Mixed-Effects Pattern-Mixture Models for Incomplete Longitudinal Data

Don Hedeker
University of Illinois at Chicago

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

Pattern-mixture models for missing data

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

- divide the 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

With subjects measured at 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 ("effect" or "sequential" coding)
- these variables are used as main effects and interactions

Combining patterns to increase interpretability

- groups based on last available measurement wave:

"recoded" pattern group	time1	time2	time3	D1	D2
1	O	M	M	1	0
2	O	O	M	0	1
2	M	O	M	0	1
3	O	O	O	0	0
3	O	M	O	0	0
3	M	O	O	0	0
3	M	M	O	0	0

- grouping of complete data vs. incomplete data,

"recoded" pattern group	time1	time2	time3	D1
1	O	O	O	0
2	O	M	M	1
2	O	O	M	1
2	M	O	M	1
2	O	M	O	1
2	M	O	O	1
2	M	M	O	1

- missing vs. present at the final timepoint

"recoded" pattern group	time1	time2	time3	D1
1	O	M	M	0
1	O	O	M	0
1	M	O	M	0
2	O	O	O	1
2	O	M	O	1
2	M	O	O	1
2	M	M	O	1

Studies with only attrition patterns - once a subject drops-out, they are gone:

pattern group	time1	time2	time3	D1	D2
1	O	M	M	1	0
2	O	O	M	0	1
3	O	O	O	0	0

$D1$ and $D2$ represent differences between each of the two dropout groups and the OOO group.

Considerations in forming groups from missing data patterns

- sparseness of the patterns
  - with large percentage of study completers, completers versus dropouts may be OK
- potential influence of the missing data pattern
  - intermittant missing may be random, while attrition is not
- interest in interactions
  - time interactions are not possible with OMM, MOM, & MMO patterns

Treatment-Related Change Across Time

NIMH Schizophrenia collaborative study on treatment related changes in overall severity (IMPS item # 79). 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

The experimental design and corresponding sample sizes:

	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

Main question of interest:

- Was there differential improvement for the drug groups relative to the control group?

Descriptive Statistics

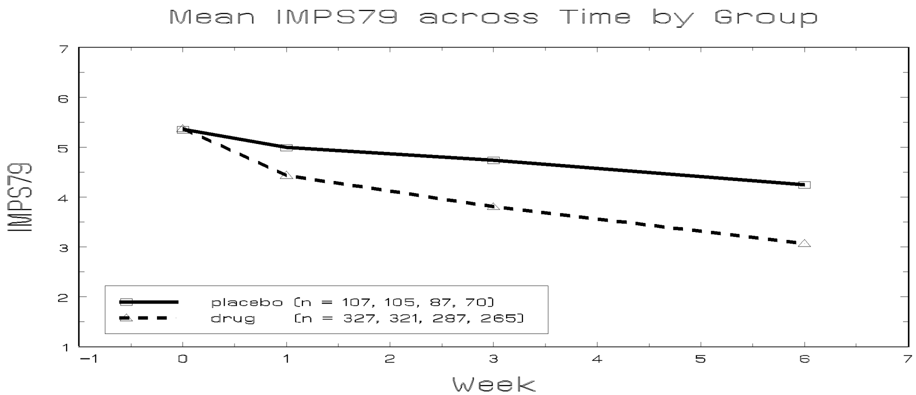
Observed IMPS79 Means, n, and sd

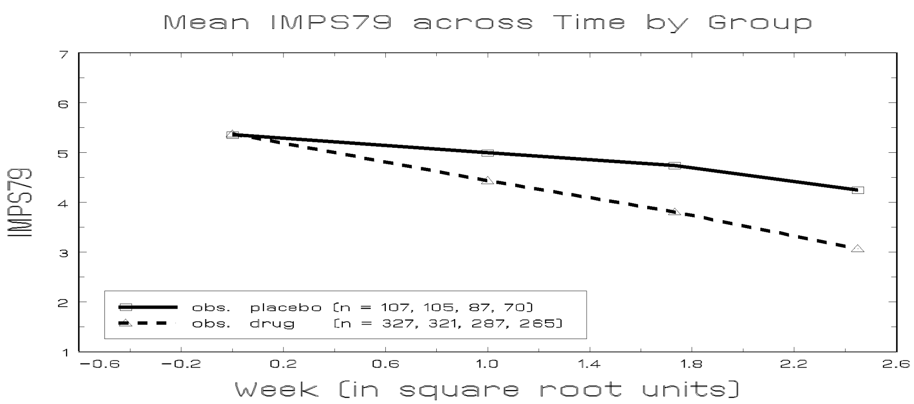
	week 0	week 1	week 3	week 6
placebo	5.35	4.99	4.74	4.25
n	107	105	87	70
drug	5.37	4.43	3.80	3.06
n	327	321	287	265
pooled sd	.87	1.23	1.44	1.48

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

Plots of Means across Time by Condition





Within-Subjects / Between-Subjects components

Within-subjects model - *level 1* $(j = 1, \ldots, n_i \text{ obs})$

$$IMPS79_{ij} = b_{0i} + b_{1i}\sqrt{Week}_j + \varepsilon_{ij}$$

Between-subjects model - *level 2* $(i = 1, \ldots, N \text{ subjects})$

$$b_{0i} = \beta_0 + \beta_2 Grp_i + \upsilon_{0i}$$

$$b_{1i} = \beta_1 + \beta_3 Grp_i + \upsilon_{1i}$$

$$\varepsilon_{ij} \sim NID(0, \sigma^2) \quad \text{level-1 residuals}$$
$$\boldsymbol{\upsilon}_i \sim NID(0, \boldsymbol{\Sigma}_\upsilon) \quad \text{level-2 residuals}$$

β_0 = average week 0 IMPS79 level for PLC patients ($Grp = 0$)

β_1 = average IMPS79 (sqrt) weekly change for PLC patients ($Grp = 0$)

β_2 = average difference in week 0 IMPS79 for DRUG patients ($Grp = 1$)

β_3 = average difference in IMPS79 (sqrt) weekly change for DRUG patients ($Grp = 1$)

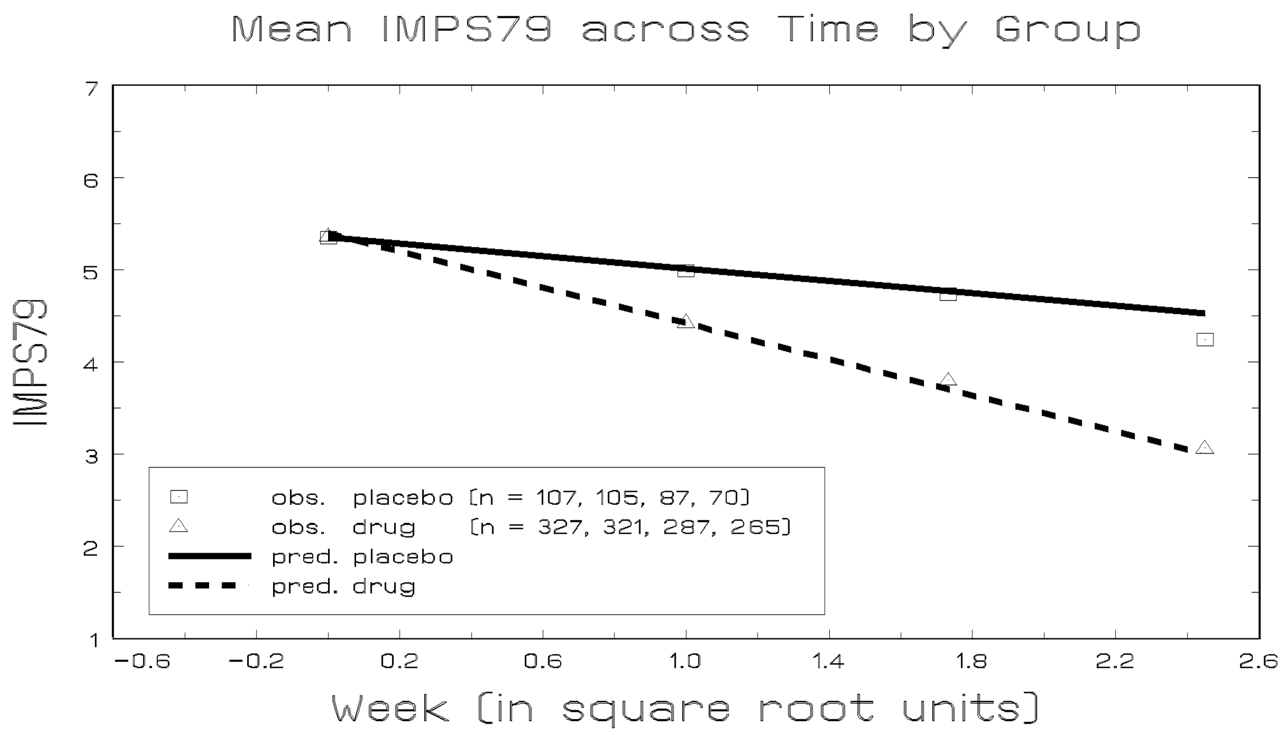
υ_{0i} = individual deviation from group intercept

υ_{1i} = individual deviation from group (sqrt) weekly change

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

	Completers $N = 335$		*All Subjects* $N = 437$	
	est.	se	est.	se
intercept	5.221	0.109	5.348	0.088
Drug (0 = plc; 1 = drug)	0.202	0.123	0.046	0.101
Time (sqrt week)	-0.393	0.073	-0.336	0.068
Drug by Time	-0.539	0.083	-0.641	0.078
Intercept var	0.398	0.068	0.369	0.060
	sd = .63		*sd =.61*	
Int-Time covar	-0.011	0.035	0.021	0.034
	r = -.04		*r =.07*	
Time var	0.205	0.031	0.242	0.032
	sd = .45		*sd =.49*	
-2 log L	3782.1		4649.0	

Fitted and Obs. Means across Time by Condition



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

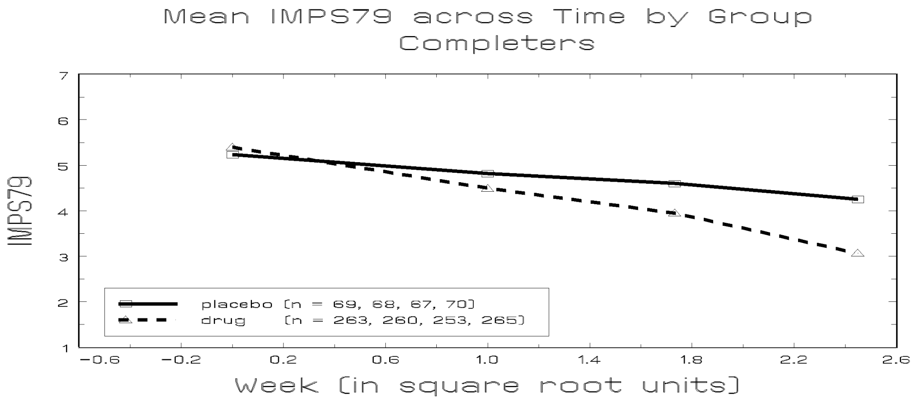
Classification of Subjects based on missing-data

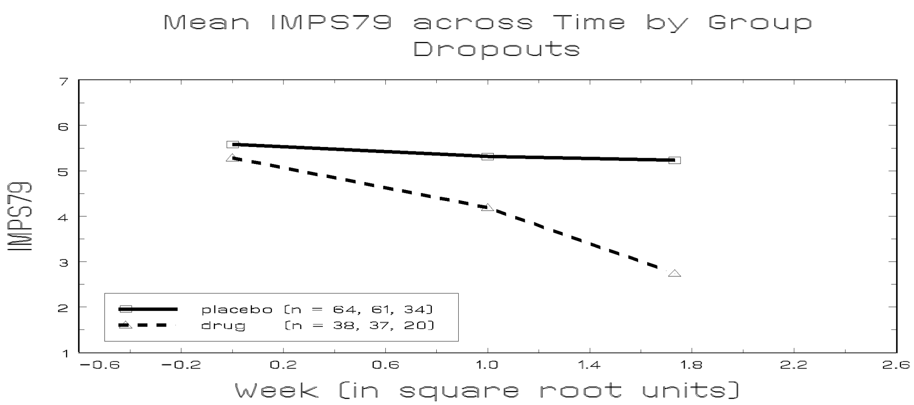
$$Dropout_i = \begin{cases} 0 & \text{subject measured at week 6 (last timepoint)} \\ 1 & \text{subject missing at week 6 (last timepoint)} \end{cases}$$

Drug group	Dropout group: completer	Dropout 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?

Means across Time by Condition by Dropout





Within-Subjects / Between-Subjects components

<u>*Within-subjects model*</u> *- level 1* $\quad (j = 1, \ldots, n_i \text{ obs})$

$$IMPS79_{ij} = b_{0i} \ + \ b_{1i}\sqrt{Week_j} \ + \ \varepsilon_{ij}$$

<u>*Between-subjects model*</u> *- level 2* $\quad (i = 1, \ldots, N \text{ subjects})$

$$b_{0i} \ = \ \beta_0 \ + \ \beta_2 Grp_i \ + \ \beta_4 Drop_i \ + \ \beta_6 (G_i \times D_i) \ + \ \upsilon_{0i}$$

$$b_{1i} \ = \ \beta_1 \ + \ \beta_3 Grp_i \ + \ \beta_5 Drop_i \ + \ \beta_7 (G_i \times D_i) \ + \ \upsilon_{1i}$$

$$\varepsilon_{ij} \ \sim \ NID(0, \sigma^2) \quad \text{level-1 residuals}$$
$$\boldsymbol{\upsilon}_i \ \sim \ NID(0, \boldsymbol{\Sigma}_\upsilon) \quad \text{level-2 residuals}$$

$$Grp_i = \begin{cases} 0 & \text{placebo} \\ 1 & \text{drug} \end{cases} \qquad Drop_i = \begin{cases} 0 & \text{completer} \\ 1 & \text{dropout} \end{cases}$$

Pattern mixture model: *Dropouts vs Completers*

	est.	se	z	$p <$
intercept *placebo completers at week 0*	5.221	0.108	48.55	.001
Drug (0=P; 1=D) *for completers at week 0*	0.202	0.121	1.67	.10
Dropout (0=N; 1=Y) *for placebo at week 0*	0.320	0.186	1.72	.09
Drug by Dropout *week 0 interaction*	-0.399	0.227	-1.76	.08
Time (sqrt week) *for placebo completers*	-0.393	0.076	-5.15	.001
Drug by Time *for completers*	-0.539	0.086	-6.28	.001
Dropout by Time *for placebo*	0.252	0.159	1.58	.12
3 Way	-0.635	0.196	3.24	.002

Estimated Trends by Subgroups

Placebo Completers

$$\hat{Y} \;=\; 5.22 - .39\,Time$$

Drug Completers

$$\hat{Y} \;=\; (5.22 + .20) - (.39 + .54)\,Time \;\;=\; 5.42 - .93\,Time$$

⇒ *difference in slope = -.54*

Placebo Dropouts

$$\hat{Y} \;=\; (5.22 + .32) - (.39 + .25)\,Time \;\;=\; 5.54 - .14\,Time$$

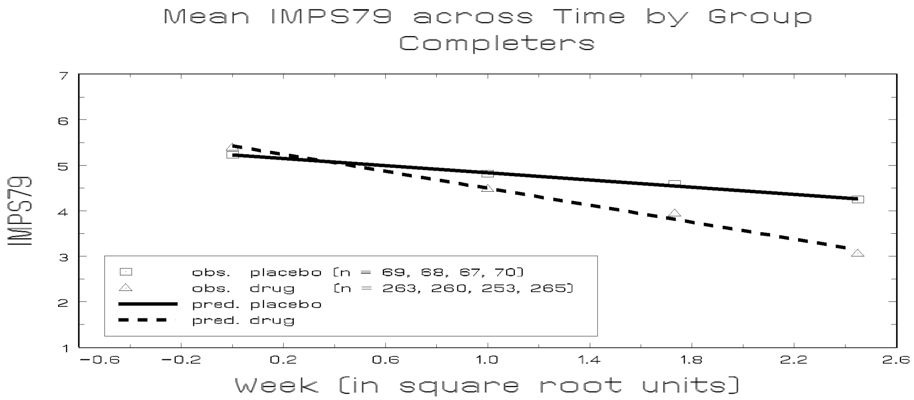
Drug Dropouts

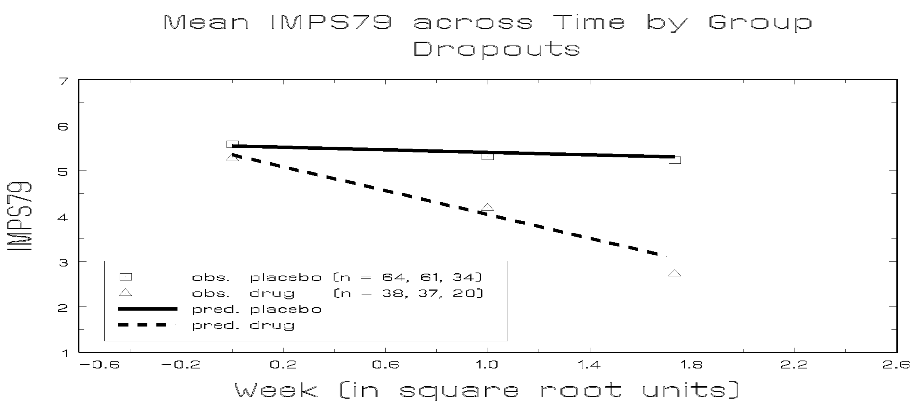
$$\hat{Y} \;=\; (5.22 + .20 + .32 - .40) - (.39 + .54 - .25 + .64)\,Time$$
$$=\; 5.34 - 1.32\,Time$$

⇒ *difference in slope = -1.17*

⇒ *difference in slope difference = -.64*
degree that Drug by Time interaction varies by Dropout

Obs & Est Means across Time by Cond by Drop





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		**Dropout**	
placebo	*70/108*	*placebo*	*38/108*
drug	*265/329*	*drug*	*64/329*
	335/437		**102/437**

Pattern-mixture averaged results

$$\hat{\hat{\boldsymbol{\beta}}} \; = \; \hat{\pi}_c \, \hat{\boldsymbol{\beta}}_c \; + \; \hat{\pi}_d \, \hat{\boldsymbol{\beta}}_d$$

or

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

note:

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

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

Placebo Time effect

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

Drug Intercept difference

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

Drug Time difference

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

Delta Method for estimating asymptotic variance of averaged estimates

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

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

where

$$\frac{\partial \hat{\bar{\boldsymbol{\beta}}}}{\partial \hat{\boldsymbol{\beta}}_c} = 1 \qquad \frac{\partial \hat{\bar{\boldsymbol{\beta}}}}{\partial \hat{\boldsymbol{\beta}}_\Delta} = \hat{\pi}_d \qquad \frac{\partial \hat{\bar{\boldsymbol{\beta}}}}{\partial \hat{\pi}_d} = \hat{\boldsymbol{\beta}}_\Delta$$

Thus,

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

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

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

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

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

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 $\hat{V}(\hat{\bar{\boldsymbol{\beta}}})_F$ to get correct standard errors

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

parameter	$\hat{\bar{\beta}}$	$\hat{V}(\hat{\bar{\beta}})_F$	$\hat{\beta}_\Delta$	Augment	$\hat{V}(\hat{\bar{\beta}})$	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 × 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

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

Placebo Time effect

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

Drug Intercept difference

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

Drug Time difference

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

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

parameter	$\hat{\bar{\beta}}$	$\hat{V}(\hat{\bar{\beta}})_F$	$\hat{\beta}_\Delta$	Augment	$\hat{V}(\hat{\bar{\beta}})$	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 × time	-.6621	$(.0772)^2 = .00596$	-.6348	.000192	.00615	.0784

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

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

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

NIMH Schizophrenia Study: Severity across Time

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

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

Conclusions

- Mixed-effects models 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 models maybe more useful
  - adds missing-data pattern as between-subjects factor
  - assesses degree to which "missingness" influences (available) outcomes
  - assesses degree to which "missingness" interacts with model terms
    * intervention group
    * intervention group by time

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