Mixed-Effects Models for Categorical Outcomes

Don Hedeker
University of Illinois at Chicago

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

Mixed-effects models for categorical outcomes

- dichotomous outcomes
  - mixed-effects logistic regression

- ordinal outcomes
  - mixed-effects ordinal logistic regression
    * proportional odds model
    * partial or non-proportional odds model

- discrete or grouped time-to-event data
  - mixed-effects dichotomous or ordinal regression
  - replace logistic link with complementary log-log link to yield proportional (and non-proportional) hazards models

Logistic Regression Model

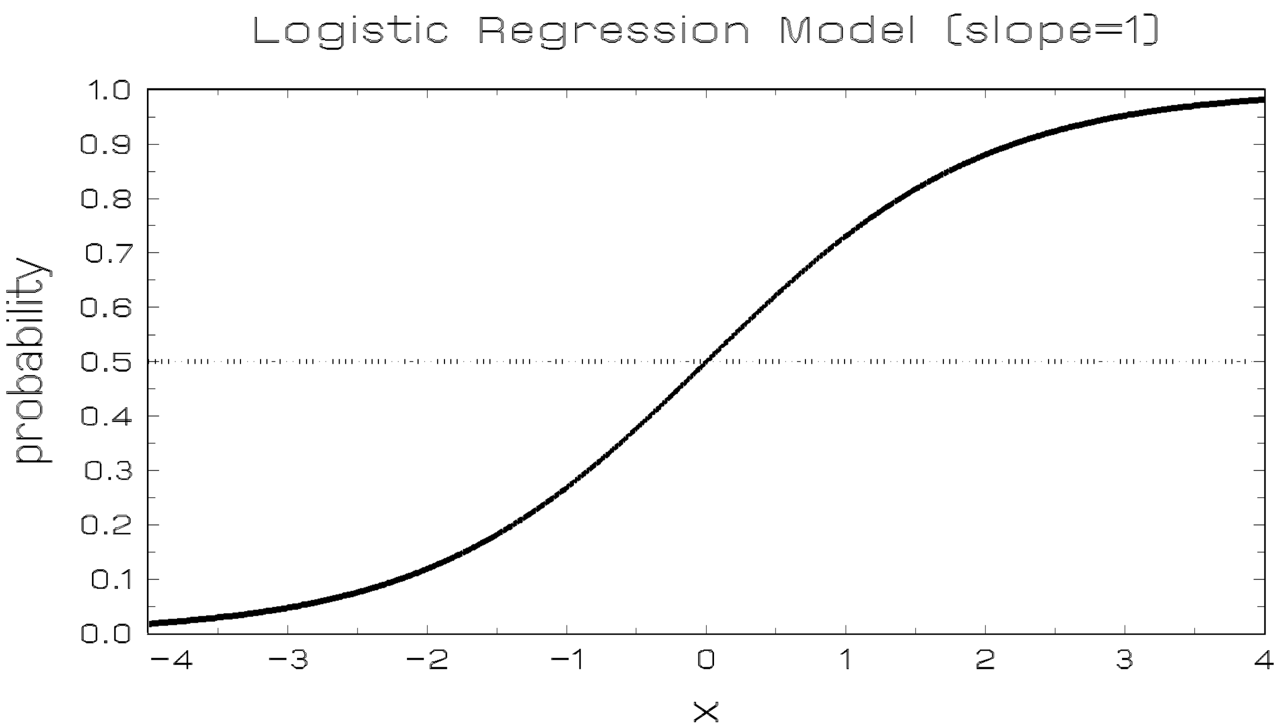
$$\log\left[\frac{P(\mathsf{y}_i = 1)}{1 - P(\mathsf{y}_i = 1)}\right] = \boldsymbol{x}_i'\boldsymbol{\beta}$$

- Dichotomous outcome ($\mathsf{y} = 0$ absence, $\mathsf{y} = 1$ presence).
- Function that links probabilities to regressors is the logit (or log odds) function $\log\left[P/(1-P\right]$. Logit is called the link function.

The model can be written in terms of probabilities:

$$P(\mathsf{y}_i = 1) = \frac{1}{1 + \exp(-\boldsymbol{x}_i'\boldsymbol{\beta})}$$

- Model is a linear model for the logits, not for the probabilities. Logits can take on any values between negative and positive infinity, probabilities can only take on values between 0 and 1.


Logistic Regression Model (slope=1)
probability
1.0
0.9
0.8
0.7
0.6
0.5
0.4
0.3
0.2
0.1
0.0
−4
−3
−2
−1
0
1
2
3
4
X


Logistic Regression Model (slope=1)
logit
4
3
2
1
0
−1
−2
−3
−4
−4
−3
−2
−1
0
1
2
3
4
X

The model can also be written in terms of the odds:

$$\frac{P(\mathsf{y}_i = 1)}{1 - P(\mathsf{y}_i = 1)} = \exp(\boldsymbol{x}_i'\boldsymbol{\beta})$$

$\exp\beta$ = change in odds for y per unit change of x

- $\beta = 0$ yields no effect on the odds
- $\beta > 0$ increases odds y is present with increasing x
- $\beta < 0$ decreases odds y is present with increasing x

It is interesting to compare the odds between two levels of x, *e.g.*, x is a dummy-coded variable indicating group (=1) or control (=0) intervention. The odds ratio for the group versus control condition is:

$$\frac{\exp(\beta_0 + \beta_1 x)}{\exp(\beta_0)} = \frac{\exp(\beta_0)\exp(\beta_1 x)}{\exp(\beta_0)} = \exp(\beta_1)$$

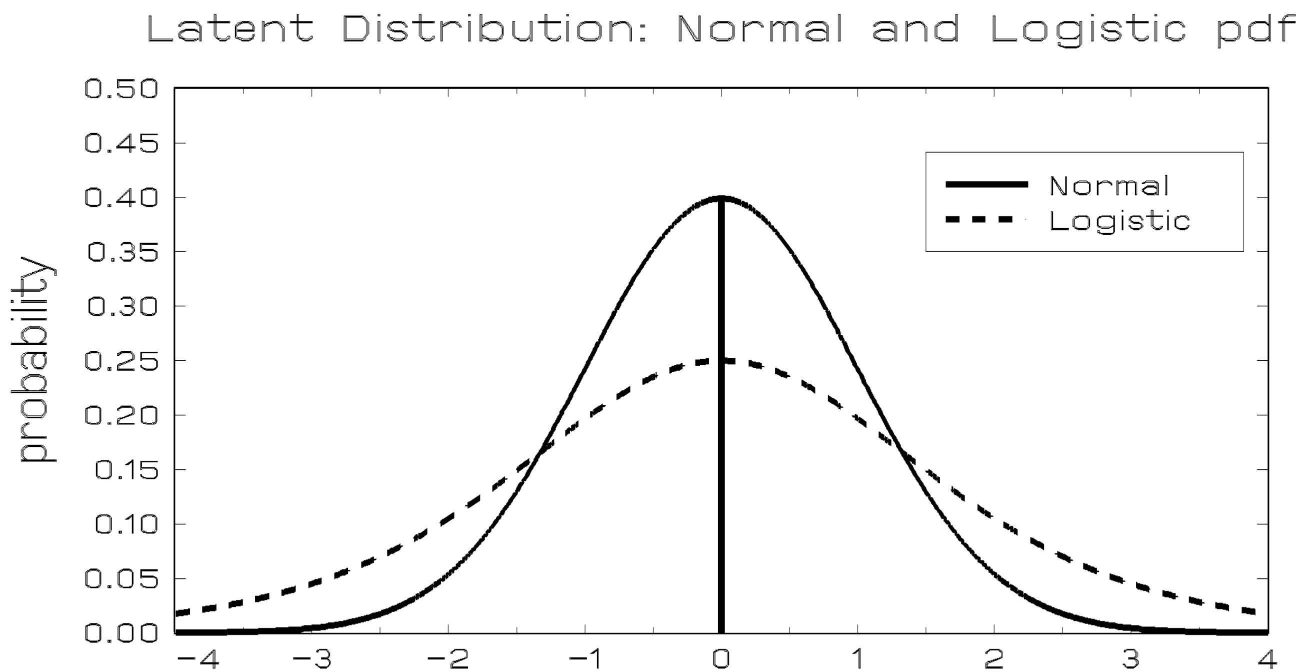
More generally, comparing $x_1 = a$ versus $x_1 = b$:

$$\exp\left[\beta_1(a - b)\right]$$

Dichotomous Response and Threshold Concept

Continuous y_i - an unobservable latent variable - related to dichotomous response y_i via "threshold concept"

Response occurs ($\mathsf{y}_i = 1$) if $\gamma < y_i$
otherwise, a response does not occur ($\mathsf{y}_i = 0$)



The Threshold Concept in Practice

"How was your day?" (what is your satisfaction level today?)

- Satisfaction may be continuous, but we usually emit a dichotomous response:

Model for Latent Continuous Responses

Consider the model with p covariates for the latent response strength y_i ($i = 1, 2, \ldots, N$):

$$y_i = \boldsymbol{x}_i'\boldsymbol{\beta} + \varepsilon_i$$

- probit: $\varepsilon_i \sim$ standard normal (mean=0, variance=1)
- logistic: $\varepsilon_i \sim$ standard logistic (mean=0, variance=$\pi^2/3$)

$\Rightarrow$ $\boldsymbol{\beta}$ estimates from logistic regression are larger (in abs. value) than from probit regression by approximately $\sqrt{\pi^2/3} = 1.8$

Underlying latent variable

- useful way of thinking of the problem
- not an essential assumption of the model

Random-intercept Logistic Regression Model

Consider the model with p covariates for the response y_{ij} for subject j ($j = 1, 2, \ldots, n_i$) in cluster i ($i = 1, 2, \ldots, N$):

$$\log\left[\frac{P(\mathsf{y}_{ij}=1)}{1-P(\mathsf{y}_{ij}=1)}\right] = \boldsymbol{x}'_{ij}\boldsymbol{\beta} + \upsilon_{0i}$$

where

$$\begin{aligned}
\mathsf{y}_{ij} &= \text{dichotomous response for subject } j \text{ in cluster } i \\
\boldsymbol{x}_{ij} &= (p+1) \times 1 \text{ covariate vector (includes 1 for intercept)} \\
\boldsymbol{\beta} &= (p+1) \times 1 \text{ vector of unknown parameters} \\
\upsilon_{0i} &= \text{cluster effects distributed } \mathcal{NID}(0, \sigma_\upsilon^2)
\end{aligned}$$

Characteristics of $v_{0i} \sim \mathcal{NID}(0, \sigma_v^2)$

- separates model from ususal (fixed-effects) multiple logistic regression model
- takes on $i = 1, 2, \ldots, N$ values
- assess impact of cluster i on individual outcome $logit_{ij}$, represents effect of subject clustering
- common for each cluster member, but changes for each cluster
- if $v_{0i} = 0$, then cluster has no effect for cluster i
- if $v_{0i} = 0$ for all clusters, cluster structure has no impact on individual data ($\sigma_v^2 = 0$)
  - no need for multilevel approach
  - ordinary logistic regression is OK
- if subject clustering has strong effect, estimates of $v_{0i} \neq 0$ and σ_v^2 will increase from 0

Model for Latent Continuous Responses

Consider the model with p covariates for the $n_i \times 1$ latent response strength y_{ij}:

$$y_{ij} = \boldsymbol{x}'_{ij}\boldsymbol{\beta} + \upsilon_{0i} + \varepsilon_{ij}$$

where assuming

- $\varepsilon_{ij} \sim$ standard normal (mean 0 and $\sigma^2 = 1$) leads to multilevel probit regression
- $\varepsilon_{ij} \sim$ standard logistic (mean 0 and $\sigma^2 = \pi^2/3$) leads to multilevel logistic regression

Underlying latent variable

- not an essential assumption of the model
- useful for obtaining intra-class correlation (r)

$$r = \frac{\sigma_v^2}{\sigma_v^2 + \sigma^2}$$

and for design effect (d)

$$d = \frac{\sigma_v^2 + \sigma^2}{\sigma^2} = 1/(1 - r)$$

ratio of actual variance to the variance that would be obtained by simple random sampling (holding sample size constant)

Scaling of regression coefficients

Fixed-effects model

$\boldsymbol{\beta}$ estimates from logistic regression are larger (in abs. value) than from probit regression by approximately

$$\sqrt{\frac{\pi^2/3}{1}} = 1.8$$

because

- $V(y) = \sigma^2 = \pi^2/3$ for logistic
- $V(y) = \sigma^2 = 1$ for probit

Mixed-effects model

$\boldsymbol{\beta}$ estimates from mixed-effects model are larger (in abs. value) than from fixed-effects model by approximately

$$\sqrt{d} = \sqrt{\frac{\sigma_v^2 + \sigma^2}{\sigma^2}}$$

because

- $V(y) = \sigma_v^2 + \sigma^2$ in mixed-effects model
- $V(y) = \sigma^2$ in fixed-effects model

difference depends on size of random-effects variance σ_v^2

Within-Clusters / Between-Clusters models

Within-clusters model - *level 1* $\quad (j = 1, \ldots, n_i)$

$$\underline{\text{observed response}}$$

$$\log\left[\frac{P(\mathsf{y}_{ij}=1)}{1-P(\mathsf{y}_{ij}=1)}\right] = b_{0i} \; + \; b_{1i}\, Sex_{ij}$$

$$\underline{\text{latent response}}$$

$$y_{ij} = b_{0i} \; + \; b_{1i}\, Sex_{ij} \; + \; \varepsilon_{ij}$$

Between-clusters model - *level 2* $\quad (i = 1, \ldots, N)$

$$b_{0i} = \beta_0 \; + \; \beta_2\, Grp_i \; + \; \upsilon_{0i}$$
$$b_{1i} = \beta_1 + \; \beta_3\, Grp_i$$

$$\upsilon_{0i} \sim \mathcal{NID}(0, \sigma_\upsilon^2) \quad \text{and} \quad \varepsilon_{ij} \sim \mathcal{LID}(0, \pi^2/3)$$

Effects of a School-based Intervention

The Television School and Family Smoking Prevention and Cessation Project (Flay, *et al.*, 1988); a subsample:

- *sample* - 1600 7th-graders - 135 classes - 28 schools
  - 1 to 13 classes per school, 2 to 28 students per class
- *outcome* - knowledge of the effects of tobacco use
- *timing* - students tested at pre and post-intervention
- *design* - schools exposed to
  - a social-resistance classroom curriculum (CC)
  - a media (television) intervention (TV)
  - CC combined with TV
  - a no-treatment control group

Main question of interest:

- Influence of the intervention on the tobacco health knowledge scores (THKS) ?

Challenges in the analysis:

- outcome variable (THKS) is number correct of 7 items
- controlling for intra-school and intra-class variability
- potential explanatory variables are at different levels

Tobacco and Health Knowledge Scale
Post-Intervention Scores ≥ 3 (out of 7)
Subgroup Descriptive Statistics

	CC = no		CC = yes	
	TV=no	TV=yes	TV=no	TV=yes
n	421	416	380	383
proportions	.416	.483	.632	.603
odds	.711	.935	1.714	1.520
logits	-.341	-.067	.539	.419

Within-Subjects / Between-Subjects components

Within-clusters model - *level 1* $\quad (j = 1, \ldots, n_i$ subjects$)$

$$logit_{ij} = b_{0i}$$

Between-clusters model - *level 2* $\quad (i = 1, \ldots, N$ clusters$)$

$$b_{0i} \ = \ \beta_0 \ + \ \beta_1 CC_i \ + \ \beta_2 TV_i \ + \ \beta_3 (CC_i \times TV_i) \ + \ \upsilon_{0i}$$

$$\upsilon_{0i} \sim \mathcal{NID}(0, \sigma_\upsilon^2)$$

$$\beta_0 \; = \; \text{THKS logit for CC=no TV=no subgroup}$$

$$\beta_1 \; = \; \text{logit diff. between CC=yes vs CC=no (for TV=no)}$$

$$b_{0i} \; = \; \beta_0 \; + \; (\beta_1 + \beta_3 TV_i) CC_i \; + \; \beta_2 TV_i \; + \; v_{0i}$$

$$\beta_2 \; = \; \text{logit diff. between TV=yes vs TV=no (for CC=no)}$$

$$b_{0i} \; = \; \beta_0 \; + \; (\beta_2 + \beta_3 CC_i) TV_i \; + \; \beta_1 CC_i \; + \; v_{0i}$$

$$\beta_3 \; = \; \text{difference in logit attributable to interaction}$$

$$v_{0i} \; = \; \text{random cluster deviation}$$

note: interpretation depends on coding of variables

3-level model

Within-classrooms (and schools) model - *level 1*
($k = 1, \ldots, n_{ij}$ students)

$$logit_{ijk} = b_{0ij}$$

Between-classrooms (within-schools) model - *level 2*
($j = 1, \ldots, n_i$ classrooms)

$$b_{0ij} \;=\; b_{0i} \;+\; \upsilon_{0ij}$$

Between-schools model - *level 3* ($i = 1, \ldots, N$ schools)

$$b_{0i} \;=\; \beta_0 \;+\; \beta_1 CC_i \;+\; \beta_2 TV_i \;+\; \beta_3 (CC_i \times TV_i) \;+\; \upsilon_{0i}$$

$$\upsilon_{0ij} \sim \mathcal{NID}(0, \sigma^2_{\upsilon(2)}) \quad \text{and} \quad \upsilon_{0i} \sim \mathcal{NID}(0, \sigma^2_{\upsilon(3)})$$

$$
\begin{aligned}
\beta_0 &= \text{THKS logit for CC=no TV=no subgroup} \\
\beta_1 &= \text{logit diff. between CC=yes vs CC=no (for TV=no)} \\
\beta_2 &= \text{logit diff. between TV=yes vs TV=no (for CC=no)} \\
\beta_3 &= \text{difference in logit attributable to interaction} \\
\upsilon_{0ij} &= \text{random classroom deviation} \\
\upsilon_{0i} &= \text{random school deviation}
\end{aligned}
$$

THKS Post-Int (dichotomized) Scores - LR Estimates (std errs)

	Ordinary LR		*Mixed LR models*					
intercept	-.341	***	-.384	***	-.389	***	-.391	***
	(.099)		(.160)		(.303)		(.316)	
CC	.880	***	.887	***	1.042	**	.979	**
	(.145)		(.211)		(.411)		(.425)	
TV	.273	**	.232		.362		.323	
	(.139)		(.207)		(.334)		(.339)	
CC× TV	-.394	*	-.324		-.601		-.501	
	(.204)		(.294)		(.468)		(.475)	
class sd			.524				.411	
			(.095)				(.133)	
school sd					.402		.347	
					(.124)		(.138)	
-2 log L	2162.53		2138.15		2141.88		2133.70	

***$p < .01$ **$p < .05$ *$p < .10$ (Wald-tests not done for sds)

Calculation of ICC - 2 level models

$$r = \frac{\sigma_\upsilon^2}{\sigma_\upsilon^2 + \sigma^2}$$

Random classrooms model

$$r = \frac{.524^2}{.524^2 + \pi^2/3} = .077$$

Random schools model

$$r = \frac{.402^2}{.402^2 + \pi^2/3} = .047$$

Calculation of ICC - 3 level model

Level-3 (likeness of students in the same school)

$$r = \frac{\sigma^2_{\upsilon(3)}}{\sigma^2_{\upsilon(3)} + \sigma^2_{\upsilon(2)} + \sigma^2} = \frac{.347^2}{.347^2 + .411^2 + \pi^2/3} = .034$$

Level-2 (likeness of students in same classroom & school)

$$r = \frac{\sigma^2_{\upsilon(3)} + \sigma^2_{\upsilon(2)}}{\sigma^2_{\upsilon(3)} + \sigma^2_{\upsilon(2)} + \sigma^2} = \frac{.347^2 + .411^2}{.347^2 + .411^2 + \pi^2/3} = .081$$

Level-2 (likeness of classes in the same school)

$$r = \frac{\sigma^2_{\upsilon(3)}}{\sigma^2_{\upsilon(3)} + \sigma^2_{\upsilon(2)}} = \frac{.347^2}{.347^2 + .411^2} = .416$$

- $r < .5$: the school level contributes slightly less to variability than the class level
- average classroom post THKS scores are moderately similar within schools

CC	TV	logistic $\Psi(z) = \frac{1}{1+\exp(-z)}$	estimate
Fixed-effects model			
0	0	$\Psi(-.341)$	.416
0	1	$\Psi(-.341 + .273)$	.483
1	0	$\Psi(-.341 + .880)$	.632
1	1	$\Psi(-.341 + .273 + .880 - .394)$	.603
Random-classrooms model $\hat{d} = (.524^2 + \pi^2/3)/(\pi^2/3)$			
0	0	$\Psi((-.384)/\sqrt{\hat{d}})$	.409
0	1	$\Psi((-.384 + .232)/\sqrt{\hat{d}})$	.464
1	0	$\Psi((-.384 + .887)/\sqrt{\hat{d}})$	.619
1	1	$\Psi((-.384 + .232 + .887 - .324)/\sqrt{\hat{d}})$	.597
Random-schools model $\hat{d} = (.402^2 + \pi^2/3)/(\pi^2/3)$			
0	0	$\Psi((-.389)/\sqrt{\hat{d}})$	.406
0	1	$\Psi((-.389 + .362)/\sqrt{\hat{d}})$	.493
1	0	$\Psi((-.389 + 1.042)/\sqrt{\hat{d}})$	.654
1	1	$\Psi((-.389 + .362 + 1.042 - .601)/\sqrt{\hat{d}})$	.600
3-level model $\hat{d} = (.347^2 + .411^2 + \pi^2/3)/(\pi^2/3)$			
0	0	$\Psi((-.391)/\sqrt{\hat{d}})$	.407
0	1	$\Psi((-.391 + .323)/\sqrt{\hat{d}})$	.484
1	0	$\Psi((-.391 + .979)/\sqrt{\hat{d}})$	.637
1	1	$\Psi((-.391 + .323 + .979 - .501)/\sqrt{\hat{d}})$	.597

d = design effect $= (\sigma_v^2 + \sigma^2)/\sigma^2$ or $= (\sigma_{v(3)}^2 + \sigma_{v(2)}^2 + \sigma^2)/\sigma^2$

Within-Clusters / Between-Clusters components

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

$$logit_{ij} = b_{0i} + b_{1i} PRETHKS_{ij}$$

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

$$\begin{aligned} b_{0i} &= \beta_0 \ + \ \beta_2 CC_i \ + \ \beta_3 TV_i \ + \ \beta_4 (CC_i \times TV_i) \ + \ \upsilon_{0i} \\ b_{1i} &= \beta_1 \end{aligned}$$

$$\upsilon_{0i} \sim \mathcal{NID}(0, \sigma_\upsilon^2)$$

β_0 = (PRETHKS adjusted) logit for CC=no TV=no subgroup

β_1 = effect of PRETHKS on POSTTHKS

β_2 = (PRETHKS adjusted) logit diff. between CC=yes vs CC=no (for TV=no)

β_3 = (PRETHKS adjusted) logit diff. between TV=yes vs TV=no (for CC=no)

β_4 = (PRETHKS adjusted) difference in logit attributable to interaction

υ_{0i} = random cluster deviation

3-level model

Within-classrooms (and schools) model - *level 1*
($k = 1, \ldots, n_{ij}$ students)

$$logit_{ijk} = b_{0ij} + b_{1ij} PRETHKS_{ijk}$$

Between-classrooms (within-schools) model - *level 2*
($j = 1, \ldots, n_i$ classrooms)

$$\begin{aligned} b_{0ij} &= b_{0i} + \upsilon_{0ij} \\ b_{1ij} &= b_{1i} \end{aligned}$$

Between-schools model - *level 3* ($i = 1, \ldots, N$ schools)

$$\begin{aligned} b_{0i} &= \beta_0 + \beta_2 CC_i + \beta_3 TV_i + \beta_4 (CC_i \times TV_i) + \upsilon_{0i} \\ b_{1i} &= \beta_1 \end{aligned}$$

$$\upsilon_{0ij} \sim \mathcal{NID}(0, \sigma^2_{\upsilon(2)}) \quad \text{and} \quad \upsilon_{0i} \sim \mathcal{NID}(0, \sigma^2_{\upsilon(3)})$$

THKS Post-Int (dichotomized) Scores - LR Estimates (std err)

	Ordinary LR		*Mixed LR models*					
intercept	-1.217	***	-1.253	***	-1.228	***	-1.246	***
	(.140)		(.190)		(.267)		(.272)	
PRETHKS	.400	***	.401	***	.387	***	.395	***
	(.043)		(.046)		(.043)		(.046)	
CC	.973	***	.988	***	1.089	***	1.038	***
	(.151)		(.207)		(.354)		(.358)	
TV	.316	**	.287		.374		.333	
	(.144)		(.199)		(.281)		(.285)	
CC× TV	-.413	**	-.369		-.558		-.464	
	(.219)		(.284)		(.408)		(.409)	
class sd			.468				.406	
			(.094)				(.132)	
school sd					.326		.251	
					(.115)		(.143)	
-2 log L	2073.32		2057.18		2063.20		2055.70	

$^{***}p < .01$ $^{**}p < .05$ $^{*}p < .10$ (Wald-tests not done for sds)

Calculation of ICC - 2 level models

$$r = \frac{\sigma_\upsilon^2}{\sigma_\upsilon^2 + \sigma^2}$$

Random classrooms model

$$r = \frac{.468^2}{.468^2 + \pi^2/3} = .062$$

Random schools model

$$r = \frac{.326^2}{.326^2 + \pi^2/3} = .031$$

Calculation of ICC - 3 level model

Level-3 (likeness of students in the same school)

$$r = \frac{\sigma^2_{\upsilon(3)}}{\sigma^2_{\upsilon(3)} + \sigma^2_{\upsilon(2)} + \sigma^2} = \frac{.251^2}{.251^2 + .406^2 + \pi^2/3} = .018$$

Level-2 (likeness of students in same classroom & school)

$$r = \frac{\sigma^2_{\upsilon(3)} + \sigma^2_{\upsilon(2)}}{\sigma^2_{\upsilon(3)} + \sigma^2_{\upsilon(2)} + \sigma^2} = \frac{.251^2 + .406^2}{.251^2 + .406^2 + \pi^2/3} = .065$$

Level-2 (likeness of classes in the same school)

$$r = \frac{\sigma^2_{\upsilon(3)}}{\sigma^2_{\upsilon(3)} + \sigma^2_{\upsilon(2)}} = \frac{.251^2}{.251^2 + .406^2} = .276$$

- $r < .5$: the school level contributes slightly less to variability than the class level
- average classroom post THKS scores are fairly similar within schools

CC	TV	logistic $\Psi(z) = \frac{1}{1+\exp(-z)}$	estimate
Fixed-effects model			
0	0	$\Psi(-1.217 + 2.152 \times .400)$	.412
0	1	$\Psi(-1.217 + .316 + 2.087 \times .400)$	.483
1	0	$\Psi(-1.217 + .973 + 2.050 \times .400)$	.640
1	1	$\Psi(-1.217 + .316 + .973 - .413 + 1.979 \times .400)$	.610
Random-classrooms model $\hat{d} = (.468^2 + \pi^2/3)/(\pi^2/3)$			
0	0	$\Psi((-1.253 + 2.152 \times .401)/\sqrt{\hat{d}})$	.407
0	1	$\Psi((-1.253 + .287 + 2.087 \times .401)/\sqrt{\hat{d}})$	.469
1	0	$\Psi((-1.253 + .988 + 2.050 \times .401)/\sqrt{\hat{d}})$	.632
1	1	$\Psi((-1.253 + .287 + .988 - .369 + 1.979 \times .401)/\sqrt{\hat{d}})$	.606
Random-schools model $\hat{d} = (.326^2 + \pi^2/3)/(\pi^2/3)$			
0	0	$\Psi((-1.228 + 2.152 \times .387)/\sqrt{\hat{d}})$	.404
0	1	$\Psi((-1.228 + .374 + 2.087 \times .387)/\sqrt{\hat{d}})$	.489
1	0	$\Psi((-1.228 + 1.089 + 2.050 \times .387)/\sqrt{\hat{d}})$	.656
1	1	$\Psi((-1.228 + .374 + 1.089 - .558 + 1.979 \times .387)/\sqrt{\hat{d}})$	.607
3-level model $\hat{d} = (.251^2 + .406^2 + \pi^2/3)/(\pi^2/3)$			
0	0	$\Psi((-1.246 + 2.152 \times .395)/\sqrt{\hat{d}})$	.405
0	1	$\Psi((-1.246 + .333 + 2.087 \times .395)/\sqrt{\hat{d}})$	.479
1	0	$\Psi((-1.246 + 1.038 + 2.050 \times .395)/\sqrt{\hat{d}})$	.642
1	1	$\Psi((-1.246 + .333 + 1.038 - .464 + 1.979 \times .395)/\sqrt{\hat{d}})$	.605

d = design effect = $(\sigma_v^2 + \sigma^2)/\sigma^2$ or $= (\sigma_{v(3)}^2 + \sigma_{v(2)}^2 + \sigma^2)/\sigma^2$

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

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?

Challenges in the analysis:

- controlling for intra-subject variability
- varying numbers of observations per subject
- potential problem due to dropouts
- may want to treat outcome as dichotomous or ordinal

Observed proportions $\geq$ "moderately ill"

	week 0	week 1	week 3	week 6
placebo	.98	.91	.89	.71
drug	.99	.82	.66	.42

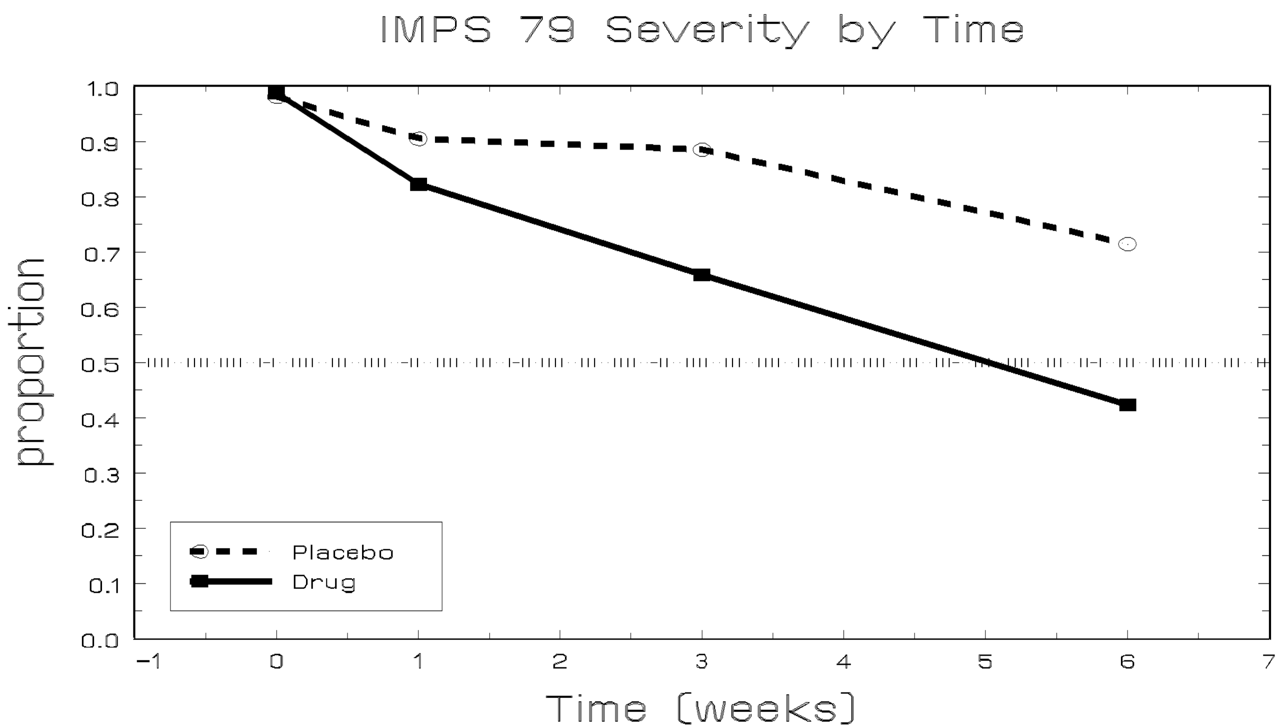
Observed odds $\geq$ "moderately ill"

	week 0	week 1	week 3	week 6
placebo	52.5	9.50	7.70	2.50
drug	80.8	4.63	1.93	.73
ratio	.65	2.05	3.99	3.42

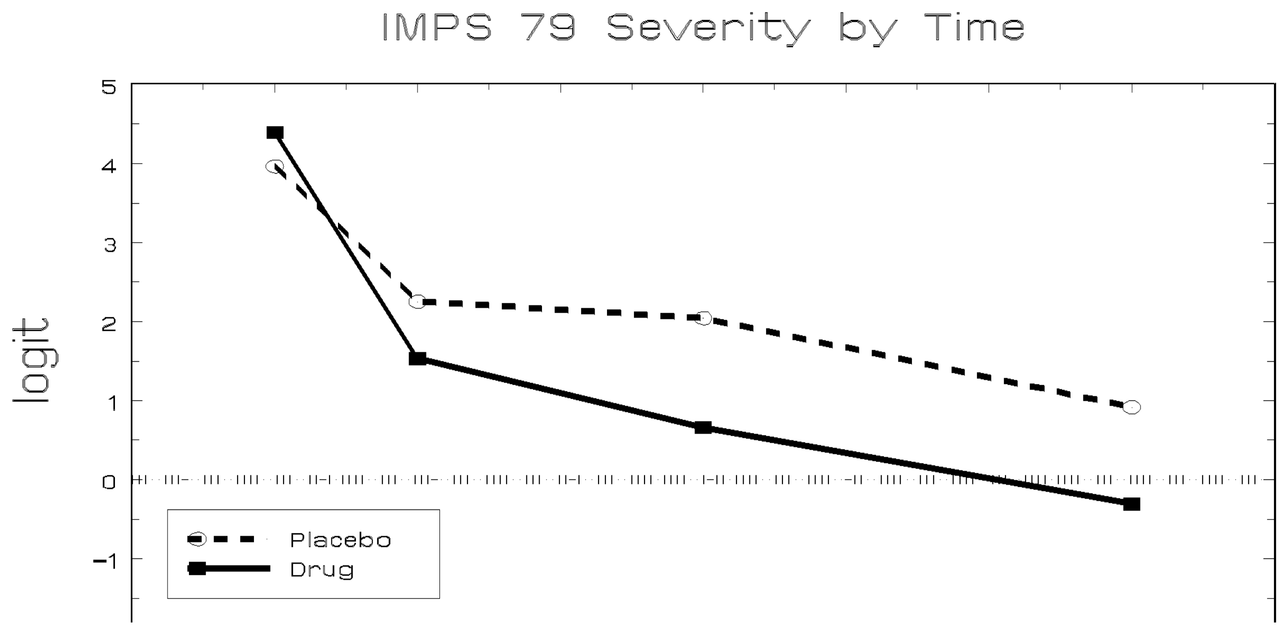
Observed log odds $\geq$ "moderately ill"

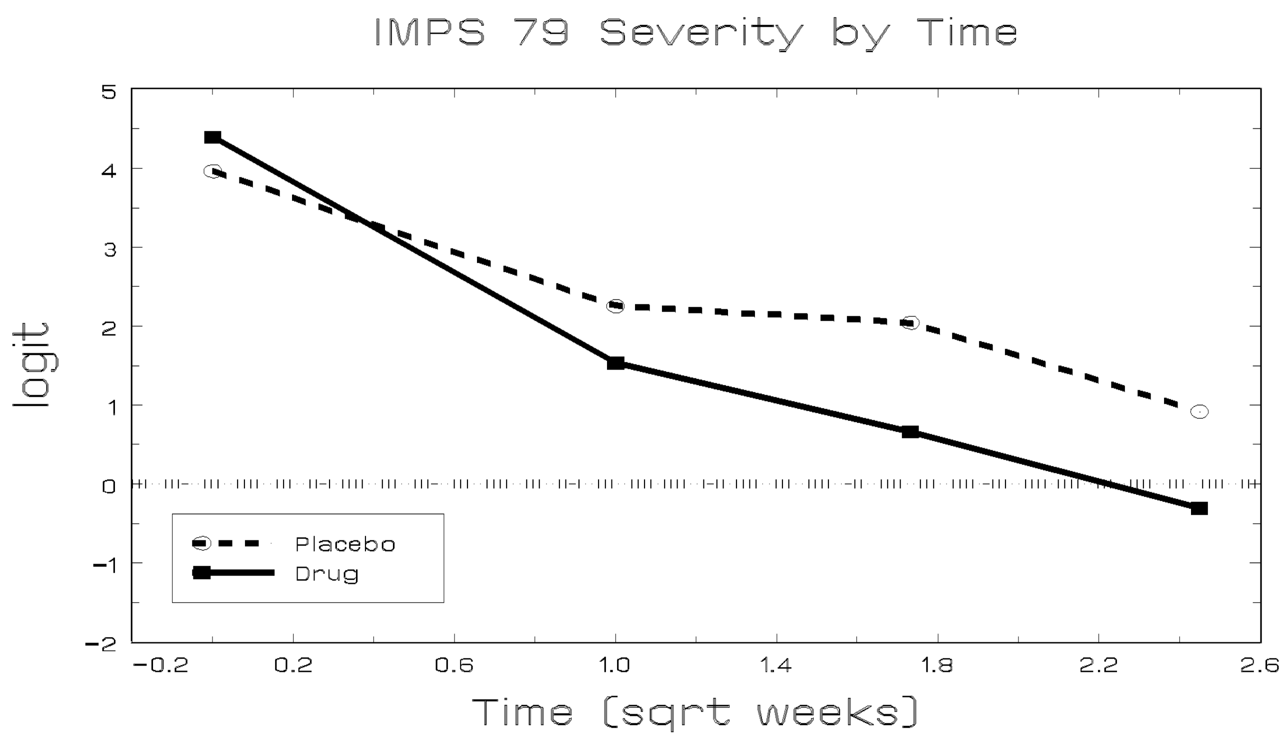
	week 0	week 1	week 3	week 6
placebo	3.96	2.25	2.04	.92
drug	4.39	1.53	.66	-.31
difference	-.43	.72	1.38	1.23
exp (odds ratio)	.65	2.05	3.99	3.42

Observed Proportions across Time by Condition



- model is not linear in terms of probabilites


IMPS 79 Severity by Time
logit
5
4
3
2
1
0
−1
Placebo
Drug


IMPS 79 Severity by Time
logit
5
4
3
2
1
0
−1
−2
−0.2
0.2
0.6
1.0
1.4
1.8
2.2
2.6
Time (sqrt weeks)
Placebo
Drug

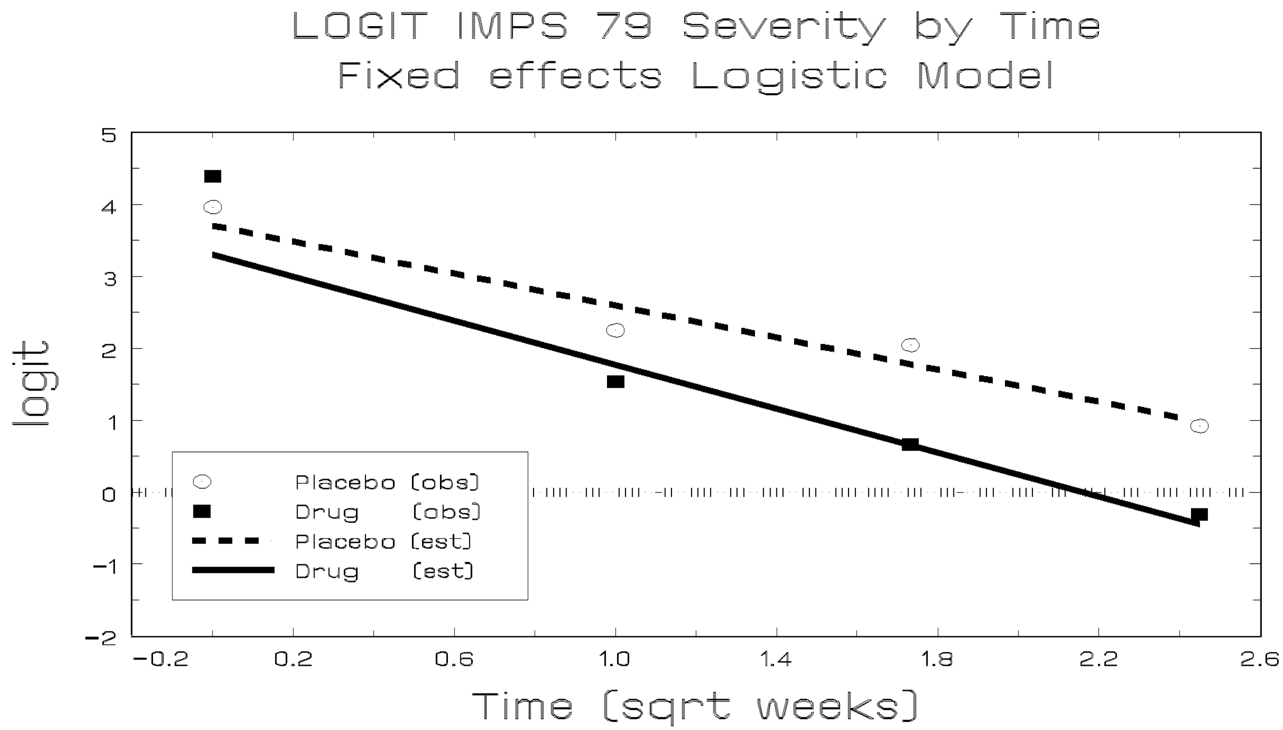
NIMH Schizophrenia Study - Severity of Illness (N = 437)
(Dichotomous) LR Estimates - *Fixed effects (marginal) model*

	ML estimate	se	z	$p <$
intercept	3.703	0.442	8.38	.001
Drug (0 = plc; 1 = drug)	-0.405	0.494	-0.82	.41
Time (sqrt week)	-1.113	0.234	-4.76	.001
Drug by Time	-0.418	0.262	-1.60	.11
$-2 \log L = 1362.06$				

Appropriate model if data were cross-sectional longitudinal or if $\sigma_\upsilon = 0$

Fitted Logits across Time by Condition

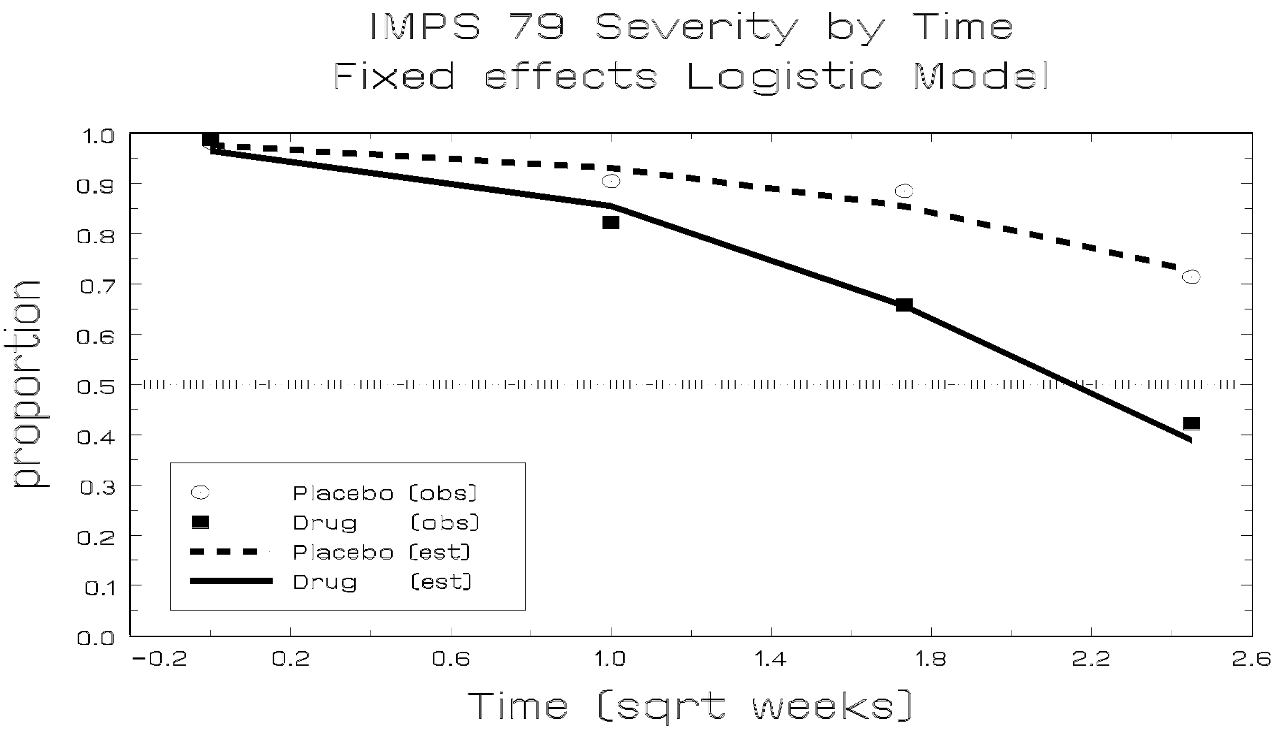
fixed-effects logistic regression model



$$\log\left[\frac{P(\mathsf{y}_{ij}=1)}{1-P(\mathsf{y}_{ij}=1)}\right] \;=\; 3.70 - .41\,Drug_i - 1.11\,Time_j - .42\,(D_i \times T_j)$$

Fitted Proportions across Time by Condition

fixed-effects logistic regression model



$$P(\mathsf{y}_{ij} = 1) \; = \; 1/\left\{1 + \exp\left[-\left(3.70 - .41\,D_i - 1.11\,T_j - .42\,D_i T_j\right)\right]\right\}$$

Within-Subjects / Between-Subjects components

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

$$logit_{ij} = b_{0i} \; + \; b_{1i}\sqrt{Week}_j$$

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

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

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

$$\upsilon_{0i} \sim \mathcal{NID}(0, \sigma_\upsilon^2)$$

$$
\begin{aligned}
\beta_0 &= \text{week 0 IMPS79 logit for PLC patients } (Grp = 0) \\
\beta_1 &= \text{IMPS79 (sqrt) weekly logit change for PLC patients } (Grp = 0) \\
\beta_2 &= \text{difference in week 0 IMPS79 logit for DRUG patients } (Grp = 1) \\
\beta_3 &= \text{difference in IMPS79 (sqrt) weekly logit change for DRUG patients } (Grp = 1) \\
\upsilon_{0i} &= \text{individual deviation from group trend}
\end{aligned}
$$

NIMH Schiz Study - Severity of Illness (N = 437)
(Dichotomous) Mixed LR Estimates (se)

	ML estimates	se	z	$p <$
intercept	5.387	0.535	10.07	.001
Drug (0 = plc; 1 = drug)	-0.025	0.601	-0.04	.97
Time (sqrt week)	-1.500	0.228	-6.59	.001
Drug by Time	-1.015	0.274	-3.70	.001
Intercept sd	2.116	0.215		

$$\textit{Intra-person correlation} = 2.116^2/(2.116^2 + \pi^2/3) = .58$$

$$-2 \log L = 1249.74 \qquad \chi^2_1 = 112.3$$

Observed log odds $\geq$ "moderately ill"

	week 0	week 1	week 3	week 6
placebo	3.96	2.25	2.04	.92
drug	4.39	1.53	.66	-.31
difference	-.43	.72	1.38	1.23
exp (odds ratio)	.65	2.05	3.99	3.42

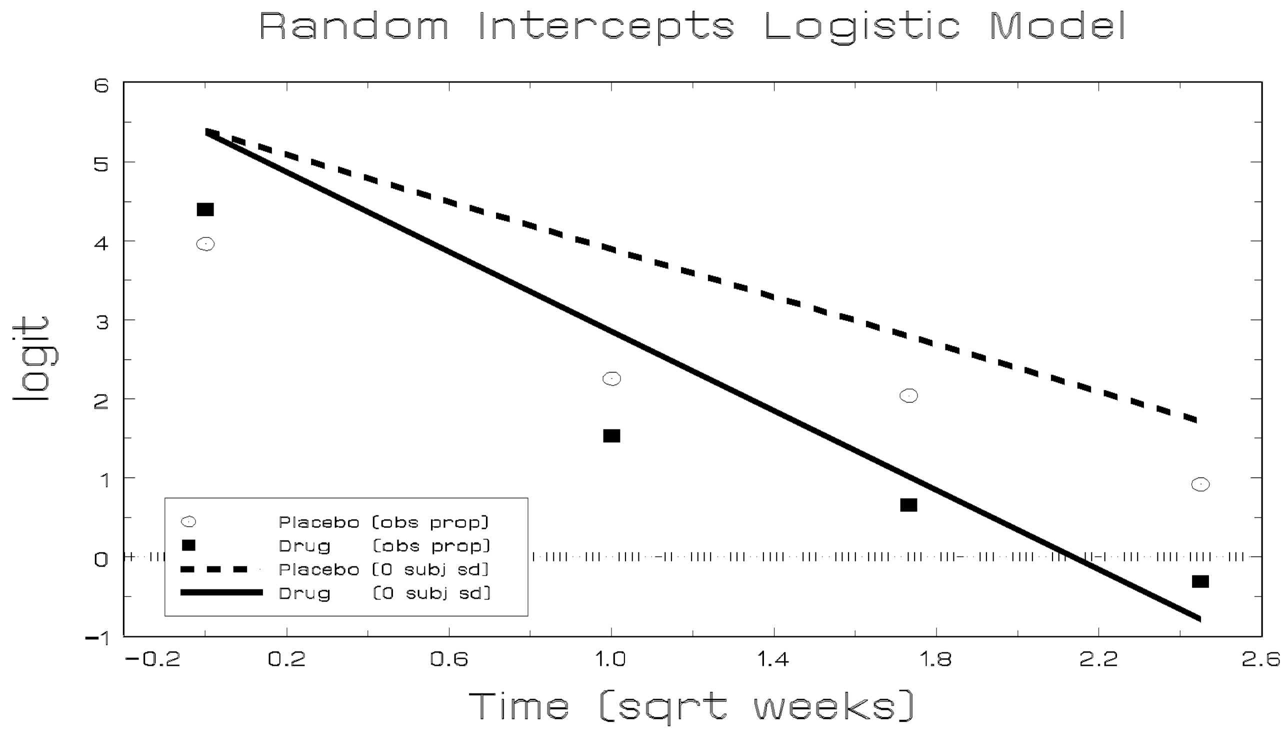
Estimated log odds $\geq$ "moderately ill"

	week 0	week 1	week 3	week 6
placebo	5.39	3.89	2.79	1.71
drug	5.36	2.85	1.01	-.80
difference	.03	1.04	1.78	2.51
exp (odds ratio)	1.03	2.83	5.95	12.3

Estimated logits are much larger (in absolute value)!

Fitted Logits across Time by Condition

random-intercepts logistic regression model

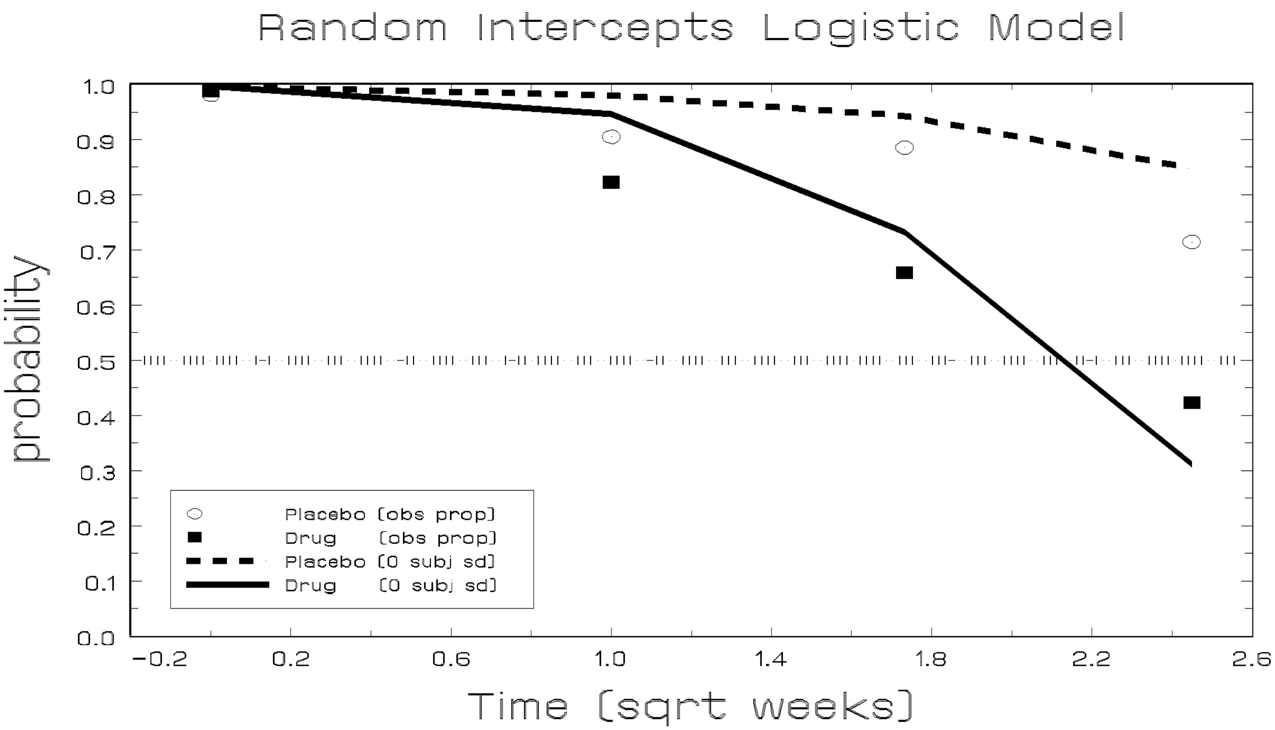


$$\log\left[\frac{P(\mathsf{y}_{ij}=1)}{1-P(\mathsf{y}_{ij}=1)}\right] = 5.39 - .03\, D_i - 1.50\, T_j - 1.02\, (D_i \times T_j) + \upsilon_{0i}$$

where $\upsilon_{0i} \sim \mathcal{NID}(0, \hat{\sigma}_\upsilon = 2.12)$

Fitted Proportions across Time by Condition

random-intercepts logistic regression model



$$P(\mathsf{y}_{ij} = 1) \; = \; 1/\left\{1 + \exp\left[-\left(5.39 - .03\, D_i - 1.50\, T_j - 1.02\, D_i T_j + \upsilon_{0i}\right)\right]\right\}$$

where $\upsilon_{0i} \sim \mathcal{NID}(0, \hat{\sigma}_\upsilon = 2.12)$

Random-intercepts Logistic Regression

$$logit_{ij} = \boldsymbol{x}'_{ij}\boldsymbol{\beta} + \upsilon_{0i}$$

- every subject has their own propensity for response (υ_{0i})
- the influence of covariates $\boldsymbol{x}$ is determined controlling (or adjusting) for the subject effect
- the covariance structure, or dependency, of the repeated observations is explicitly modeled

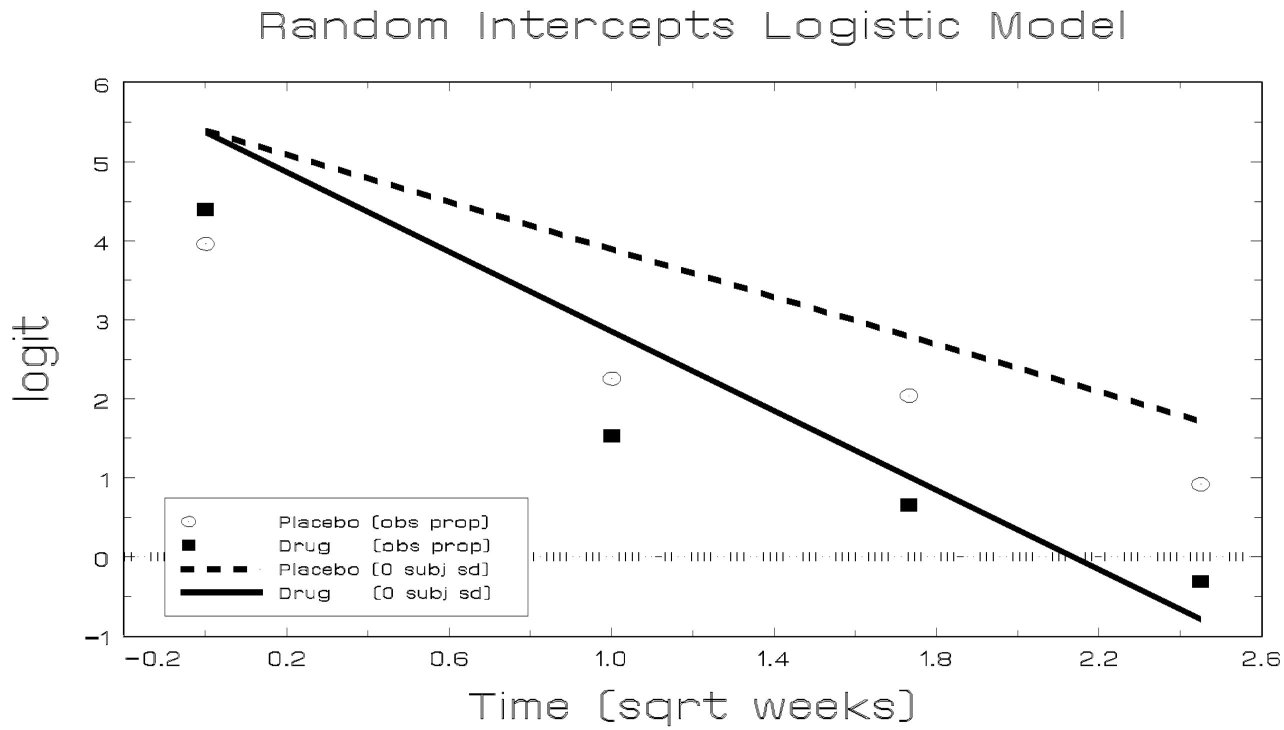
β_0 = log odds of response for a typical subject with $\boldsymbol{x} = 0$ and $\upsilon_{0i} = 0$

$\boldsymbol{\beta}$ = log odds ratio for response associated with unit changes in $\boldsymbol{x}$ for the same subject value υ_{0i}
$\Rightarrow$ on average, how a *subject's* response probability depends on $\boldsymbol{x}$

σ_υ = degree of heterogeneity across subjects in the probability of response not attributable to $\boldsymbol{x}$

- most useful when the objective is to make inference about *subjects* rather than the population average
- interest is in the heterogeneity of subjects

Estimated (subject-specific) Logits across Time by Condition:
random-intercepts model



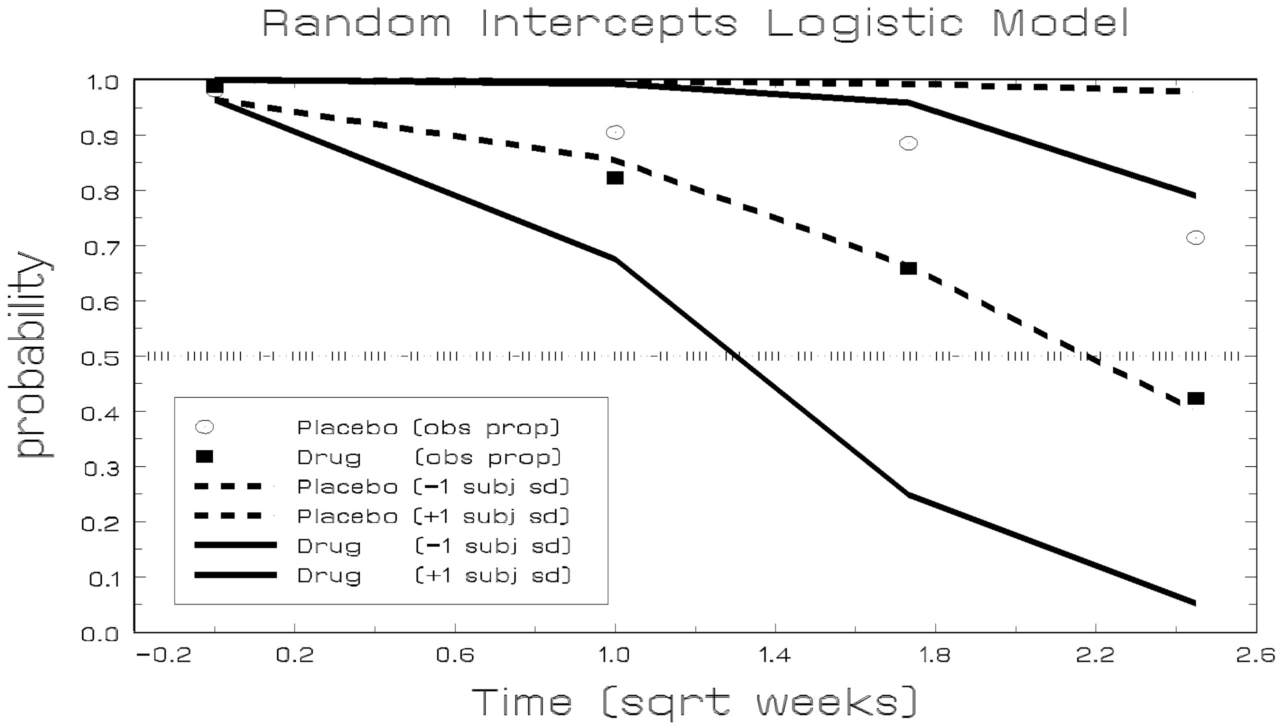
$$\log\left[\frac{P(\mathsf{y}_{ij}=1)}{1-P(\mathsf{y}_{ij}=1)}\right] \;=\; 5.39 - .03\,Drug_i - 1.50\,Time_j - 1.02\,(D_i \times T_j) + \upsilon_{0i}$$

$$\upsilon_{0i} \sim \mathcal{NID}(0, \hat{\sigma}_\upsilon = 2.12)$$

$\hat{\boldsymbol{\beta}}$ assesses change in (conditional) logit due to $\boldsymbol{x}$ for subjects with the same value of υ_{0i} (can be thought of as an omitted subject covariate)

Estimated Subject-Specific Probabilites

random-intercepts logistic regression model



$$P(\mathsf{y}_{ij} = 1) \;=\; 1/\left\{1 + \exp\left[-\left(5.39 - .03\, D_i - 1.50\, T_j - 1.02\, D_i T_j + \upsilon_{0i}\right)\right]\right\}$$

$$\text{where } \upsilon_{0i} = \begin{cases} -1\sigma_\upsilon \\ \;\;1\sigma_\upsilon \end{cases} \text{ and } \hat{\sigma}_\upsilon = 2.12$$

Model fit of observed marginal proportions

1. $\hat{\boldsymbol{y}}_i = \boldsymbol{X}_i \hat{\boldsymbol{\beta}}$

2. calculate "marginalization" vector

$$\hat{\boldsymbol{s}} = \frac{1}{\sigma}\left[\text{Diag}(\hat{V}(\boldsymbol{y}_i))\right]^{1/2}$$

- $\hat{V}(\boldsymbol{y}_i) = \boldsymbol{Z}_i \hat{\boldsymbol{\Sigma}}_\upsilon \boldsymbol{Z}_i' + \sigma^2 \boldsymbol{I}_i$
- $\sigma = 1$ for probit and $\sigma = \pi/\sqrt{3}$ for logistic
- $\boldsymbol{Z}_i$ = design matrix for random effects
- for random-intercepts model $\boldsymbol{Z}_i = \mathbf{1}_i$, and so, $\hat{s} = \sqrt{\hat{\sigma}_\upsilon^2/\sigma^2 + 1}$

3. perform element-wise division

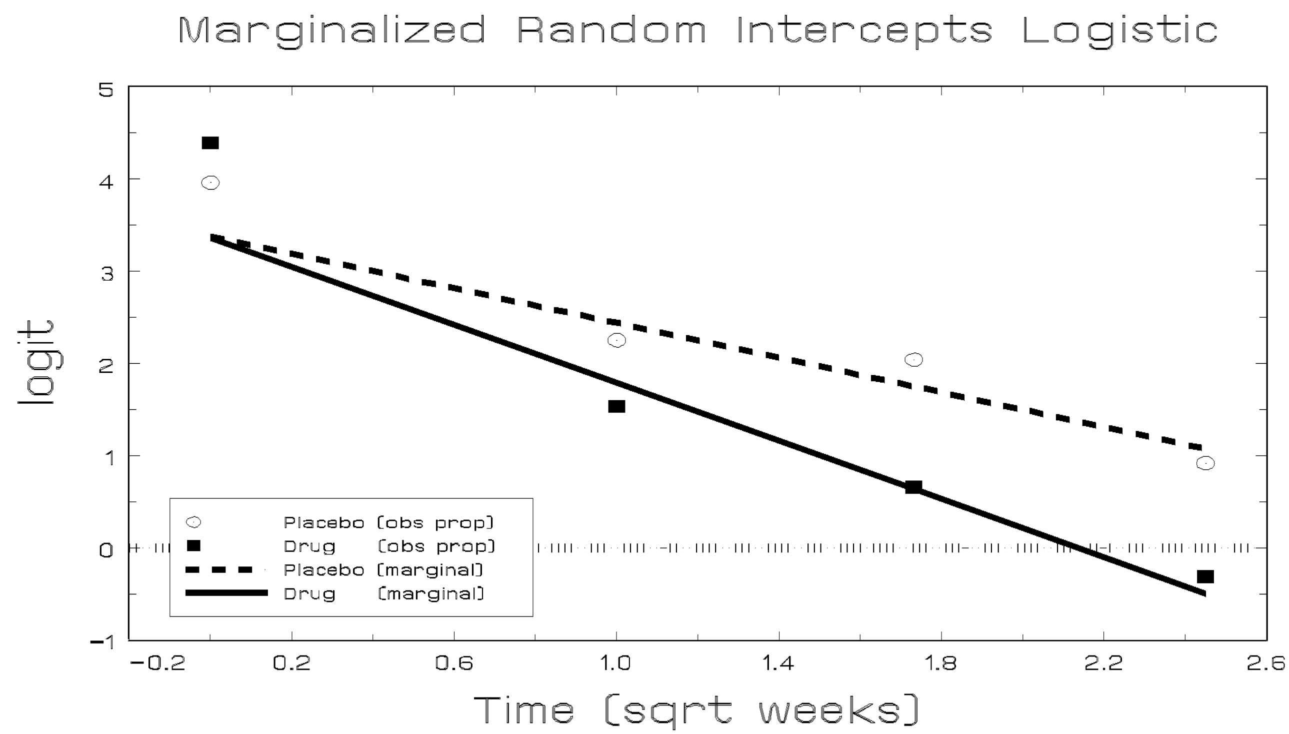
$$\hat{\mathbf{z}}_i = \hat{\boldsymbol{y}}_i \; /. \; \hat{\boldsymbol{s}}$$

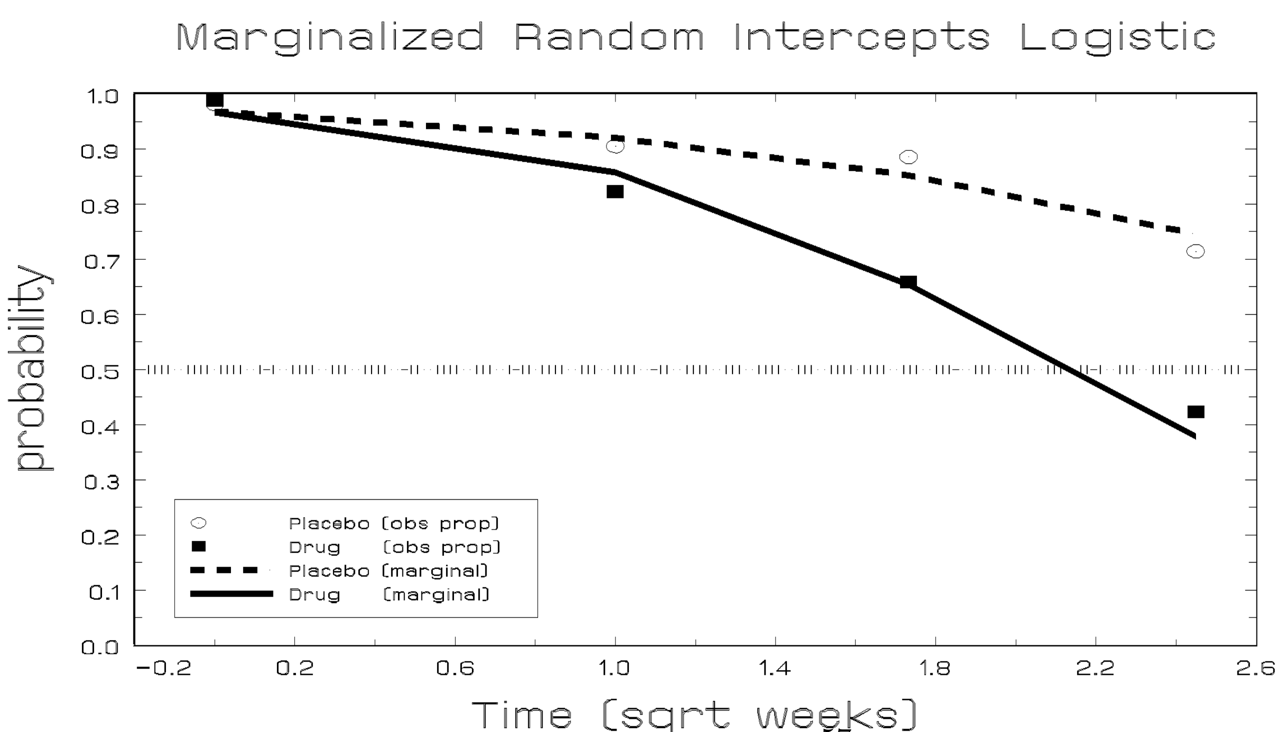
4. $\hat{\boldsymbol{p}}_i = \Phi(\hat{\mathbf{z}}_i)$ for probit and $\hat{\boldsymbol{p}}_i = \Psi(\hat{\mathbf{z}}_i)$ for logistic

5. In practice, for logistic, $(15\pi)/(16\sqrt{3})$ works better than $\pi/\sqrt{3}$ as σ (Zeger *et al.*, 1988, Biometrics)
6. Logistic is approximate; relies on cumulative Gaussian approximation to the logistic function

```
TITLE1 'NIMH Schizophrenia Data - Estimated Marginal Probabilities';
PROC IML;
/* Results from MIXOR analysis:  random intercept model */;
 x0  =  {  0 0.00000 0,
           0 1.00000 0,
           0 1.73205 0,
           0 2.44949 0};
 x1  =  {  1 0.00000 0.00000,
           1 1.00000 1.00000,
           1 1.73205 1.73205,
           1 2.44949 2.44949};
 int   =  {5.387};
 sd    =  {2.116};
 beta  =  {-.025, -1.500, -1.015};

/* Approximate Marginalization Method */;
 pi    =  3.141592654;
 nt    =  4;
 ivec  =  J(nt,1,1);
 zvec  =  J(nt,1,1);
 evec  =  (15/16)**2 * (pi**2)/3 * ivec;

/* nt by nt matrix with evec on the diagonal and zeros elsewhere */;
emat = diag(evec);
/* variance-covariance matrix of underlying latent variable */;
vary = zvec * sd * T(sd) * T(zvec) + emat;
sdy = sqrt(vecdiag(vary) / vecdiag(emat));
```

```
z0 = (int + x0*beta) / sdy ;
z1 = (int + x1*beta) / sdy;

grp0 = 1 / ( 1 + EXP(0 - za));
grp1 = 1 / ( 1 + EXP(0 - za));

print 'Random intercept model';
print 'Approximate Marginalization Method';
print 'marginal prob for group 0 - response' grp0 [FORMAT=8.4];
print 'marginal prob for group 1 - response' grp1 [FORMAT=8.4];
```


Marginalized Random Intercepts Logistic
logit
Time (sqrt weeks)
Placebo (obs prop)
Drug (obs prop)
Placebo (marginal)
Drug (marginal)


Marginalized Random Intercepts Logistic
probability
Time (sqrt weeks)
Placebo (obs prop)
Drug (obs prop)
Placebo (marginal)
Drug (marginal)

Within-Subjects / Between-Subjects components

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

$$logit_{ij} = b_{0i} \; + \; b_{1i}\sqrt{Week}_j$$

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

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

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

$$\boldsymbol{\upsilon}_i \sim \mathcal{NID}(\mathbf{0}, \boldsymbol{\Sigma}_{\upsilon} = \boldsymbol{TT'})$$

$$
\begin{aligned}
\beta_0 &= \text{week 0 IMPS79 logit for PLC patients } (Grp = 0) \\
\beta_1 &= \text{IMPS79 (sqrt) weekly logit change for PLC patients } (Grp = 0) \\
\beta_2 &= \text{difference in week 0 IMPS79 logit for DRUG patients } (Grp = 1) \\
\beta_3 &= \text{difference in IMPS79 (sqrt) weekly logit change for DRUG patients } (Grp = 1) \\
\upsilon_{0i} &= \text{individual deviation from group intercept} \\
\upsilon_{1i} &= \text{individual deviation from group (sqrt) weekly change}
\end{aligned}
$$

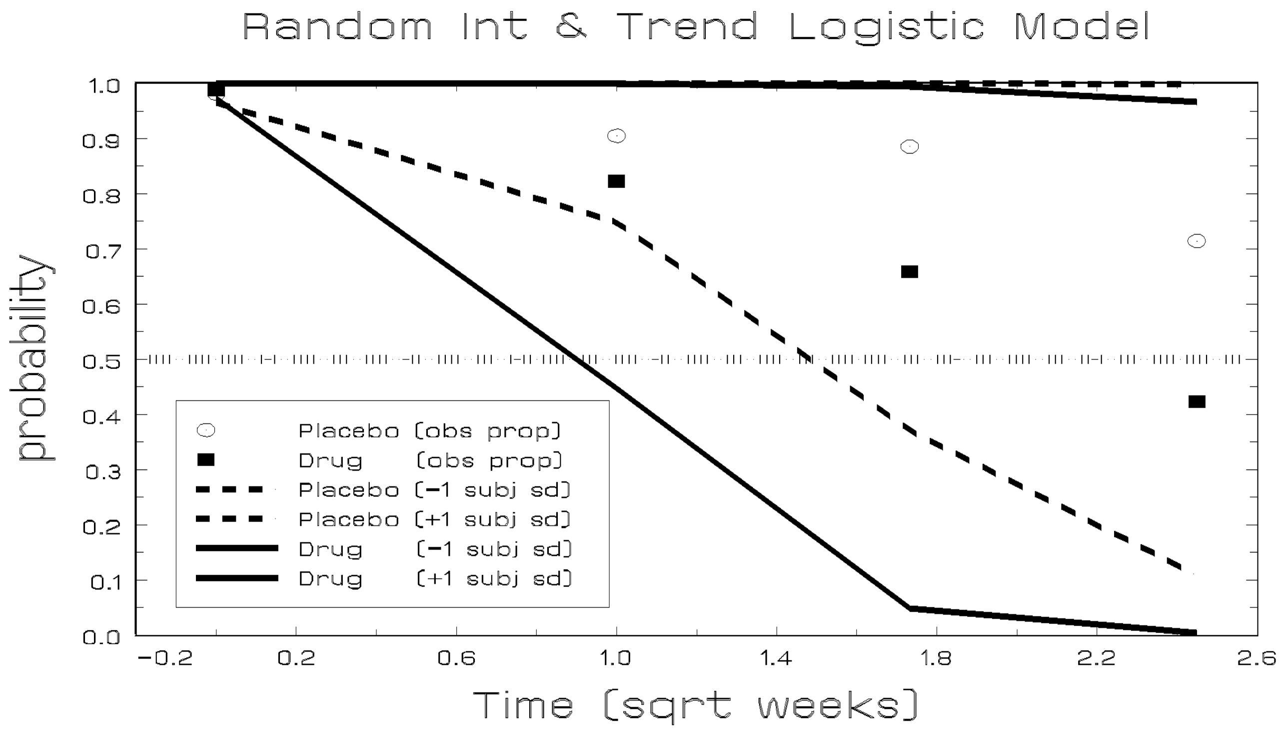
NIMH Schiz Study - Severity of Illness ($N = 437$)
LR Estimates (se) - random intercept and trend model

	ML estimates	se	z	$p <$
intercept	6.025	0.918	6.56	.001
Drug (0 = plc; 1 = drug)	0.281	0.761	0.37	.71
Time (sqrt week)	-1.477	0.451	-3.27	.001
Drug by Time	-1.587	0.479	-3.31	.001
<u>*Cholesky*</u>				
Intercept sd	2.726	0.597	$(\hat{\sigma}^2_{\upsilon_0} = 7.43)$	
Int-Time covar	-0.829	0.352	$(r_{\upsilon_0 \upsilon_1} = -.47)$	
Time sd	1.561	0.248	$(\hat{\sigma}^2_{\upsilon_1} = 3.12)$	

$-2 \log L = 1227.43, \quad \chi^2_2 = 22.31, p < .001$

Estimated (subject-specific) Logits across Time

Random intercepts and trends model

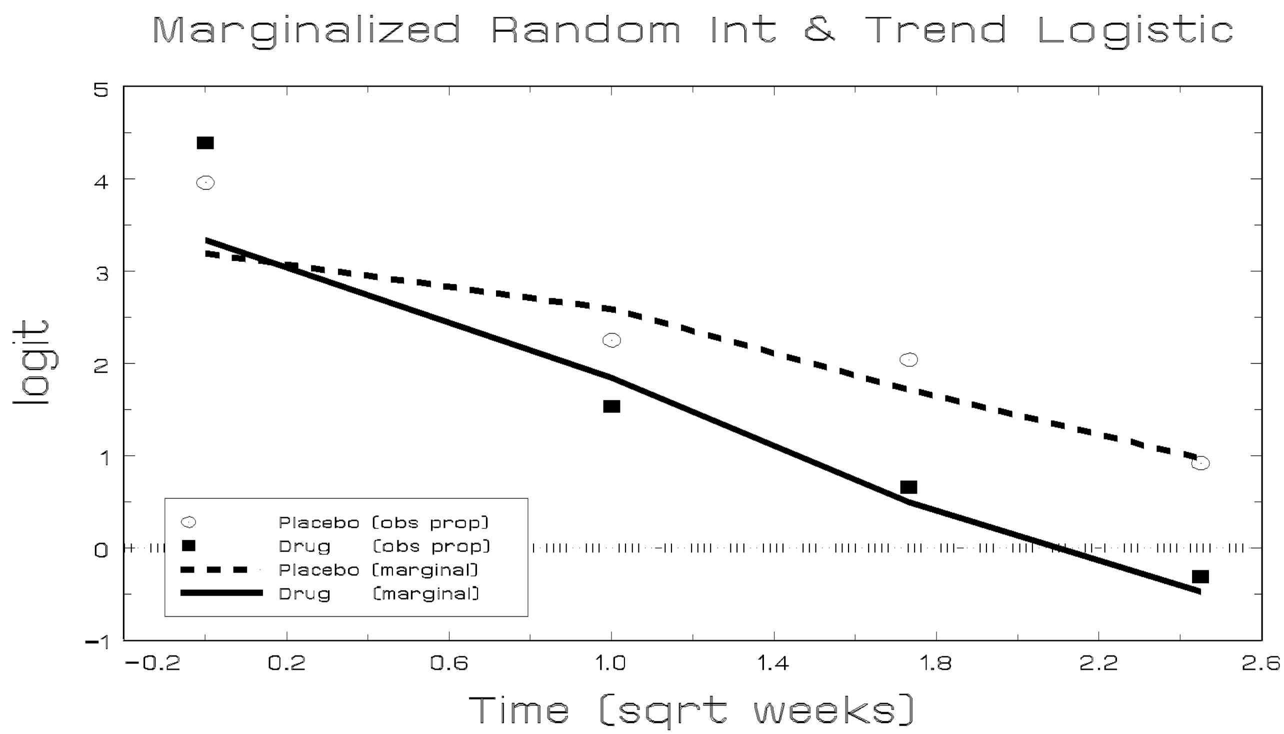


$$P(\mathsf{y}_{ij} = 1) \;=\; 1/\{1 + \exp[-(6.03 + .28\, D_i - 1.48\, T_j - 1.59\, D_i T_j + \upsilon_{0i} + \upsilon_{1i}\, T_j)]\}$$

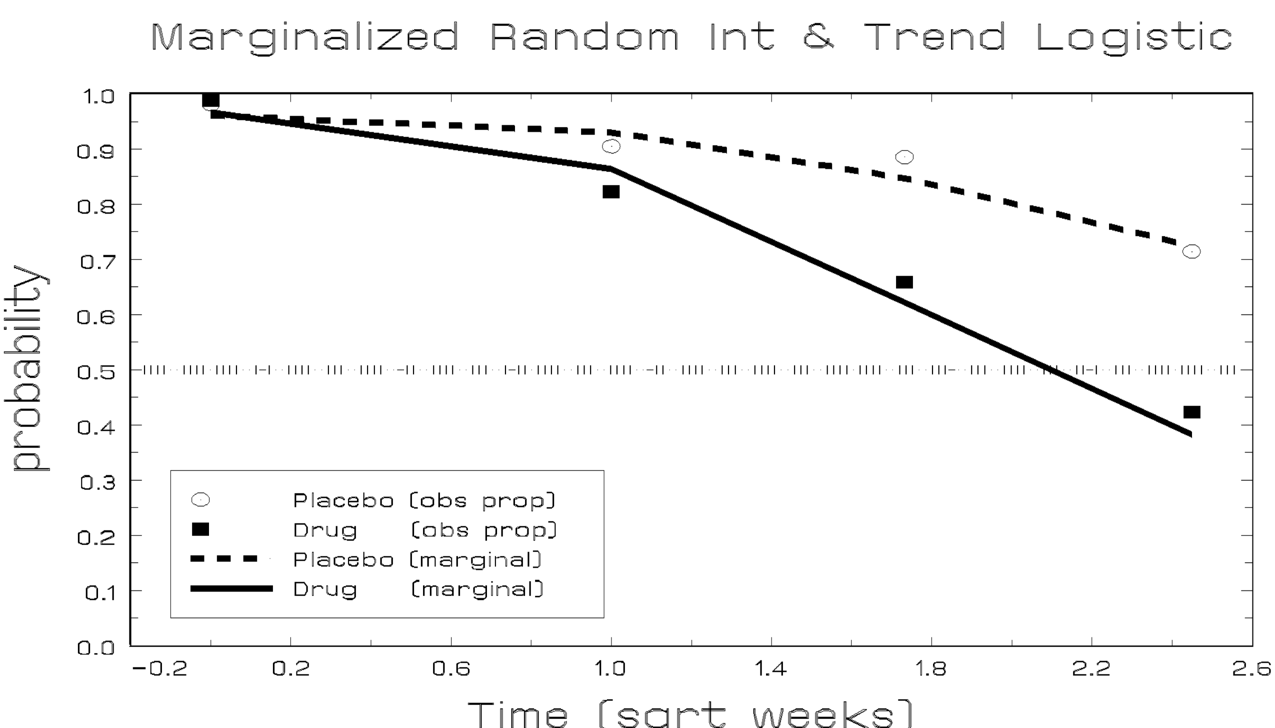
$$\begin{bmatrix} \upsilon_{0i} \\ \upsilon_{1i} \end{bmatrix} \sim \mathcal{NID} \left\{ \begin{bmatrix} 0 \\ 0 \end{bmatrix}, \hat{\boldsymbol{\Sigma}}_\upsilon = \hat{\boldsymbol{T}} \hat{\boldsymbol{T}}' \right\} \quad \hat{\boldsymbol{T}} = \begin{bmatrix} 2.73 & 0 \\ -.83 & 1.56 \end{bmatrix}$$

```
TITLE1 'NIMH Schizophrenia Data - Estimated Marginal Probabilities';
PROC IML;
/* Results from MIXOR analysis:  random intercept & trend model */;
 x0  =  {  0 0.00000 0,
           0 1.00000 0,
           0 1.73205 0,
           0 2.44949 0};
 x1  =  {  1 0.00000 0.00000,
           1 1.00000 1.00000,
           1 1.73205 1.73205,
           1 2.44949 2.44949};
 int   =  {6.025};
 chol  =  {2.726 0,
           -.829 1.561};
 beta  =  { .281, -1.477, -1.587};
/* Approximate Marginalization Method */;
 pi    =  3.141592654;
 nt    =  4;
 ivec  =  J(nt,1,1);
 zmat  =  {1 0.00000,
           1 1.00000,
           1 1.73205,
           1 2.44949};
 evec  =  (15/16)**2 * (pi**2)/3 * ivec;
/* nt by nt matrix with evec on the diagonal and zeros elsewhere */;
emat = diag(evec);
/* variance-covariance matrix of underlying latent variable */;
vary = zmat * chol * T(chol) * T(zmat) + emat;
sdy = sqrt(vecdiag(vary) / vecdiag(emat));
```

```
z0 = (int + x0*beta) / sdy ;
z1 = (int + x1*beta) / sdy;

grp0 = 1 / ( 1 + EXP(0 - za));
grp1 = 1 / ( 1 + EXP(0 - za));

print 'Random intercept and trend model';
print 'Approximate Marginalization Method';
print 'marginal prob for group 0 - response' grp0 [FORMAT=8.4];
print 'marginal prob for group 1 - response' grp1 [FORMAT=8.4];
```



Marginalized Random Int & Trend Logistic
logit
Time (sqrt weeks)
Placebo (obs prop)
Drug (obs prop)
Placebo (marginal)
Drug (marginal)



Marginalized Random Int & Trend Logistic
probability
Time (sqrt weeks)
Placebo (obs prop)
Drug (obs prop)
Placebo (marginal)
Drug (marginal)