Missing Data in Longitudinal Studies

Hedeker D & Gibbons RD (1997). Application of random-effects pattern-mixture models for missing data in longitudinal studies. Psychological Methods, 2, 64-78.

Chapter 14 in Hedeker & Gibbons (2006), Longitudinal Data Analysis, Wiley.

Missing Data Patterns and Mechanisms

Pattern: which values are missing?

Mechanism: Why? Is missingness related to study variables?

$$\boldsymbol{y} = \text{complete data matrix}$$

$$\begin{aligned} \boldsymbol{y}^{(O)} &= \text{observed part of } \boldsymbol{y} \\ \boldsymbol{y}^{(M)} &= \text{missing part of } \boldsymbol{y} \end{aligned}$$

$$\begin{aligned} \boldsymbol{R} &= \text{missing data indicator matrix} \\ R_{ij} &= \begin{cases} 1 & y_{ij} \text{ missing} \\ 0 & y_{ij} \text{ observed} \end{cases} \end{aligned}$$

Consider the joint distribution of $\boldsymbol{Y}$ and $\boldsymbol{R}$ (both are random variables)

- Pattern concerns the distribution of $\boldsymbol{R}$
- Mechanism concerns the distribution of $\boldsymbol{R}$ given $\boldsymbol{Y}$
  how does this distribution depend on $\boldsymbol{Y}$

Rubin (1976) - typology for missing data

- Missing Completely At Random (MCAR)

$$P(\boldsymbol{R} \mid \boldsymbol{y}) = P(\boldsymbol{R}) \text{ for all } \boldsymbol{y}$$

$\Rightarrow \boldsymbol{R}$ is independent of both $\boldsymbol{y}^{(O)}$ and $\boldsymbol{y}^{(M)}$

(covariate-dependent missingness special case of MCAR)

- Missing At Random (MAR)

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

$\Rightarrow \boldsymbol{R}$ is independent of $\boldsymbol{y}^{(M)}$

- Missing Not At Random (MNAR)

$$P(\boldsymbol{R} \mid \boldsymbol{y}) \text{ depends on } \boldsymbol{y}^{(M)}$$

$\Rightarrow$ sometimes called "non-ignorable non-response" or "informative missingness"

Properties of methods strongly influenced by assumptions made about mechanism

Important to consider nature of mechanism in a particular application

General strategies

- Complete-case (CC) analysis (*i.e.*, discard cases with incomplete data)
  - CC may be a biased sample
  - inefficient
  - weighting of complete cases can be used to correct for bias

- Imputation
  - naive imputations (*e.g.*, means, LOCF) can be worse than CC analysis
  - multiple imputation good (Schafer & Graham, 2002)

- Analyze as incomplete (*e.g.*, MRM, GEE, CPM)
  - different methods have different mechanism assumptions

Incomplete Data Models

Distinguish between dependent variable

$$\boldsymbol{y} \;\langle\; \begin{array}{ll} \boldsymbol{y}^{(O)} & R = 0 \\ \boldsymbol{y}^{(M)} & R = 1 \end{array}$$

and independent variables (all observed)

$$\boldsymbol{X} \quad \text{time, group, } \ldots$$

- GEE1 (*i.e.*, ordinary GEE) assumes special case of MCAR

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

  $\Rightarrow$ conditional on covariates, $\boldsymbol{R}$ is independent of both $\boldsymbol{y}^{(O)}$ and $\boldsymbol{y}^{(M)}$ ("covariate-dependent missingness")

- Likelihood-based methods (MRM, CPM) assume MAR

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

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

"ignorable non-response" (Laird, 1988)

Predictors of Missingness

Both MCAR and MAR allow missingness to depend on $\boldsymbol{X}$ variables

- Very important to include appropriate $\boldsymbol{X}$ variables which predict missingness in longitudinal data analysis model

- Demirtas and Schafer (2003) advise including a variable like "How likely is it that you will remain in this study through the next measurement period?" in longitudinal questionnaires. To the extent that this question is related to subsequent missingness, including this variable as a covariate in analyses could convert a non-ignorable situation to one that is essentially ignorable.

Simulation Study

Data from 5000 subjects were simulated according to:

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

$T_j = 0, 1, 2, 3, 4$ for five timepoints
$G_i =$ dummy-code (0 or 1) with half in each group

Regression coefficients:
$\beta_0 = 25$, $\beta_1 = -1$, $\beta_2 = 0$, and $\beta_3 = -1$

$\Rightarrow$ the population means are:
25, 24, 23, 22, and 21 for $Grp = 0$
25, 23, 21, 19, and 17 for $Grp = 1$

Variance parameters:
$\sigma^2_{\upsilon_0} = 4$, $\sigma^2_{\upsilon_1} = .25$, $\sigma_{\upsilon_{01}} = -.1$ ($\rho = -.1$), $\sigma^2 = 4$

The population variance-covariance matrix, $V(\boldsymbol{y}) = \boldsymbol{Z}\boldsymbol{\Sigma}_v\boldsymbol{Z}' + \sigma^2\boldsymbol{I}$

$$V(\boldsymbol{y}) = \begin{bmatrix} 8.00 & 3.90 & 3.80 & 3.70 & 3.60 \\ 3.90 & 8.05 & 4.20 & 4.35 & 4.50 \\ 3.80 & 4.20 & 8.60 & 5.00 & 5.40 \\ 3.70 & 4.35 & 5.00 & 9.65 & 6.30 \\ 3.60 & 4.50 & 5.40 & 6.30 & 11.20 \end{bmatrix},$$

or expressed as a correlation matrix:

$$\begin{bmatrix} 1.00 & 0.49 & 0.46 & 0.42 & 0.38 \\ 0.49 & 1.00 & 0.50 & 0.49 & 0.47 \\ 0.46 & 0.50 & 1.00 & 0.55 & 0.55 \\ 0.42 & 0.49 & 0.55 & 1.00 & 0.61 \\ 0.38 & 0.47 & 0.55 & 0.61 & 1.00 \end{bmatrix}$$

Scenarios

complete data: no missing data.

50% random missing: 50% missing data at every timepoint; completely random and unrelated to any variable.

time related dropout: dropout rates of 0%, 25%, 50%, 75%, 87.5% for the five timepoints. If a subject was missing at a timepoint, then they were also missing at all later timepoints; these rates indicate the percentage of the original sample that were missing at each of these timepoints.

group by time related dropout: dropout rates of
0%, 23%, 46%, 70% and 83% for G=0
0%, 27%, 55%, 81% and 91% for G=1

$\Rightarrow$ notice that these missing data scenarios are all MCAR
(as long as analysis model includes time, G, and G by time)

MCAR Simulation Results - MRM estimates (standard errors)

	β_0	β_1	β_2	β_3	$\sigma^2_{v_0}$	$\sigma_{v_{01}}$	$\sigma^2_{v_1}$	σ^2
simulated value:	25	-1	0	-1	4	-.1	.25	4
complete data	24.969	-.994	-.001	-.986	3.918	-.057	.239	3.991
	(.050)	(.016)	(.071)	(.023)	(.129)	(.032)	(.014)	(.046)
50% random	24.991	-1.024	-.087	-.933	3.811	-.020	.199	4.070
missing	(.063)	(.023)	(.089)	(.032)	(.193)	(.056)	(.025)	(.083)
Time-related	24.989	-.968	.019	-1.021	3.853	-.062	.229	4.000
dropout	(.053)	(.028)	(.075)	(.040)	(.150)	(.060)	(.032)	(.074)
Group by Time	24.991	-.977	.041	-1.014	3.872	-.048	.234	3.994
related dropout	(.053)	(.026)	(.075)	(.041)	(.150)	(.059)	(.031)	(.073)

- GEE results were near-identical for the regression coefficients
- Standard errors do indicate the effect of missing data; these are larger when missing data are present

$\Rightarrow$ If missing data mechanism is MCAR, analysis by GEE or MRM yields valid results, provided that the model includes the predictors of missingness

MAR and MNAR Scenarios

MAR(a): if the value of the dependent variable was lower than 23, then the subject dropped out at the next timepoint (*i.e.*, they were missing at the next and all subsequent timepoints).

MAR(b): the MAR specification was different for the two groups. For $Grp = 1$, if the dependent variable was lower than 23, then the subject dropped out at the next timepoint, however for $Grp = 0$, if the dependent variable was greater than 25.5 then the subject dropped out at the next timepoint.

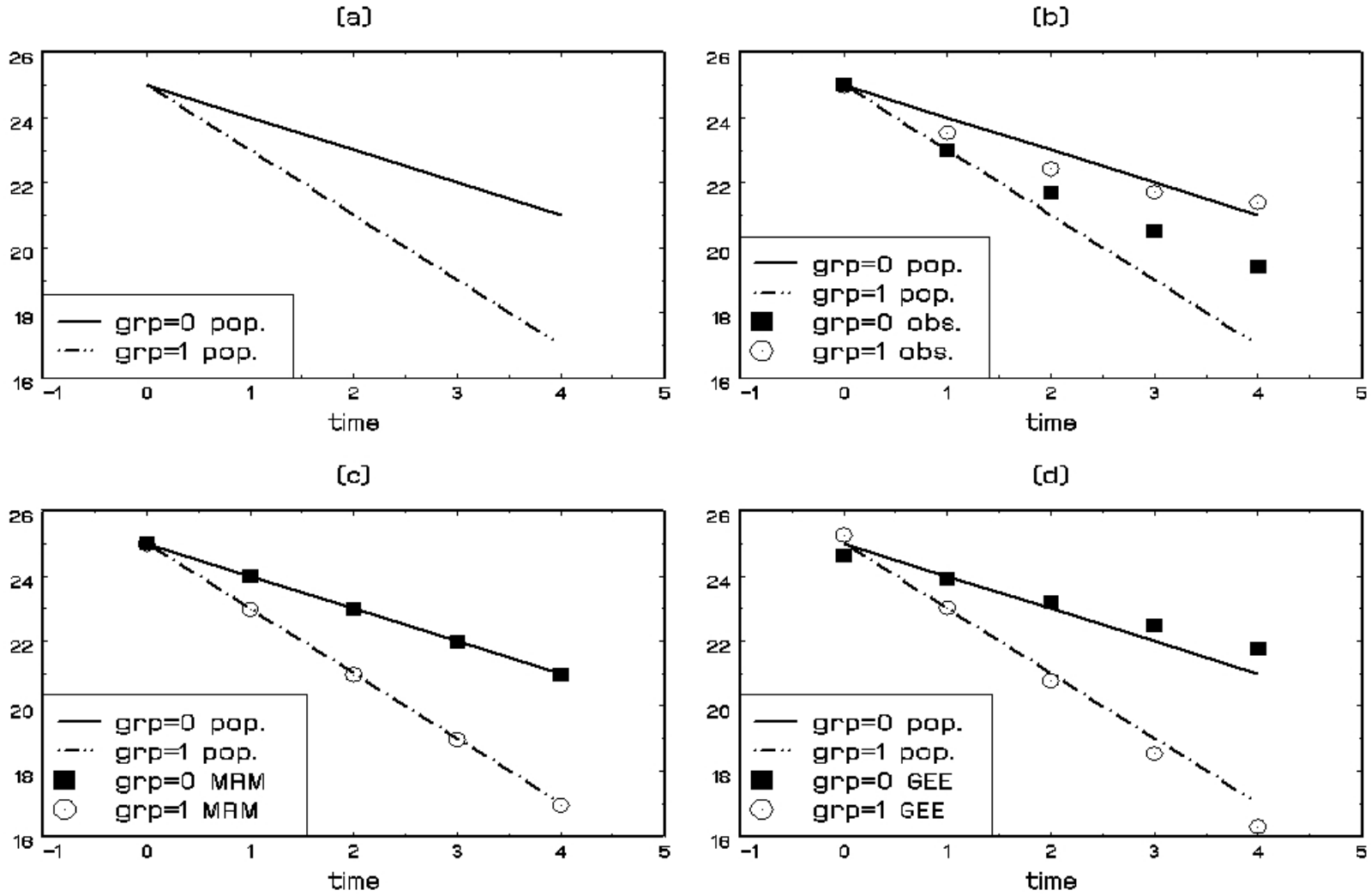
MNAR: after the first timepoint, if the value of the dependent variable was lower than 21.5, then the subject was missing *at that timepoint* and all subsequent timepoints.

All of these cutoff values were selected to yield approximate dropout rates of of 0%, 25%, 50%, 75%, and 87.5% for the five timepoints, so the amount of missing data is very similar to the earlier MCAR simulations.

MAR and MNAR Simulation Results - Estimates (standard errors)

	β_0	β_1	β_2	β_3	$\sigma^2_{v_0}$	$\sigma_{v_{01}}$	$\sigma^2_{v_1}$	σ^2
simulated value:	25	-1	0	-1	4	-.1	.25	4
MAR(a)								
MRM	24.996	-1.039	-.010	-.969	3.981	-.064	.233	3.873
	(.053)	(.025)	(.075)	(.041)	(.158)	(.065)	(.032)	(.078)
GEE1	25.281	-1.164	.019	-1.001				
	(.058)	(.037)	(.097)	(.085)				
MAR(b)								
MRM	24.999	-1.003	-.016	-1.004	4.050	-.082	.229	3.812
	(.053)	(.022)	(.075)	(.039)	(.154)	(.064)	(.027)	(.073)
GEE1	24.635	-.714	.634	-1.532				
	(.055)	(.030)	(.097)	(.090)				
MNAR								
MRM	24.956	-.233	.027	-.552	3.856	-.943	.319	3.020
	(.049)	(.020)	(.070)	(.035)	(.131)	(.051)	(.025)	(.053)
GEE1	25.051	-.386	.016	-.583				
	(.049)	(.020)	(.071)	(.034)				

- MAR: MRM does well, GEE1 not so well; for MNAR: neither does well



MAR(b) simulation:
(a) population means across time by group; *(b)* observed versus population means across time by group; *(c)* MRM estimates vs pop. means across time by group; *(d)* GEE estimates vs pop. means across time by group

Notice, observed means *(b)* very misleading in terms of the population trends

Mis-specification of Variance-Covariance

Re-analysis of the simulated MAR-generated data, however using only a random-intercepts model:

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

This is a mis-specified model for the variance-covariance structure because the random trend term was omitted from the analysis.

Mis-specified MAR Simulation Results - Estimates (standard errors)

	β_0	β_1	β_2	β_3	$\sigma^2_{\upsilon_0}$	$\sigma_{\upsilon_{01}}$	$\sigma^2_{\upsilon_1}$	σ^2
simulated value:	25	-1	0	-1	4	-.1	.25	4
MRM with MAR(a)	24.938	-.891	-.009	-.949	3.722			4.329
	(.053)	(.021)	(.075)	(.036)	(.136)			(.072)
MRM with MAR(b)	24.998	-1.048	-.069	-.805	3.880			4.288
	(.053)	(.018)	(.076)	(.035)	(.138)			(.070)

- these random-intercepts analyses yield biased results, in particular for the time-related parameters β_1 and β_3.
- Performing any full-likelihood analysis, even with missing data following an MAR mechanism, does not guarantee that the correct results will be obtained (need to have the mean structure and variance-covariance structure of $\boldsymbol{y}$ correctly modeled).

Testing MCAR

- if MCAR, then either GEE or MRM is fine (provided that the covariate matrix $\boldsymbol{X}_i$ includes predictors of missingness)

- if MAR, then GEE does not perform well, whereas MRM analysis is acceptable (as long as the mean and variance-covariance structures are correctly modeled)

- useful to determine whether MCAR is acceptable or not

- distinction between MCAR and MAR is that missingness cannot depend on observed values of the dependent variable, $\boldsymbol{y}_i^O$, in the former, but can in the latter

- Tests of MCAR are based on analyses involving $\boldsymbol{y}_i^O$

Testing MCAR in 2-timepoint study

- suppose all subjects have data at time 1, but some are missing at time 2

- Define $D_i = 0$ for subjects with data at both timepoints, $D_i = 1$ for subjects with data at first timepoint only

- Compare y_1 between these two groups ($D_i = 0$ vs $D_i = 1$); MCAR posits y_1 data should not differ

Logistic regression model

$$\ln\left[\frac{\mathrm{P}(D_i = 1)}{1 - \mathrm{P}(D_i = 1)}\right] = \alpha_0 + \alpha_1 y_{i1} + \boldsymbol{\alpha}_2 \boldsymbol{x}_i + \boldsymbol{\alpha}_3 (y_{i1} \times \boldsymbol{x}_i)$$

$\boldsymbol{\alpha}_2$ and $\boldsymbol{\alpha}_3$ are vectors of regression coefficients for $\boldsymbol{x}_i$ and their interactions with y_{i1} $\quad \Rightarrow$ MCAR dictates that $\alpha_1 = \boldsymbol{\alpha}_3 = 0$

Testing MCAR in multiple timepoint study

$$\ln\left[\frac{\mathrm{P}(D_i = 1)}{1 - \mathrm{P}(D_i = 1)}\right] = \alpha_0 + \alpha_1 y_{i1} + \boldsymbol{\alpha}_2 \boldsymbol{x}_i + \boldsymbol{\alpha}_3 (y_{i1} \times \boldsymbol{x}_i)$$

can be generalized both in terms of the left and right side

Left side generalization - dropout recast as time to dropout

Right side generalization

- final observed measurement of y_{ij} instead of y_{i1}
- function of observed dependent variable values, $h(\boldsymbol{y}_i^O)$

Some possible functions of $\boldsymbol{y}_i^O$

- an average

$$h(\boldsymbol{y}_i^O) = \frac{1}{n_i} \sum_{j=1}^{n_i} y_{ij}$$

- a weighted average

$$h(\boldsymbol{y}_i^O) = \sum_{j=1}^{n_i} w_j y_{ij}$$

- moving average

$$\begin{aligned} h(\boldsymbol{y}_{i1}^O) &= y_{i1} \\ h(\boldsymbol{y}_{i2}^O) &= 1/2(y_{i1} + y_{i2}) \\ h(\boldsymbol{y}_{i3}^O) &= 1/3(y_{i1} + y_{i2} + y_{i3}) \\ &\ldots \end{aligned}$$

Discrete- or Grouped-Time Survival Analysis

$$\log\left[\frac{\mathrm{P}(D_i = j \mid D_i \geq j)}{1 - \mathrm{P}(D_i = j \mid D_i \geq j)}\right] = \alpha_{0j} + \alpha_1 h(\boldsymbol{y}_{ij}^O) + \boldsymbol{\alpha}_2 \boldsymbol{x}_{ij} + \boldsymbol{\alpha}_3 (h(\boldsymbol{y}_{ij}^O) \times \boldsymbol{x}_{ij})$$

- D_i is time of dropout (last measured timepoint)
- timepoints are indexed by $j = 1, \ldots n$
- define completers as $D_i = n$ (n is the last timepoint)
- the logit is for the probability of dropout at a given timepoint, given that it has not already occurred
- estimated with standard logistic regression software by creating person-period dataset (or using ordinal regression if no time-varying predictors)

$\Rightarrow$ MCAR is rejected if $\alpha_1 = \boldsymbol{\alpha}_3 = 0$ is rejected

Example of a person-period dataset

study with four timpoints (or periods); D_i = time of dropout; y_{ij} = dichotomous outcome for fixed-effects logistic regression

ID	D_i	period	y_{ij}
101	1	1	1
—	—	—	—
102	2	1	0
102	2	2	1
—	—	—	—
103	3	1	0
103	3	2	0
103	3	3	1
—	—	—	—
104	4	1	0
104	4	2	0
104	4	3	0

Person-period dataset with covariates

ID	D_i	period	y_{ij}	Sex	Stress	y average
101	1	1	1	Sex_{101}	$\text{Stress}_{101,1}$	$y_{101,1}$
—	—	—	—	—	—	—
102	2	1	0	Sex_{102}	$\text{Stress}_{102,1}$	$y_{102,1}$
102	2	2	1	Sex_{102}	$\text{Stress}_{102,2}$	$(y_{102,1} + y_{102,2})/2$
—	—	—	—	—	—	—
103	3	1	0	Sex_{103}	$\text{Stress}_{103,1}$	$y_{103,1}$
103	3	2	0	Sex_{103}	$\text{Stress}_{103,2}$	$(y_{103,1} + y_{103,2})/2$
103	3	3	1	Sex_{103}	$\text{Stress}_{103,3}$	$(y_{103,1} + y_{103,2} + y_{103,3})/3$
—	—	—	—			
104	4	1	0	Sex_{104}	$\text{Stress}_{104,1}$	$y_{104,1}$
104	4	2	0	Sex_{104}	$\text{Stress}_{104,2}$	$(y_{104,1} + y_{104,2})/2$
104	4	3	0	Sex_{104}	$\text{Stress}_{104,3}$	$(y_{104,1} + y_{104,2} + y_{104,3})/3$

- stress and y average are time-varying predictors
- Once the data are organized in this way, a logistic regression analysis can be performed regressing the dropout indicators y_{ij} on period, sex, stress, y average, and interactions

Complementary log-log link

- Although it is common to use logistic regression, and therefore the logit link, the complementary log-log (clog-log) link function can also be used

- Use of the clog-log link is advantageous because it yields a grouped-time proportional hazards

This model is written as:

$$\log(-\log(1 - \mathrm{P}(D_i = j \mid D_i \geq j)))$$

$$= \alpha_{0j} + \alpha_1 h(\boldsymbol{y}_{ij}^{O}) + \boldsymbol{\alpha}_2 \boldsymbol{x}_{ij} + \boldsymbol{\alpha}_3 (h(\boldsymbol{y}_{ij}^{O}) \times \boldsymbol{x}_{ij})$$

- The only change to the model is that a different link function is specified, namely clog-log instead of logit

- In practice, it often doesn't matter greatly if the logit or clog-clog link is selected

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

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

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
- almost all subjects were measured at baseline (434 of 437) and no subjects were only measured at baseline (*i.e.*, all subjects had at least one post-baseline measurement)
- some intermittent missingness, but dropout was a much more common pattern of missingness
- 102 of 437 subjects did not complete the trial (*i.e.*, were not measured at week 6)

Crosstabulation of treatment group (denoted `Drug`) by last measurement 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 $p < .0013$

MCAR Test: Schizophrenia Person-Period Dataset

- `Maxweek` $= D_i$ (time to dropout; last measured timepoint)
- possible dropout times were weeks 1 to 5
- a subject could have up to five records in the dataset
  - if a subject completed the trial (`Maxweek` $= 6$), they contributed five records with $y_{ij} = 0$ for all five
  - if a subject was measured at week 5 but not at week 6 (`Maxweek` $= 5$), they contributed five records with $y_{i1} = y_{i2} = y_{i3} = y_{i4} = 0$ and $y_{i5} = 1$
  - if a subject was last measured at week 1 (`Maxweek` $= 1$), they only contributed a single record with $y_{i1} = 1$

Covariates included:

`Week` (4 indicator variables for the 5 timepoints)
`Drug` (indicator of placebo =0 or drug = 1)
`MeanY` (a subject's cumulative mean of the observed values of the dependent variable at a given timepoint)

and interactions

Example: Analysis of NIMH Schizophrenia dataset for time to dropout using discrete-time survival analysis. Shows how to create the person-period dataset. (SAS code)

http://www.uic.edu/classes/bstt/bstt513/schz_mcartest.sas.txt

http://www.uic.edu/~hedeker/long.html/schz_mcartest.sas.txt

```
DATA one; INFILE 'c:\schizrep.dat'; INPUT id imps79 week tx 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)
tx 0=placebo 1=drug (chlorpromazine, fluphenazine, or thioridazine)
sex 0=female 1=male
*/

PROC FORMAT; VALUE tx 0 = 'placebo' 1 = 'drug';

/* calculate the maximum value of WEEK for each subject */
PROC MEANS NOPRINT; CLASS id; VAR week tx;
OUTPUT OUT=two MAX(week tx) = maxweek drug;
RUN;

DATA three; SET two; IF id NE .;

/* crosstab of treatment group by maxweek */
PROC FREQ;
TABLES drug*maxweek / CHISQ;
FORMAT drug tx.;
RUN;
```

```
/* set up data in multivariate form to get subject-level variables */
DATA t0 (KEEP = id y0); SET one; IF week EQ 0; y0 = imps79;
DATA t1 (KEEP = id y1); SET one; IF week EQ 1; y1 = imps79;
DATA t2 (KEEP = id y2); SET one; IF week EQ 2; y2 = imps79;
DATA t3 (KEEP = id y3); SET one; IF week EQ 3; y3 = imps79;
DATA t4 (KEEP = id y4); SET one; IF week EQ 4; y4 = imps79;
DATA t5 (KEEP = id y5); SET one; IF week EQ 5; y5 = imps79;
DATA t6 (KEEP = id y6); SET one; IF week EQ 6; y6 = imps79;

DATA four; MERGE t0 t1 t2 t3 t4 t5 t6 three; BY id;

/* create cumulative average of dependent variable */
mean0 = y0;
mean1 = MEAN(y0,y1);
mean2 = MEAN(y0,y1,y2);
mean3 = MEAN(y0,y1,y2,y3);
mean4 = MEAN(y0,y1,y2,y3,y4);
mean5 = MEAN(y0,y1,y2,y3,y4,y5);
```

```
/* setting up data for survival analysis of time to dropout */
drop1 = .;drop2 = .; drop3 = .; drop4 = .; drop5 = .;

IF maxweek = 1 THEN
DO; drop1 = 1; END;

IF maxweek = 2 THEN
DO; drop1 = 0; drop2 = 1; END;

IF maxweek = 3 THEN
DO; drop1 = 0; drop2 = 0; drop3 = 1; END;

IF maxweek = 4 THEN
DO; drop1 = 0; drop2 = 0; drop3 = 0; drop4 = 1; END;

IF maxweek = 5 THEN
DO; drop1 = 0; drop2 = 0; drop3 = 0; drop4 = 0; drop5 = 1; END;

IF maxweek = 6 THEN
DO; drop1 = 0; drop2 = 0; drop3 = 0; drop4 = 0; drop5 = 0; END;
```

```
/* get the data in univariate structure for grouped-time survival analysis */
FILE 'c:\schizdrop.dat';
PUT id drop1 ' 1 0 0 0 ' drug mean1
  / id drop2 ' 0 1 0 0 ' drug mean2
  / id drop3 ' 0 0 1 0 ' drug mean3
  / id drop4 ' 0 0 0 1 ' drug mean4
  / id drop5 ' 0 0 0 0 ' drug mean5;
RUN;

/* read in data for grouped-time survival analysis of time to dropout */
DATA DROPS; INFILE 'c:\schizdrop.dat';
INPUT id droptime t1 t2 t3 t4 drug meany;
RUN;
```

```
/* CLOG-LOG regression - proportional hazards models */
/* main effects model with time and drug */
PROC LOGISTIC DESCENDING;
MODEL droptime = t1 t2 t3 t4 drug meany / LINK = CLOGLOG;

/* adding week by drug interaction */
PROC LOGISTIC DESCENDING;
MODEL droptime = t1 t2 t3 t4 drug meany t1*drug t2*drug t3*drug t4*drug
      / LINK = CLOGLOG;

/* adding drug by meany interaction */
PROC LOGISTIC DESCENDING;
MODEL droptime = t1 t2 t3 t4 drug meany t1*drug t2*drug t3*drug t4*drug
      drug*meany / LINK = CLOGLOG;

/* adding week by meany interaction */
PROC LOGISTIC DESCENDING;
MODEL droptime = t1 t2 t3 t4 drug meany t1*drug t2*drug t3*drug t4*drug
      drug*meany meany*t1 meany*t2 meany*t3 meany*t4 / LINK = CLOGLOG;

/* adding 3-way interaction */
PROC LOGISTIC DESCENDING;
MODEL droptime = t1 t2 t3 t4 drug meany t1*drug t2*drug t3*drug t4*drug
      drug*meany meany*t1 meany*t2 meany*t3 meany*t4 drug*meany*t1
      drug*meany*t2 drug*meany*t3 drug*meany*t4 / LINK = CLOGLOG;
RUN;
```

Clog-log Regression - Time to Dropout Models

covariates	p	-2 log L
`Week`, `Drug`, `MeanY`	7	729.44
+ `Week` × `Drug`	11	728.13
+ `Drug` × `MeanY`	12	706.77
+ `Week` × `MeanY`	16	700.50
+ `Week` × `Drug` × `MeanY`	20	697.71

p = number of regression coefficients
(including one for the intercept)

⇒ `Drug` × `MeanY` is highly significant

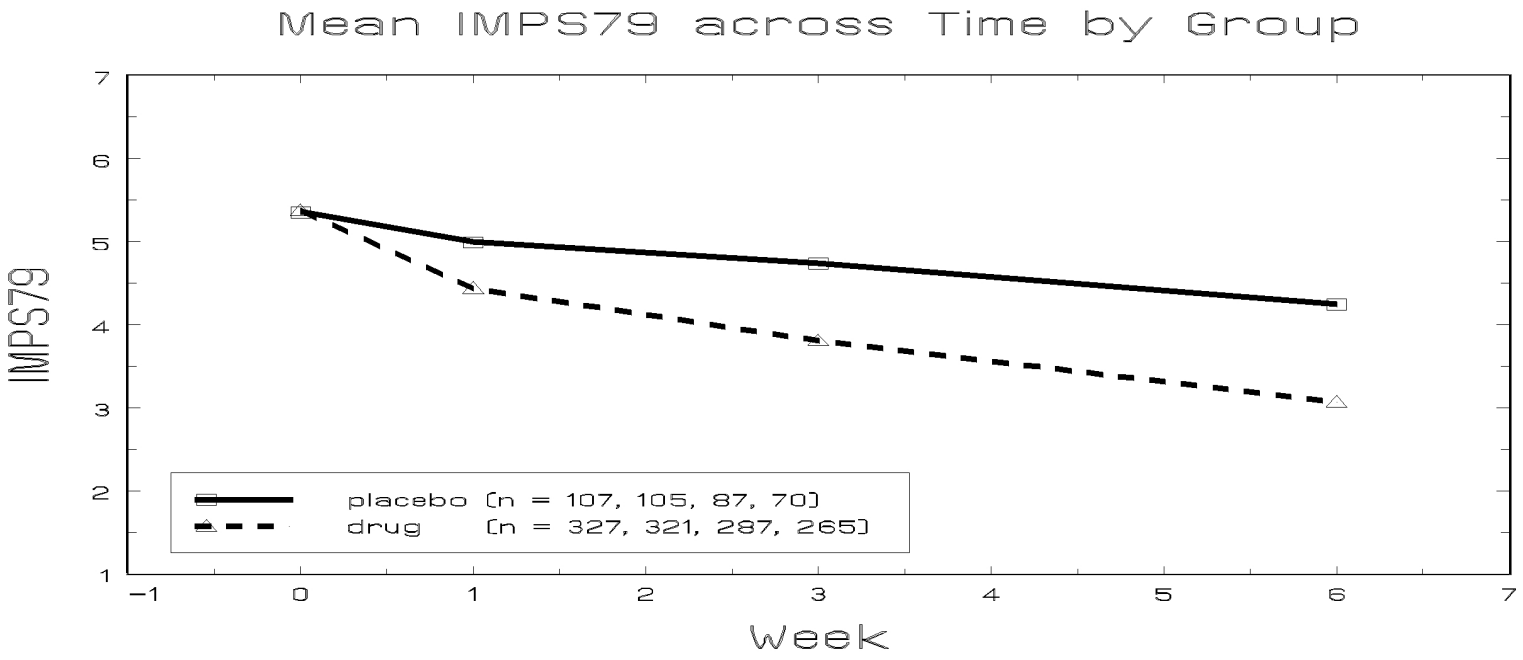
Final time to dropout model

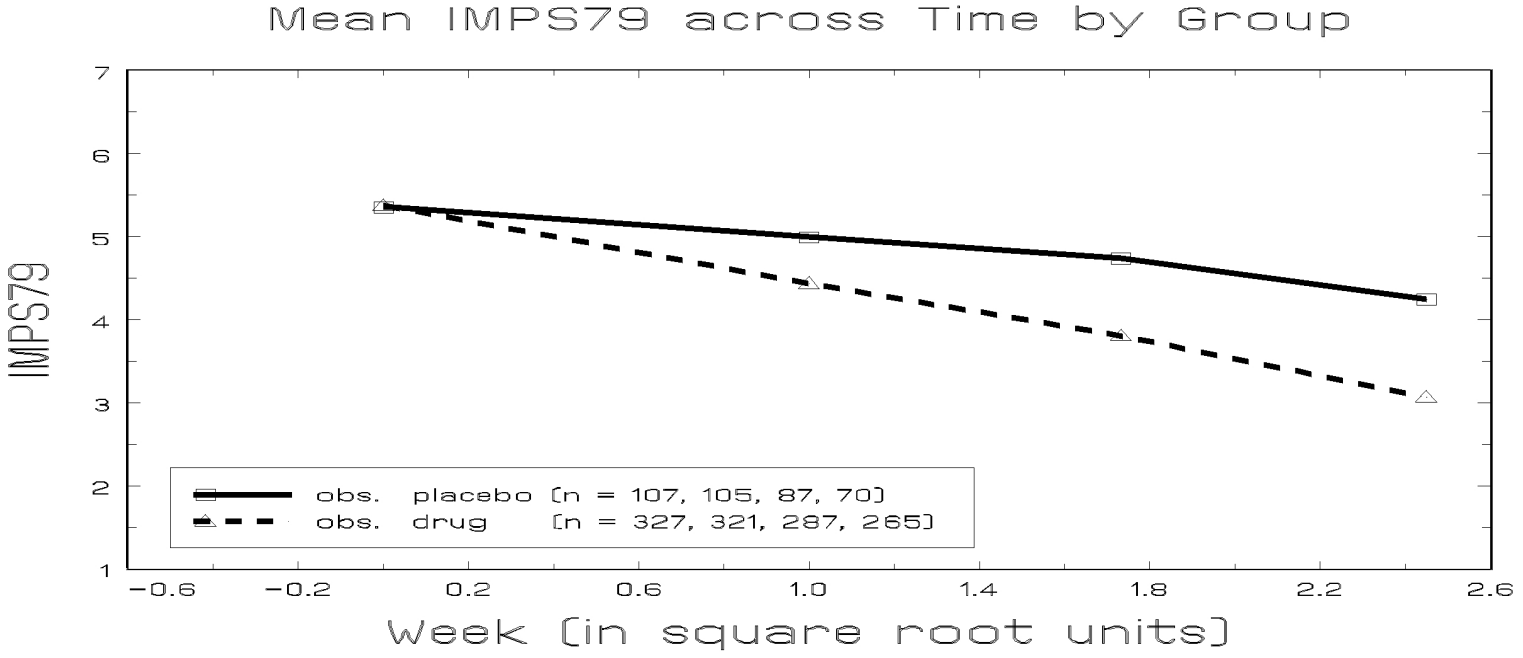
(`Week, Drug, Meany, Drug` $\times$ `MeanY`)

term	ML estimate	se	p-value
`Intercept`	-6.573	1.208	.0001
`Week 1`	1.327	.393	.0007
`Week 2`	0.096	.476	.84
`Week 3`	1.549	.386	.0001
`Week 4`	-0.494	.570	.39
`Drug`	4.765	1.297	.0002
`MeanY`	0.635	.214	.003
`Drug` $\times$ `MeanY`	-1.108	.249	.0001

- placebo: exp(.635) = 1.89; a unit *increase* in `MeanY` nearly doubles hazard of dropout
- drug: .635 - 1.108 = -.473, exp(-.473) = .623, inverted equals 1.60; a unit *decrease* in `MeanY` nearly doubles dropout hazard

$\Rightarrow$ *MCAR is rejected* (aside: in main effects model, `MeanY` is not significant, which would suggest that MCAR is reasonable!)


Mean IMPS79 across Time by Group
IMPS79
Week
placebo [n = 107, 105, 87, 70]
drug [n = 327, 321, 287, 265]
1
2
3
4
5
6
7
−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]
1
2
3
4
5
6
7
−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

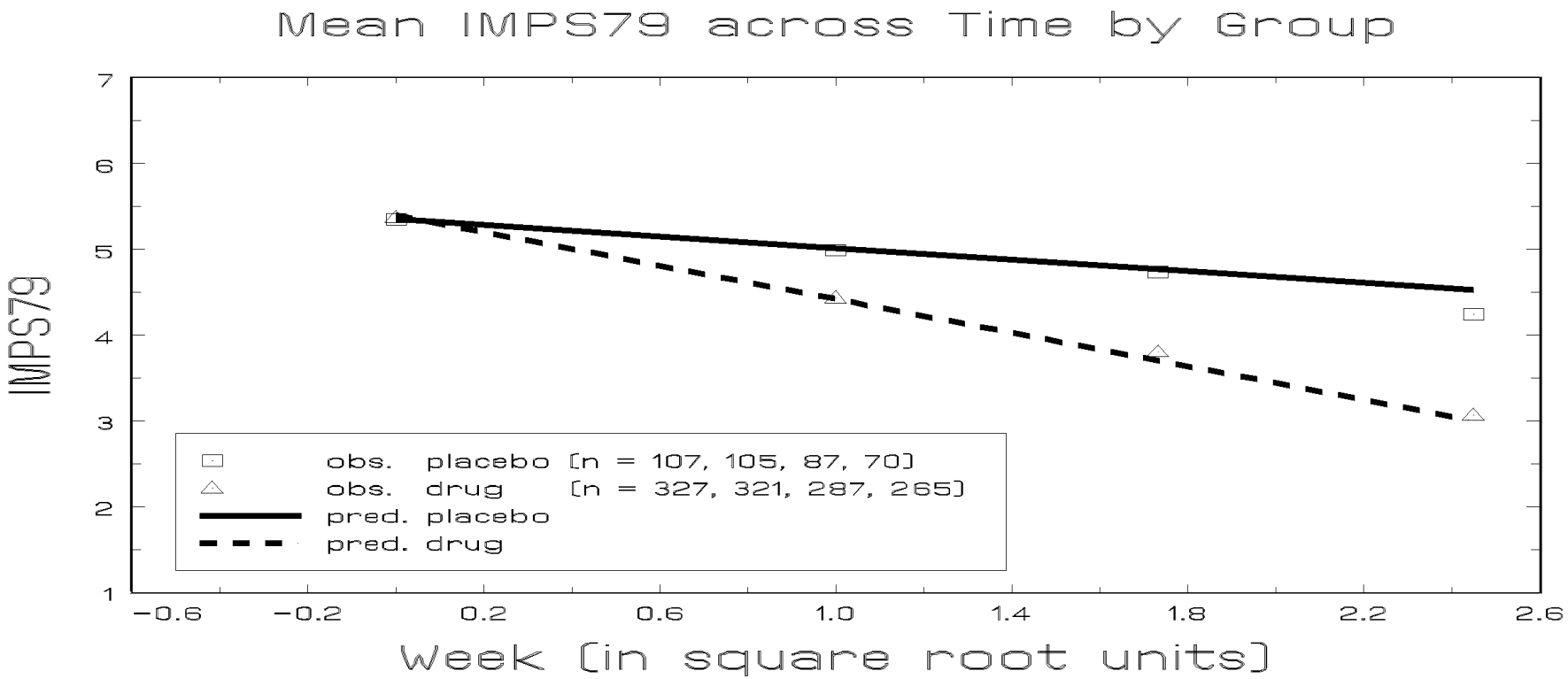
$\texttt{Drug} = 0$ for placebo, 1 for drug

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

NIMH Schizophrenia Study - IMPS79 across Time:
GEE and MRM Estimates (se)

	GEE UN			*Mixed Model*		
	est.	se	$p <$	est.	se	$p <$
intercept	5.368	0.087	.001	5.348	0.088	.001
Drug (0 = plc; 1 = drug)	0.029	0.101	.78	0.046	0.101	.65
Time (sqrt week)	-0.376	0.065	.001	-0.336	0.068	.001
Drug by Time	-0.604	0.075	.001	-0.641	0.078	.001
Intercept variance				0.369	0.060	
				sd =.61		
Int-Time covariance				0.021	0.034	
				r =.07		
Time variance				0.242	0.032	
				sd =.49		

Fitted and Obs. Means across Time by Condition



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

SAS code: GEE and MRM models

```
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
*/

/* GEE Model:  Unstructured */
PROC GENMOD;
CLASS id week;
MODEL imps79 = tx sweek txswk / DIST=NORMAL;
REPEATED SUBJECT=id / WITHIN=week CORRW TYPE=UN;
RUN;

/* Mixed Regression Model:  Random Trend */
PROC MIXED METHOD=ML COVTEST;
CLASS id;
MODEL imps79 = tx sweek txswk / SOLUTION;
RANDOM INT sweek / SUBJECT=id TYPE=UN G GCORR;
RUN;
```

Missing Not At Random (MNAR) Models

- When the data are nonignorable (*i.e.*, MNAR), standard statistical models can yield badly biased results
- In the simulation we know that the data are nonignorable, because they were created that way, however considering a real dataset, one never knows whether the data are ignorable or not
- The observed data provide no information to either confirm or refute ignorability

$\Rightarrow$ **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

$\Rightarrow$ 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

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 can be used ("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:

pattern group	time1	time2	time3	"recoded" 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

Classification of Subjects based on missing-data

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

Drug group	Drop group completer	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?

Mixed-effects pattern mixture model: Schiz data

augment the basic MRM of `IMPS79` over time:

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

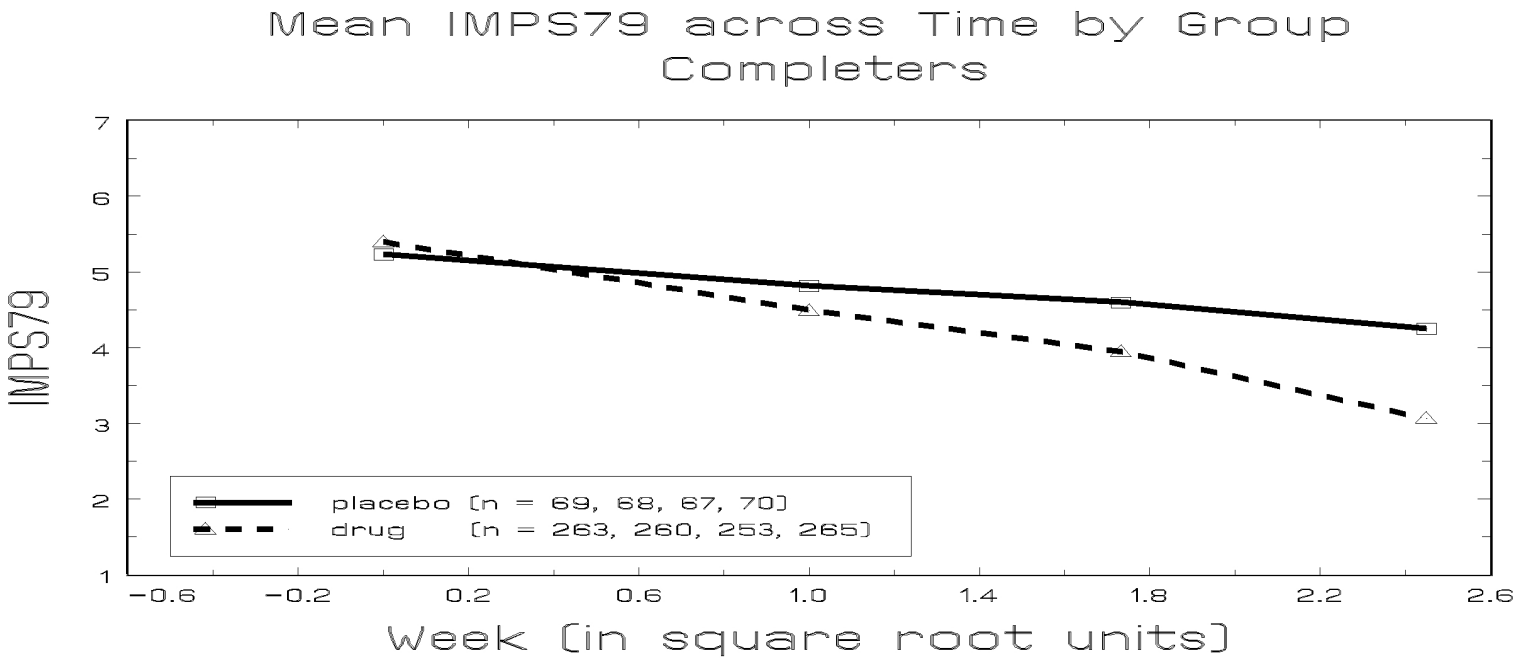
with variables based on the missing data patterns

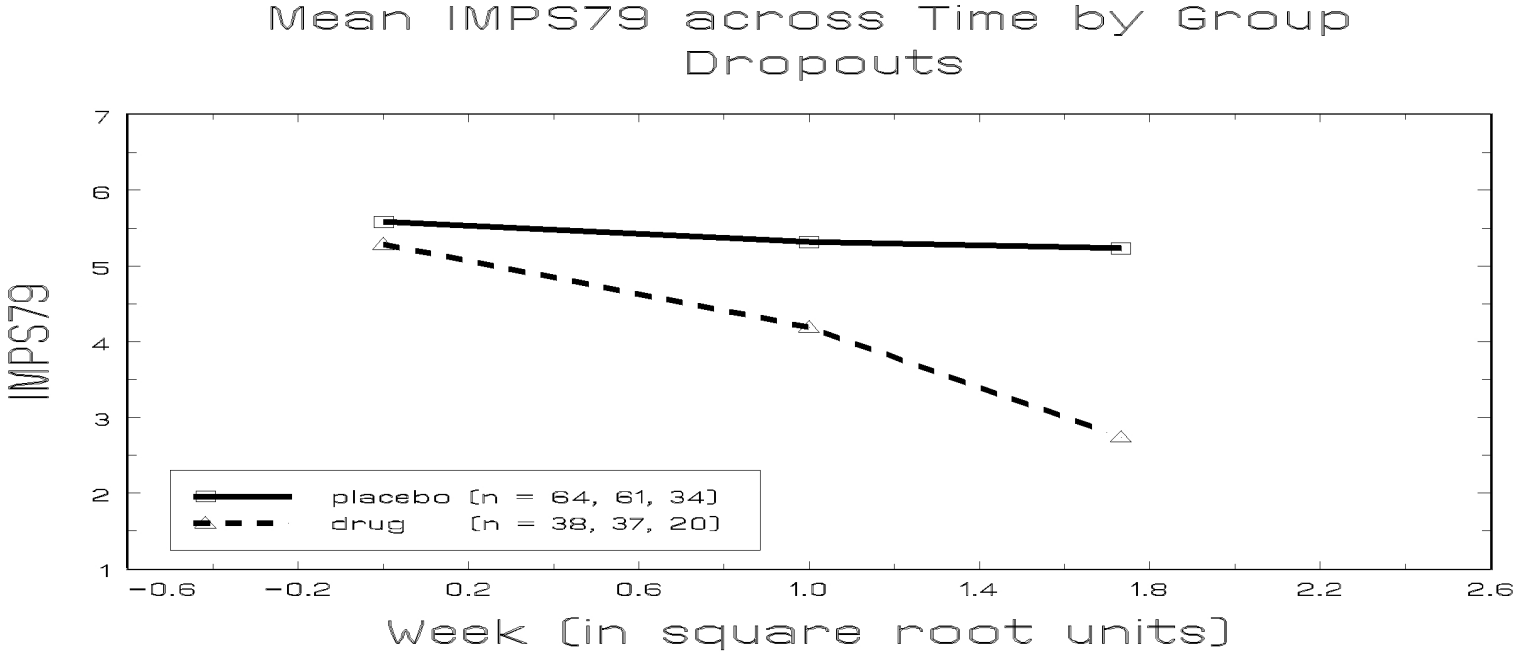
e.g., completers (N = 335) vs non-completers (N = 102)

`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)

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

- 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
7
6
5
4
3
2
1
placebo [n = 69, 68, 67, 70]
drug [n = 263, 260, 253, 265]
−0.6
−0.2
0.2
0.6
1.0
1.4
1.8
2.2
2.6
Week [in square root units]


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

Less Restrictive Pattern Mixture Model

- use week of dropout variable $\mathtt{D}_i$ in forming missing data patterns
- six missing data patterns: five dropout weeks and completers
- Let $\mathtt{D}_m = \mathtt{D}_1, \ldots, \mathtt{D}_5$ denote dummy-variables which contrast each dropout pattern to the completers

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

- β_0, β_1, β_2, and β_3 are for completers
- β_0^m, β_1^m, β_2^m, and β_3^m indicate how dropout group m differs from completers
- the β_3^m parameters are of great interest

	MRM			pattern-mixture MRMs					
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` $\times$ `SW` β_3	-.641	.078	.0001	-.539	.086	.0001	-.539	.085	.0001
Dropout				`Drop` $= 1$ $(N = 102)$			`D` $= 1$ $(N = 37)$		
`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` $\times$ `SW` β_3^1				-.635	.196	.002	-.412	.412	.32
							`D` $= 2$ $(N = 10)$		
`Int` β_0^2							.524	.437	.23
`Drug` β_1^2							-.703	.613	.25
`SWeek` β_2^2							.338	.398	.40
`Drug` $\times$ `SW` β_3^2							-.735	.562	.19
							`D` $= 3$ $(N = 42)$		
`Int` β_0^3							.047	.256	.85
`Drug` β_1^3							-.198	.318	.53
`SWeek` β_2^3							.377	.208	.07
`Drug` $\times$ `SW` β_3^3							-.835	.261	.002
							`D` $= 4$ $(N = 5)$		
`Int` β_0^4							.801	.653	.22
`Drug` β_1^4							-.237	.841	.78
`SWeek` β_2^4							-.101	.485	.84
`Drug` $\times$ `SW` β_3^4							-1.210	.625	.054
							`D` $= 5$ $(N = 8)$		
`Int` β_0^5							.337	.645	.60
`Drug` β_1^5							-.842	.746	.26
`SWeek` β_2^5							-.157	.466	.74
`Drug` $\times$ `SW` β_3^5							.231	.538	.67
Deviance	4649.0			4623.3			4607.8		

Model comparisons

4649.0 - 4607.8 = 41.2, df = 20, $p < .004$
4649.0 - 4623.3 = 25.7, df = 4, $p < .0001$
4623.3 - 4607.8 = 15.5, df = 16, $p < .49$

- both pattern mixture models fit better than ordinary MRM (this is NOT a test of ignorability)
- less-restrictive pattern mixture model does not fit statistically better than simpler pattern mixture model

Estimated Trends by Subgroups

Placebo Completers

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

Drug Completers

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

$\Rightarrow$ *difference in slope = -.54*

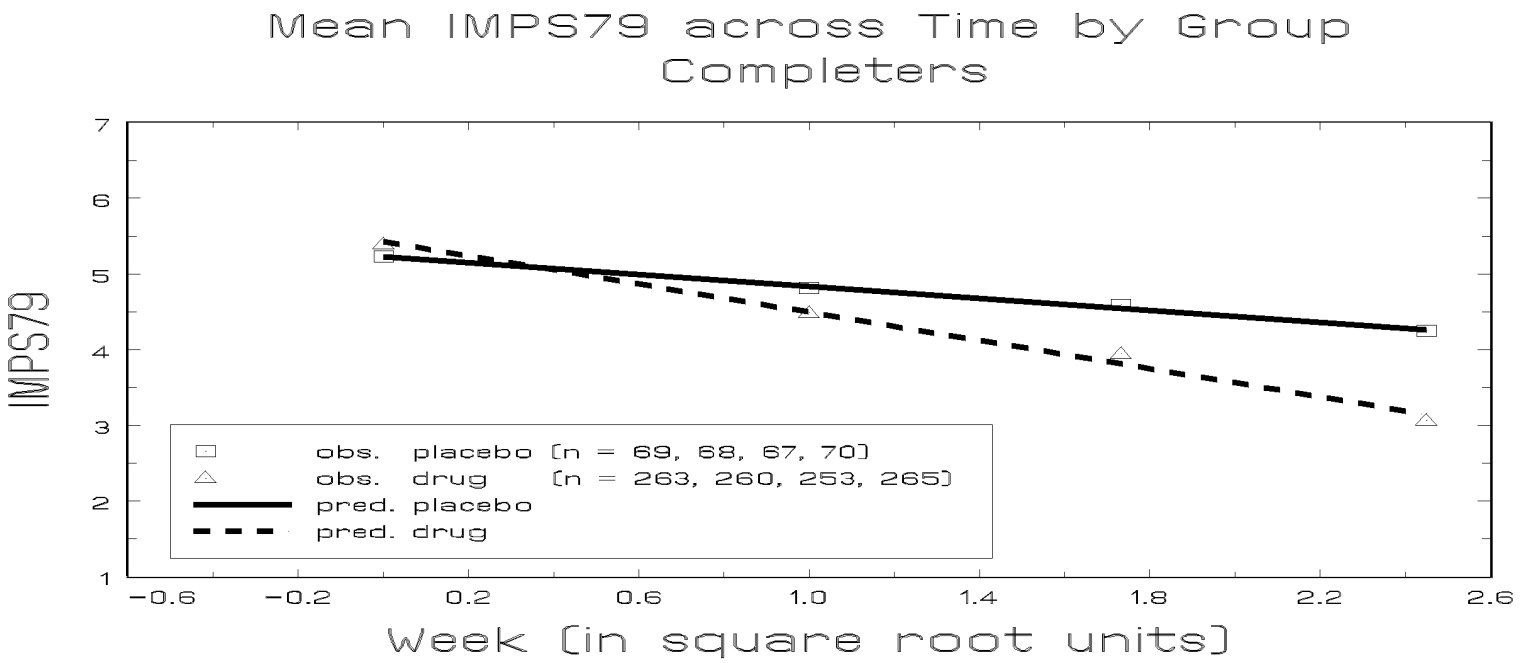
Placebo Dropouts

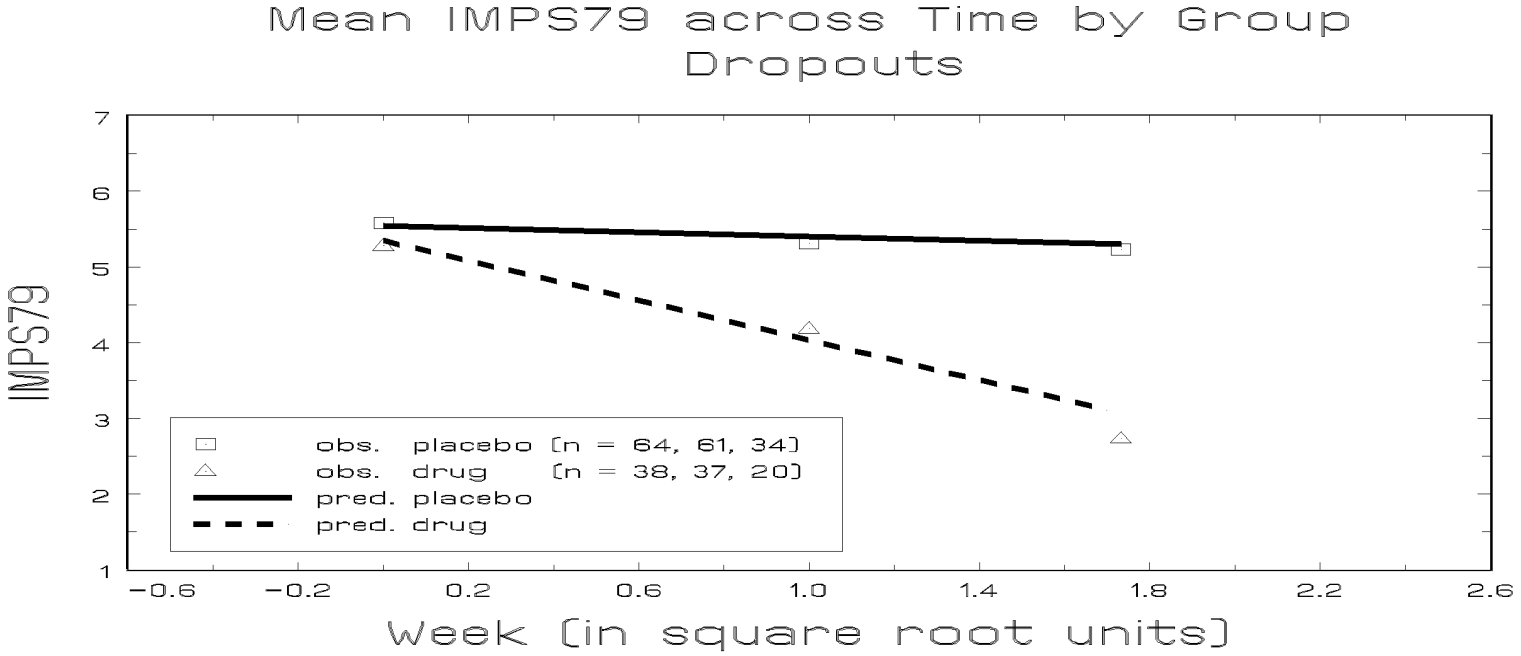
$$\hat{Y} = (5.22 + .32) - (.39 - .25)\,T \;=\; 5.54 - .14\,T$$

Drug Dropouts

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

$\Rightarrow$ *difference in slope = -1.17*
$\Rightarrow$ *difference in slope difference = -.64*
degree that Drug by Time interaction varies by Dropout


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


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

```
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 */
PROC MEANS NOPRINT; CLASS id; VAR week;
OUTPUT OUT=two MAX(week)=maxweek;
RUN;

/* determine if a subject has data at week 6
drop = 0 (for completers) or = 1 (for dropouts) */
DATA three; SET two;
drop=0;
IF maxweek LT 6 THEN drop=1;
```

```
/* dataset with all subjects (adding the drop variable) */
DATA four; MERGE one three; BY id;

/* pattern-mixture random intercept and trend model */
PROC MIXED DATA=four METHOD=ML COVTEST;
CLASS id;
MODEL imps79 = sweek drug sweek*drug drop drop*sweek
             drop*drug drop*drug*sweek / SOLUTION ;
RANDOM INTERCEPT sweek /SUB=id TYPE=UN G GCORR;
RUN;
```

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
- 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

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

- $\hat{\boldsymbol{\beta}}_c$ correspond to the coefficients in the current model formulation not involving dropout (*i.e.*, intercept, drug, time, drug by time)
- $\hat{\boldsymbol{\beta}}_\Delta = (\hat{\boldsymbol{\beta}}_d - \hat{\boldsymbol{\beta}}_c)$ 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 linear combinations of model estimates (obtained by simple arithmetic or via `ESTIMATE` statement in SAS)

Placebo Intercept

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

Placebo Time effect

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

Drug Intercept difference

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

Drug Time difference

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

SAS example using ESTIMATE

```
/* pattern-mixture random intercept and trend model
/* using marginal dropout proportion to estimate averaged results */

PROC MIXED METHOD=ML COVTEST;
CLASS id;
MODEL imps79 = sweek drug sweek*drug dropout dropout*sweek
        dropout*drug dropout*drug*sweek / SOLUTION;
RANDOM INTERCEPT sweek /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 sweek' 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 sweek*drug' INTERCEPT 0 sweek 0 drug 0 sweek*drug 1 dropout 0
        dropout*sweek 0 dropout*drug 0 dropout*drug*sweek .2334;
RUN;
```

where `dropout` =0 (completers) or =1 (dropouts) and 102/437 = .2334 (*i.e.*, dropout proportion)

Don't have (or like) SAS?

(*i.e.*, can I do pattern-mixture modeling with SPSS? .... YES!)

Weighted Effect Coding (Darlington, 1990, *Regression and Linear Models*, pp. 238-239): in ANOVA context

- yields comparisons of each cell mean, except the reference cell, with a weighted average of cell means

Effect coding

level	D
1	1
2	-1

becomes

Weighted effect coding

level	D^*
1	1
2	$-1 \times \frac{\text{weight of level 1}}{\text{weight of level 2}}$

$$Y = \beta_0 + \beta_1 D^* + e$$

β_0 = weighted average of Y means across the two levels

β_1 = difference between level 1 mean and the weighted average

SPSS example using weighted effect coding

Step 1 - do some data management to create the necessary variables

SCHIZREP.SAV

`ID`	subject's ID
`IMPS79`	severity of illness (dependent variable)
`Week`	study week (= 0, 1, 2, 3, 4, 5, 6)
`Drug`	=0 (placebo) or =1 (drug)
`SexM`	=0 (female) or =1 (male)
`SWeek`	sqrt of `Week`
`DrugSwk`	product of `Drug` by `Sweek`
`Week_max`	subject's maximum value of `Week`
`Dropout`	=0 (`Week_max` = 6) or =1 (`Week_max` < 6)
`DropW`	$= \frac{-\hat{\pi}_D}{\hat{\pi}_C} = \frac{-102}{335}$ (`Dropout` = 0) or =1 (`Dropout` = 1)
`DropWDrug`	product of `DropW` by `Drug`
`DropWSweek`	product of `DropW` by `SWeek`
`DropWDrugSwk`	product of `DropW` by `Drug` by `SWeek`

Model with weighted effect coding

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

- β_0, β_1, β_2, and β_3 are weighted averages over completers and dropouts (*i.e.*, exactly what we want!)
- β_0^{DW}, β_1^{DW}, β_2^{DW}, and β_3^{DW} how dropouts differ from the weighted averages

SPSS syntax for model with weighted effect coding

Schizpm.sps

```
MIXED
  Imps79 WITH Drug Sweek DrugSwk
              DropW DropWDrug DropWSweek DropWDrugSwk
  /FIXED = Drug Sweek DrugSwk
           DropW DropWDrug DropWSweek DropWDrugSwk
  /METHOD = ML
  /PRINT = SOLUTION TESTCOV
  /RANDOM INTERCEPT Sweek | SUBJECT(ID) COVTYPE(UN) .
```

Pattern-mixture averaged results

using either approach (ESTIMATE in SAS or weighted effect coding in SPSS), we get:

parameter	estimate	std error	p-value
`Intercept`	5.2958	.0898	.0001
`Drug`	.1086	.1029	.29
`SWeek`	-.3346	.0670	.0001
`DrugSwk`	-.6868	.0776	.0001

Great! We're done, right?

I've got some good news, and some not-so-good news

averaged estimates are fine

standard errors are not exactly correct (sample proportions are estimated, yet we treated these as known quantities)

Can use Delta Method to obtain standard errors

- uncertainty in model estimates
- uncertainty in using sample proportions as estimates

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,

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

under marginal model for completion ($i.e.$, = binomial(π_d))

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

Standard Errors for Averaged Estimates

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

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

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, `ESTIMATE` statement in SAS)

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

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

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

$\hat{\bar{\beta}}$ and $\hat{V}(\hat{\bar{\beta}})_F$ obtained from `ESTIMATE`

$\hat{\beta}_\Delta$ obtained from pattern-mixture model

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

Pattern-mixture averaged results - 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

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

```
/* 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 = sweek drug sweek*drug dropout dropout*sweek
        dropout*drug dropout*drug*sweek / SOLUTION COVB;
RANDOM INTERCEPT sweek /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 sweek' 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 sweek*drug' INTERCEPT 0 sweek 0 drug 0 sweek*drug 1
        dropout 0 dropout*sweek 0
        dropout*drug 0 dropout*drug*sweek .19453;
RUN;
```

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

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

$\hat{\bar{\beta}}$ and $\hat{V}(\hat{\bar{\beta}})_F$ obtained from `ESTIMATE`

$\hat{\beta}_\Delta$ obtained from pattern-mixture model

Augment $= \frac{n_d n_c}{N^3}\hat{\beta}^2_\Delta$, where
$\frac{n_d n_c}{N^3} = \frac{38\times70}{(108)^3} = .00211159$ for placebo
$\frac{n_d n_c}{N^3} = \frac{64\times265}{(329)^3} = .00047625$ for drug

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 and 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 $f_y(\boldsymbol{y}_i \mid \boldsymbol{v})$ represent the conditional model for the longitudinal outcome $\boldsymbol{y}_i$ given the random subject effects $\boldsymbol{v}$

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

Dropout model - grouped/discrete time survival analysis

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

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

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

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

Marginal likelihood for a subject for both components is

$$f(\boldsymbol{y}_i, D_i) = \int_{\boldsymbol{\upsilon}} f_y(\boldsymbol{y}_i \mid \boldsymbol{\upsilon}) f_D(D_i \mid \boldsymbol{\upsilon}) f(\boldsymbol{\upsilon}) d\boldsymbol{\upsilon}$$

where $f(\boldsymbol{\upsilon})$ is the random-effects distribution

- convenient to standardize the random effects, $\boldsymbol{\upsilon}_i = \boldsymbol{S}\boldsymbol{\theta}_i$
- $\boldsymbol{\theta}_i$ is distributed as a multivariate standard normal
- use Cholesky factorization, $\boldsymbol{\Sigma}_{\upsilon} = \boldsymbol{S}\boldsymbol{S}'$

The likelihood for a given individual is now of the form:

$$f(\boldsymbol{y}_i, D_i) = \int_{\boldsymbol{\theta}} f_y(\boldsymbol{y}_i \mid \boldsymbol{\theta}) f_D(D_i \mid \boldsymbol{\theta}) f(\boldsymbol{\theta}) d\boldsymbol{\theta}$$

The marginal likelihood for the sample of N subjects is then given by the sum $\log L = \sum_i^N \log f(\boldsymbol{y}_i, D_i)$, which is maximized to obtain the maximum likelihood solution

Mixed-effects selection model - Schiz study

Longitudinal model:

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

becomes

$$\begin{aligned}\texttt{IMPS79}_{ij} &= \beta_0 + \beta_1 \texttt{Drug}_i + \beta_2 \texttt{SWeek}_j + \beta_3(\texttt{Drug}_i \times \texttt{SWeek}_j) \\ &\quad + \left(\sigma_{\upsilon_0} + \left(\sigma_{\upsilon_{01}}/\sigma_{\upsilon_0}\right)\texttt{SWeek}_j\right)\theta_{0i} \\ &\quad + \left(\sqrt{\sigma^2_{\upsilon_1} - \sigma^2_{\upsilon_{01}}/\sigma^2_{\upsilon_0}}\,\texttt{SWeek}_j\right)\theta_{1i}\end{aligned}$$

because $\boldsymbol{\upsilon}_i = \boldsymbol{S}\boldsymbol{\theta}_i$, where $\boldsymbol{\Sigma}_\upsilon = \boldsymbol{S}\boldsymbol{S}'$

$$\boldsymbol{S} = \begin{bmatrix} s_0 & 0 \\ s_{01} & s_1 \end{bmatrix} = \begin{bmatrix} \sigma_{\upsilon_0} & 0 \\ \sigma_{\upsilon_{01}}/\sigma_{\upsilon_0} & \sqrt{\sigma^2_{\upsilon_1} - \sigma^2_{\upsilon_{01}}/\sigma^2_{\upsilon_0}} \end{bmatrix}$$

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

$$\begin{aligned} \upsilon_{0i} + \upsilon_{1i}\texttt{SWeek}_j &= (s_0 + s_{01}\texttt{SWeek}_j)\,\theta_{0i} + (s_1\texttt{SWeek}_j)\,\theta_{1i} \\ &= (\sigma_{\upsilon_0} + (\sigma_{\upsilon_{01}}/\sigma_{\upsilon_0})\,\texttt{SWeek}_j)\,\theta_{0i} \\ &\quad + \left(\sqrt{\sigma^2_{\upsilon_1} - \sigma^2_{\upsilon_{01}}/\sigma^2_{\upsilon_0}}\,\texttt{SWeek}_j\right)\theta_{1i} \end{aligned}$$

Dropout model (clog-log ordinal regression model):

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

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

- the 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*
- it is 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;
```

Notes:

- uppercase letters represent specific SAS syntax; lowercase letters represent user-defined information
- the SAS dataset `all` includes the $(n_i + 1) \times 1$ outcome vector $\boldsymbol{y}_i^*$, named `outcome`, which contains $\boldsymbol{y}_i$ 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 $\boldsymbol{y}_i$ 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

	Separate			Shared		
parameter	ML est	std error	p-value	ML est	std error	p-value
Outcome						
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
Dropout						
`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	

- the one separate parameter model yields identical parameter estimates and standard errors as running these two models, one for $\boldsymbol{y}_i$ and one for D_i, separately
- shared parameter model fits better, $\chi^2_4 = 30.1, p < .0001$
- 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

NIMH Schizophrenia Study (N = 437; $\Sigma\, n_i = 1603$)
ML Estimates (se) *random intercept and slope models*

	ordinary MRM	shared parameter MRM	pattern mixture MRM
intercept	5.348	5.320	5.334
	(.088)	(.088)	(.089)
Drug *(0=p; 1=d)*	0.046	0.088	0.124
	(.101)	(.102)	(.105)
Time *(sqrt wk)*	-0.336	-0.272	-0.305
	(.068)	(.073)	(.071)
Drug by Time	-0.641	-0.737	-0.662
	(.078)	(.083)	(.078)

Conclusions

- MRM 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 y

- MRM selection (*i.e.*, shared parameter) models
  - models longitudinal and missing processes simultaneously
  - indicates predictors of missingness
  - adjusts longitudinal model for missingness
  - does not assume MAR

- MRM pattern-mixture models
  - 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
  - does not assume MAR

$\Rightarrow$ These approaches do not invent data; attempt to maximize information obtained from *available* data