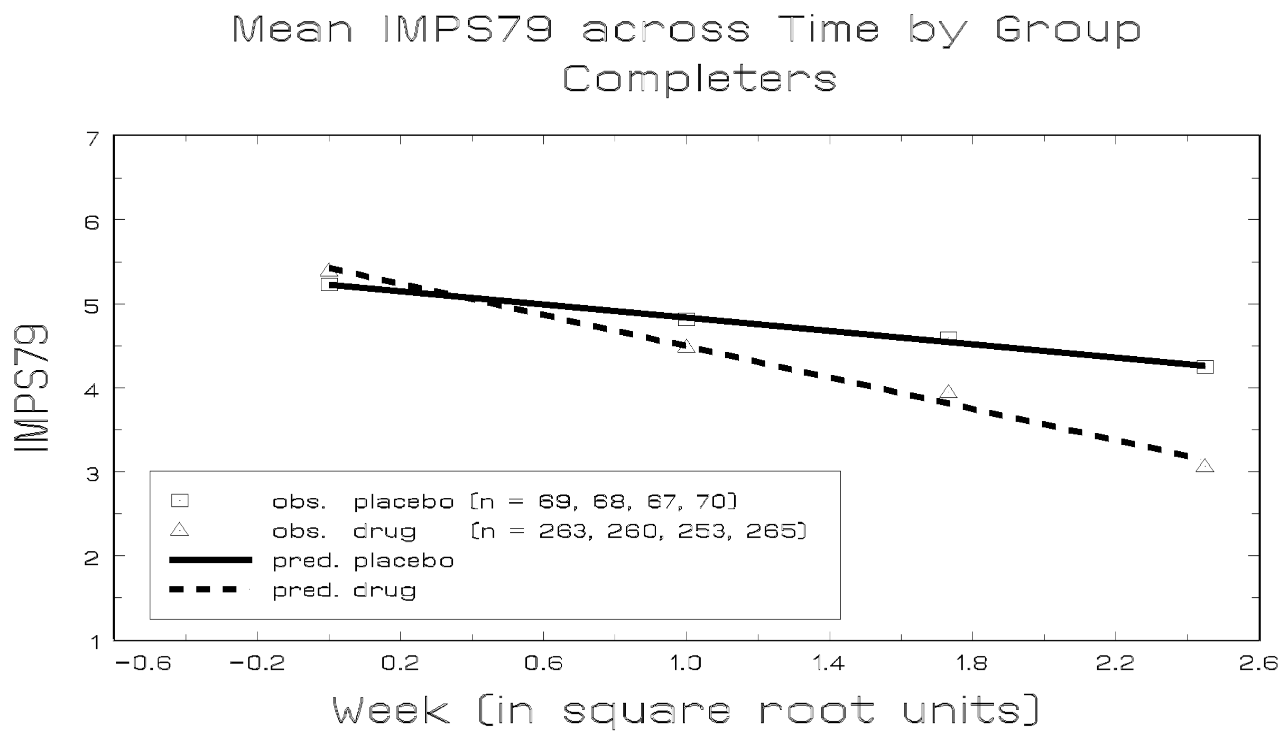
- β_0, β_1, β_2, and β_3 are for the completer subsample
- β_0^D, β_1^D, β_2^D, and β_3^D how dropouts differ from completers
- three-way interaction is of particular interest - indicates how the drug by time interaction varies with study completion

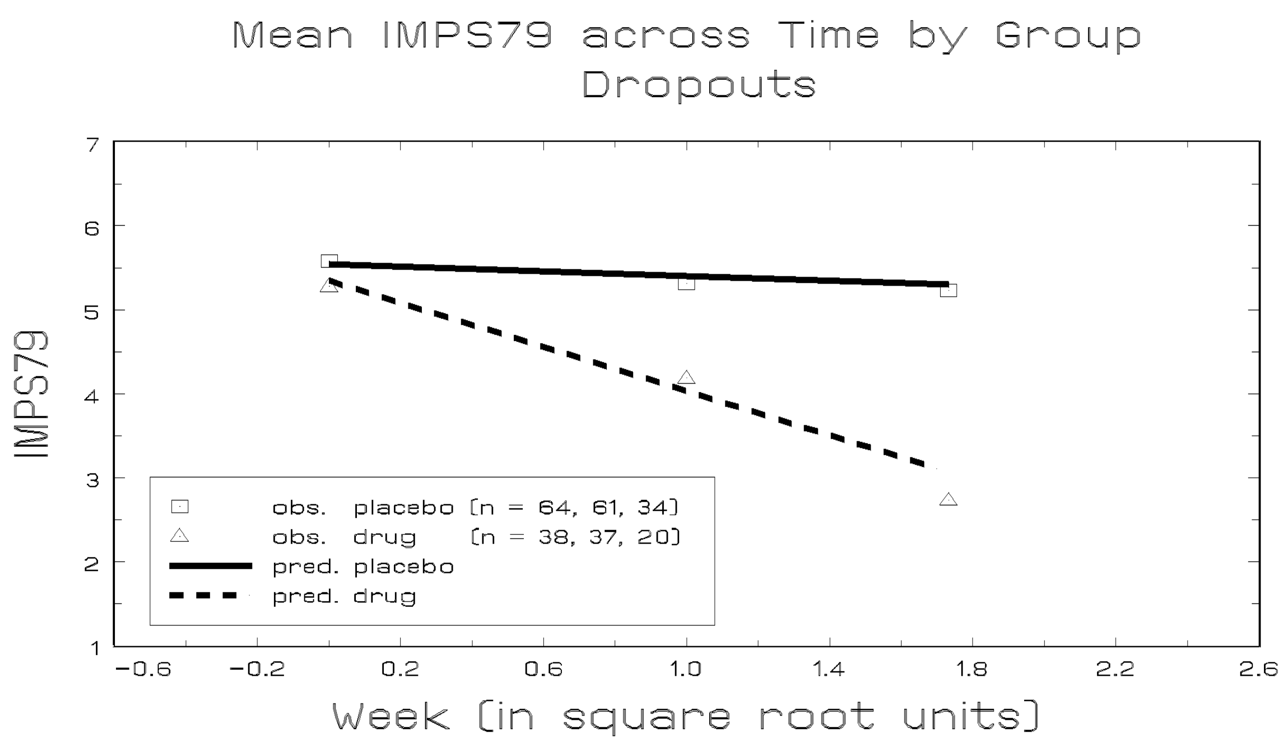

Mean IMPS79 across Time by Group
Completers
IMPS79
Week (in square root units)
placebo (n = 69, 68, 67, 70)
drug (n = 263, 260, 253, 265)
7
6
5
4
3
2
1
−0.6
−0.2
0.2
0.6
1.0
1.4
1.8
2.2
2.6


Mean IMPS79 across Time by Group
Dropouts
IMPS79
Week (in square root units)
placebo (n = 64, 61, 34)
drug (n = 38, 37, 20)
7
6
5
4
3
2
1
−0.6
−0.2
0.2
0.6
1.0
1.4
1.8
2.2
2.6

NIMH Schizophrenia Study - IMPS79 across Time: MRM models

	Ordinary			Pattern mixture			PM averaged		
parameter	est	se	$p <$	est	se	$p <$	est	se	$p <$
`Int` β_0	5.348	.088	.0001	5.221	.108	.0001	5.334	.089	.0001
`Drug` β_1	.046	.101	.65	.202	.121	.096	.124	.105	.24
`SWeek` β_2	-.336	.068	.0001	-.393	.076	.0001	-.305	.071	.0001
`Drug` $\times$ `SWeek` β_3	-.641	.078	.0001	-.539	.086	.0001	-.662	.078	.0001
`Int` β_0^D				.320	.186	.086			
`Drug` β_1^D				-.399	.227	.079			
`SWeek` β_2^D				.252	.159	.115			
`Drug` $\times$ `SWeek` β_3^D				-.635	.196	.002			
Deviance	4649.0			4623.3					


Mean IMPS79 across Time by Group
Completers
IMPS79
Week (in square root units)
obs. placebo (n = 69, 68, 67, 70)
obs. drug (n = 263, 260, 253, 265)
pred. placebo
pred. drug
7
6
5
4
3
2
1
−0.6
−0.2
0.2
0.6
1.0
1.4
1.8
2.2
2.6


Mean IMPS79 across Time by Group
Dropouts
IMPS79
Week (in square root units)
obs. placebo (n = 64, 61, 34)
obs. drug (n = 38, 37, 20)
pred. placebo
pred. drug
7
6
5
4
3
2
1
−0.6
−0.2
0.2
0.6
1.0
1.4
1.8
2.2
2.6

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

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)	

$\Rightarrow$ 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 - Schizophrenia study

Longitudinal outcome model - ordinary MRM:

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

Dropout model - grouped/discrete time survival analysis:

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

- random effects are summaries of a person's observed *and unobserved* $\boldsymbol{y}$ 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)

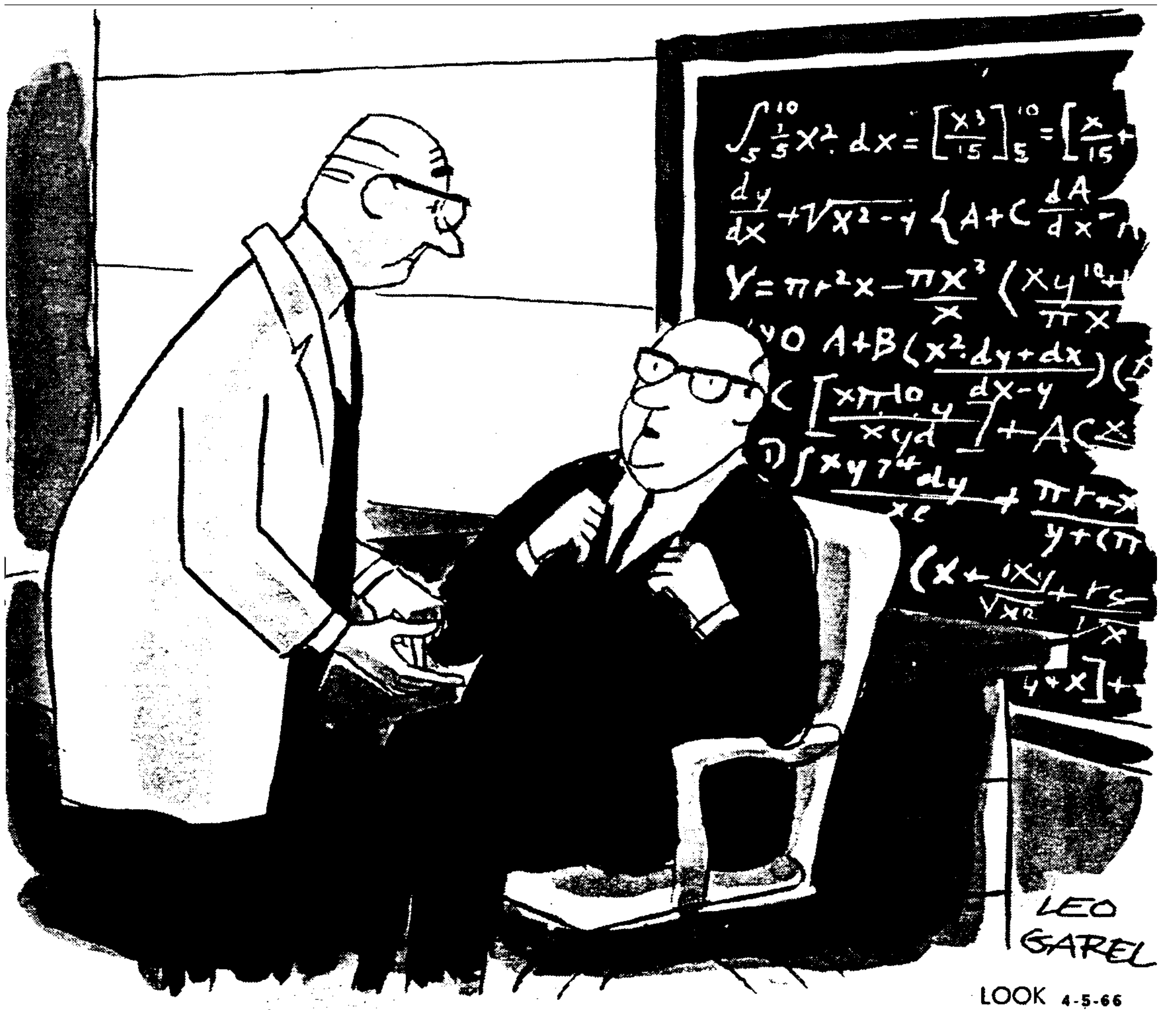
parameter	Separate ML est	se	$p <$	Shared ML est	se	$p <$
<u>Outcome</u>						
intercept β_0	5.348	.088	.0001	5.320	.088	.0001
`Drug` β_1	.046	.101	.65	.088	.102	.87
`SWeek` β_2	-.336	.068	.0001	-.272	.073	.0002
`Drug` $\times$ `Sweek` β_3	-.641	.078	.0001	-.737	.083	.0001
<u>Dropout</u>						
`Drug` 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
deviance	5380.2			5350.1		

- for longitudinal component, conclusions are same as MAR model
- marg. signif. slope - for placebo: tendency to dropout as slope increases
- signif. negative `Drug` $\times$ slope - for drug: slope effect is opposite; drug patients with more negative slopes (*i.e.*, greater improvement) more likely to drop out

NIMH Schizophrenia Study: Severity across Time

ML Estimates (se) *random intercept and slope models*

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



"I follow you up to the point where you add a pinch of oregano. . . ."

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

  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)

  Selection

  – missingness in terms of important covariates

  – missingness in terms of (shared) random subject effects

$\Rightarrow$ Does not invent data, maximizes information obtained from *available* data