Multilevel Models for Longitudinal Dichotomous Outcomes

Don Hedeker
University of Illinois at Chicago

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

Hedeker, D. (2005). Generalized linear mixed models. In B. Everitt & D. Howell (Eds.), Encyclopedia of Statistics in Behavioral Science. Wiley.

Chapter 9 in Hedeker, D. & Gibbons, R.D. (2006). Longitudinal Data Analysis. Wiley.

SMOKING
NON-
SMOKING

©1999 Humerus Cartoons Syndicate, humerusjkb@aol.com

Logistic Regression - model that relates explanatory variables (*i.e.*, covariates) to a dichotomous dependent variable

Multilevel Logistic Regression - model that relates covariates to a dichotomous dependent variable, where observations are nested

- Clustered: subjects within clusters
- Longitudinal: repeated observations within subjects

models can also be recast as probit regression models

Random-intercept Logistic Regression Model

Consider the model with p covariates for the dichotomous response Y_{ij} of subject i $(i = 1, \ldots, N)$ at timepoint j $(j = 1, \ldots, n_i)$:

$$\log\left[\frac{Pr(Y_{ij} = 1)}{1 - Pr(Y_{ij} = 1)}\right] = \boldsymbol{x}'_{ij}\boldsymbol{\beta} + \upsilon_i$$

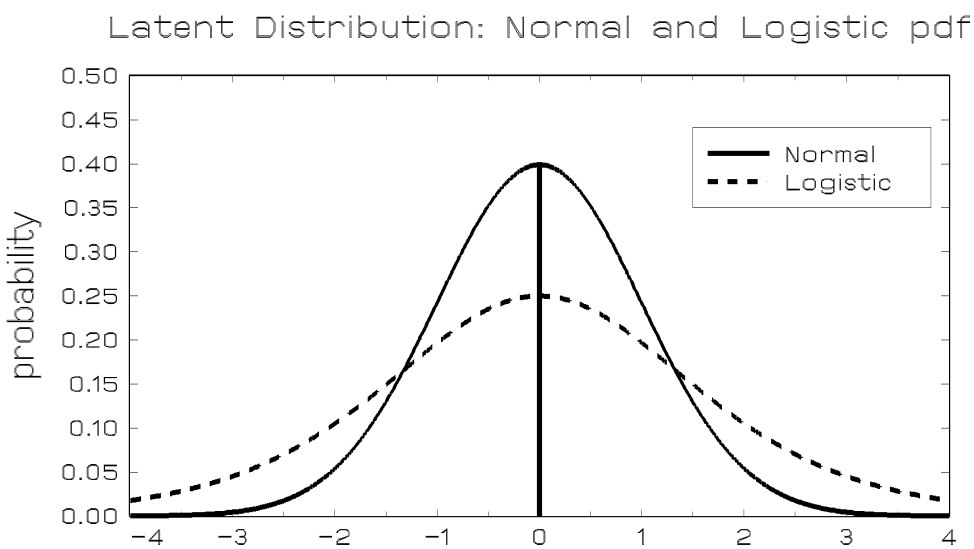
$$\begin{aligned}
Y_{ij} &= \text{dichotomous response of subject } i \text{ at timepoint } j \\
\boldsymbol{x}_{ij} &= (p+1) \times 1 \text{ vector of covariates} \\
\boldsymbol{\beta} &= (p+1) \times 1 \text{ vector of regression coefficients} \\
\upsilon_i &= \text{random subject effects distributed } \mathcal{NID}(0, \sigma_\upsilon^2)
\end{aligned}$$

Dichotomous Response and Threshold Concept

Continuous y_{ij} - an unobservable latent variable - related to dichotomous response Y_{ij} via "threshold concept"

- threshold value γ on y continuum

Response occurs $\quad Y_{ij} = 1 \quad$ if $\quad \gamma < y_{ij}$
otherwise, a response does not occur $(Y_{ij} = 0)$



The Threshold Concept in Practice

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

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

Great Day!

a day ...

Model for Latent Continuous Responses

$$y_{ij} = \boldsymbol{x}'_{ij}\boldsymbol{\beta} + \upsilon_i + \varepsilon_{ij}$$

- $\varepsilon_{ij} \sim$ std normal (mean 0, variance 1): probit regression
- $\varepsilon_{ij} \sim$ std logistic (mean 0, variance $\pi^2/3$): logistic regression

Underlying latent variable

- useful way of thinking of the problem
- not an essential assumption of the model
- used for intra-class correlation

$$ICC = \frac{\sigma_\upsilon^2}{\sigma_\upsilon^2 + 1} \quad \text{for probit (equals tetrachoric if } n = 2\text{)}$$

$$= \frac{\sigma_\upsilon^2}{\sigma_\upsilon^2 + \pi^2/3} \quad \text{for logistic}$$

Scaling of regression coefficients

Fixed-effects or marginal model - $\boldsymbol{\beta}$ estimates from logistic are larger in absolute value than from probit by

$$\approx \sqrt{\frac{\pi^2/3}{1}} = \sqrt{\frac{\text{std logistic variance}}{\text{std normal variance}}} = 1.8$$

- Amemiya (1981) suggests 1.6, Long (1997) suggests 1.7

Random-effects model - $\boldsymbol{\beta}$ estimates from random-effects model are larger in abs. value than fixed-effects or marginal model by

$$\approx \sqrt{d} = \sqrt{\frac{\sigma_\upsilon^2 + \sigma^2}{\sigma^2}} = \sqrt{\frac{\text{RE variance}}{\text{FE variance}}}$$

- d = design effect in sampling literature
- Zeger et. al. (1988) $\sigma^2 = (15/16)^2\pi^2/3$ for logistic

Random-Intercept Model *Within-Subjects / Between-Subjects models*

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

observed response

$$\log\left[\frac{Pr(Y_{ij}=1)}{1-Pr(Y_{ij}=1)}\right] = b_{0i} + b_{1i}\,Time_{ij}$$

latent response

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

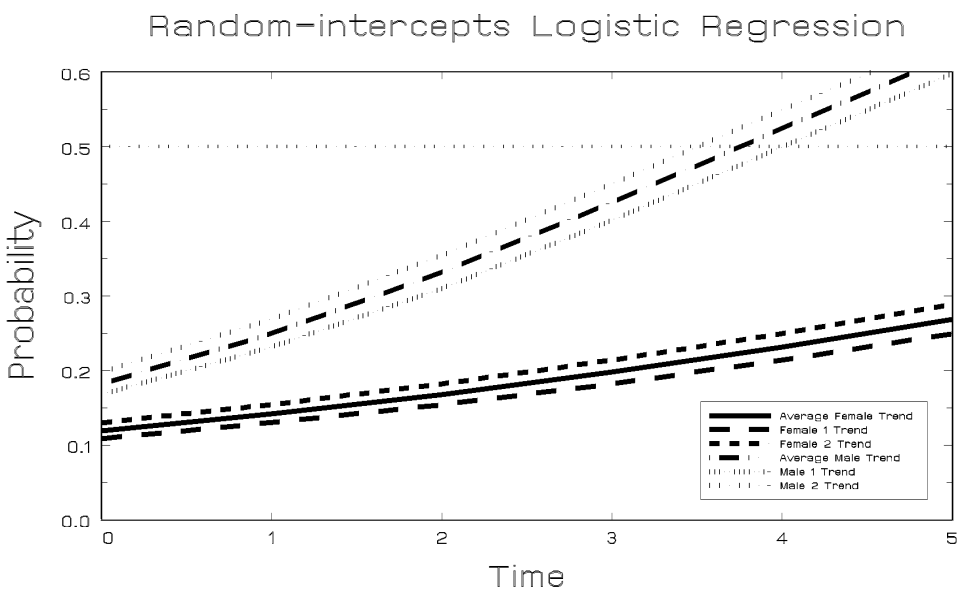
Between-subjects model - level 2 $(i = 1, \ldots, N)$

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

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

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

Random Intercept Logistic Model *in terms of probability*



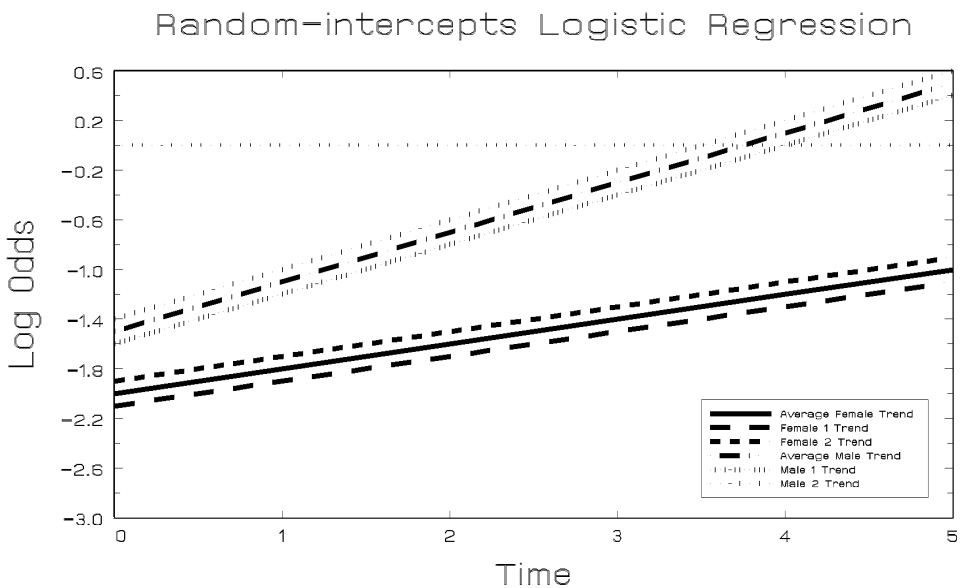
- Not linear in terms of probability

$$Pr(Y_{ij} = 1) \;=\; \frac{1}{1 + \exp\left[-\left(\beta_0 + \beta_1 G_i + \beta_2 T_j + \beta_3 (G_i \times T_j) + \upsilon_{0i}\right)\right]}$$

where G = Group $\quad T$ = Time

Random Intercept Logistic Model

in terms of log odds (logits)



- Linear in terms of log odds (logits)

$$\log\left[\frac{Pr(Y_{ij}=1)}{1-Pr(Y_{ij}=1)}\right] = \beta_0 + \beta_1 G_i + \beta_2 T_j + \beta_3 (G_i \times T_j) + \upsilon_{0i}$$

Random Intercept and Trend Model

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

latent response

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

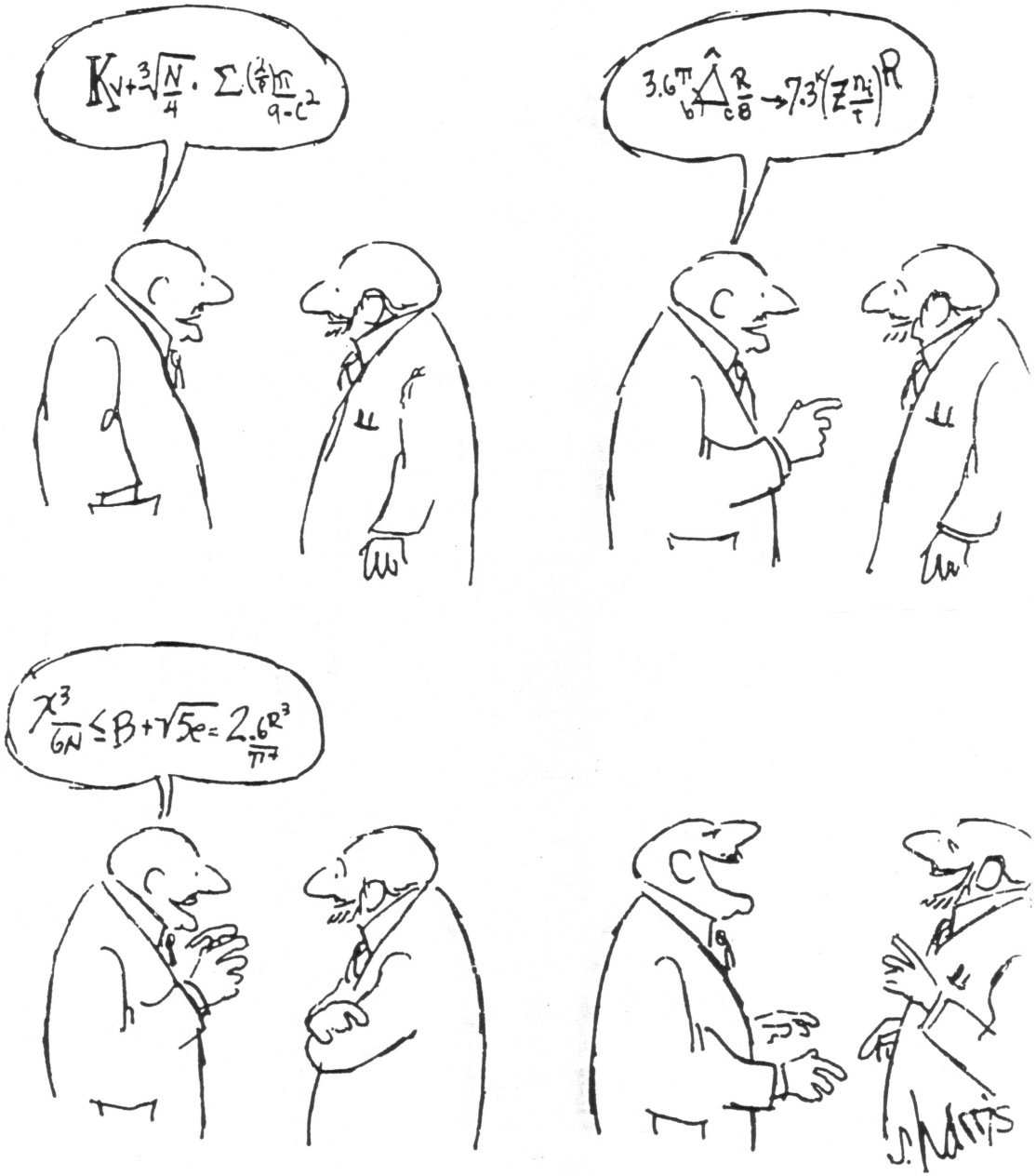
Between-subjects model - *level 2* $\quad (i = 1, \ldots, N)$

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

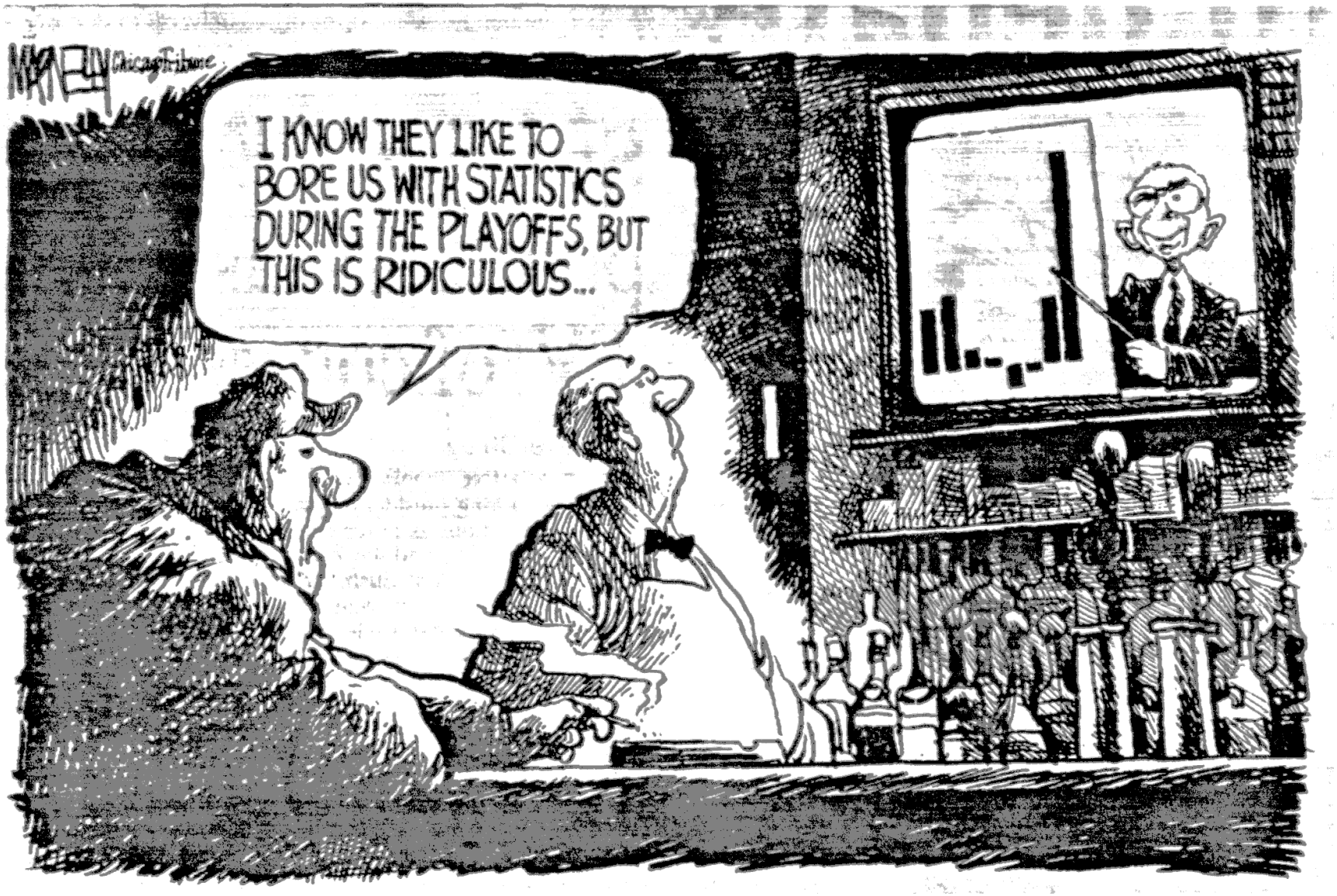
$$b_{1i} = \beta_1 + \beta_3 Grp_i + \upsilon_{1i}$$

$$\begin{bmatrix} \upsilon_{0i} \\ \upsilon_{1i} \end{bmatrix} \sim \mathcal{NID}\left\{ \begin{bmatrix} 0 \\ 0 \end{bmatrix}, \begin{bmatrix} \sigma^2_{\upsilon_0} & \sigma_{\upsilon_0 \upsilon_1} \\ \sigma_{\upsilon_0 \upsilon_1} & \sigma^2_{\upsilon_1} \end{bmatrix} \right\}$$

$$\varepsilon_{ij} \sim \mathcal{LID}(0, \pi^2/3)$$



© 1973 *The New Yorker* Magazine.


MANLEY Chicago Tribune
I KNOW THEY LIKE TO BORE US WITH STATISTICS DURING THE PLAYOFFS, BUT THIS IS RIDICULOUS...

Treatment-Related Change Across Time

NIMH Schizophrenia collaborative study on treatment related changes in overall severity (IMPS item # 79). Item 79, *Severity of Illness*, was scored as:

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

The experimental design and corresponding sample sizes:

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

Drug = Chlorpromazine, Fluphenazine, or Thioridazine

Main question of interest:

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

Descriptive Statistics

Observed proportions $\geq$ "moderately ill"

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

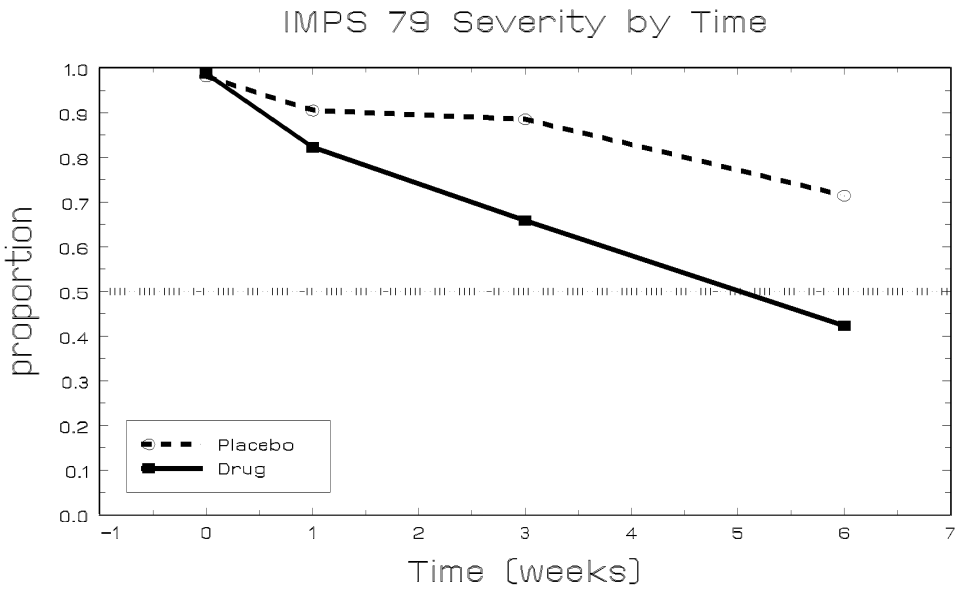
Observed odds $\geq$ "moderately ill"

	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"

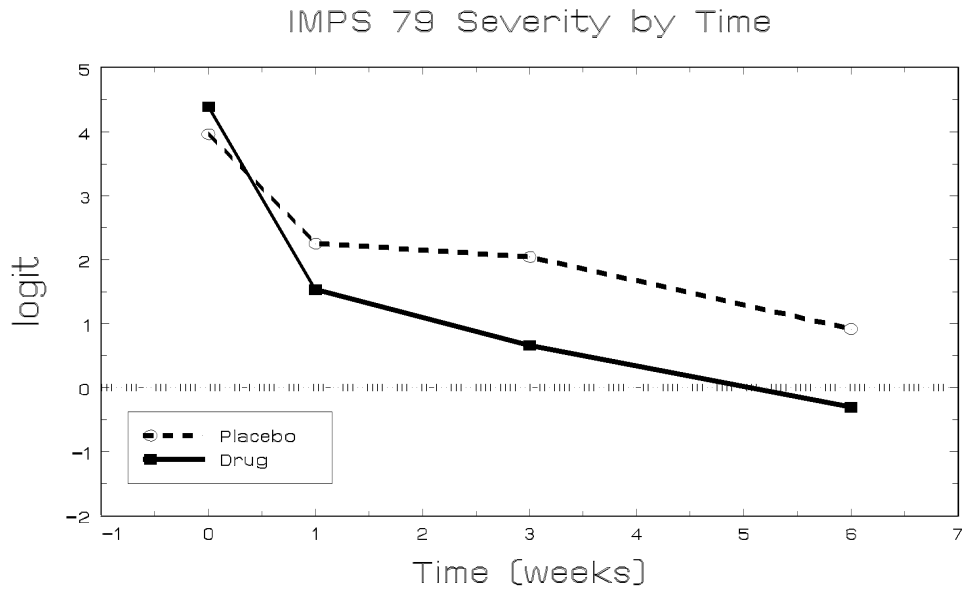
	week 0	week 1	week 3	week 6
placebo	3.96	2.25	2.04	.92
drug	4.39	1.53	.66	-.31
difference	-.43	.72	1.38	1.23
exp (odds ratio)	.65	2.05	3.99	3.42

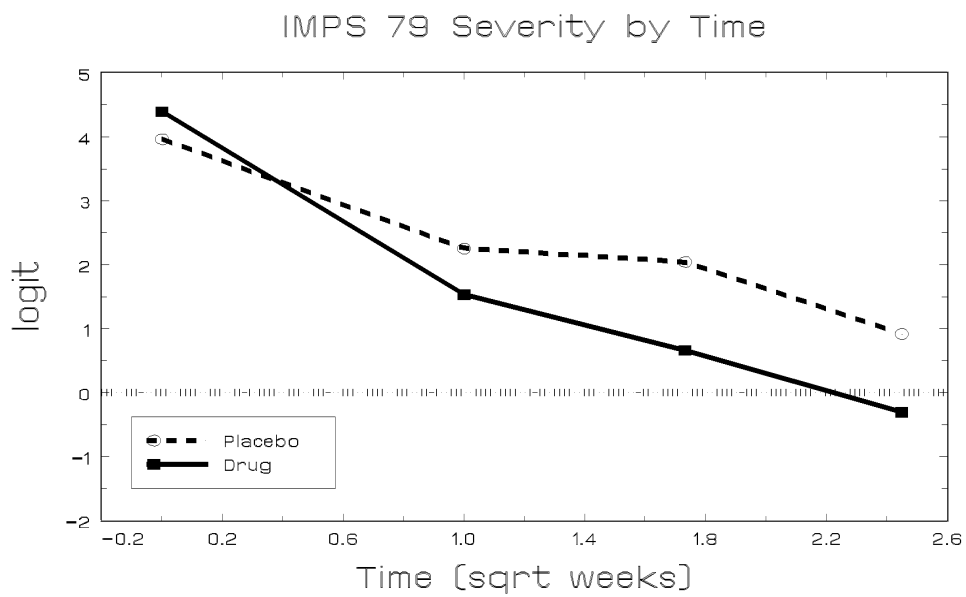
Observed Proportions across Time by Condition



- model is not linear in terms of probabilites

Observed Logits across Time by Condition



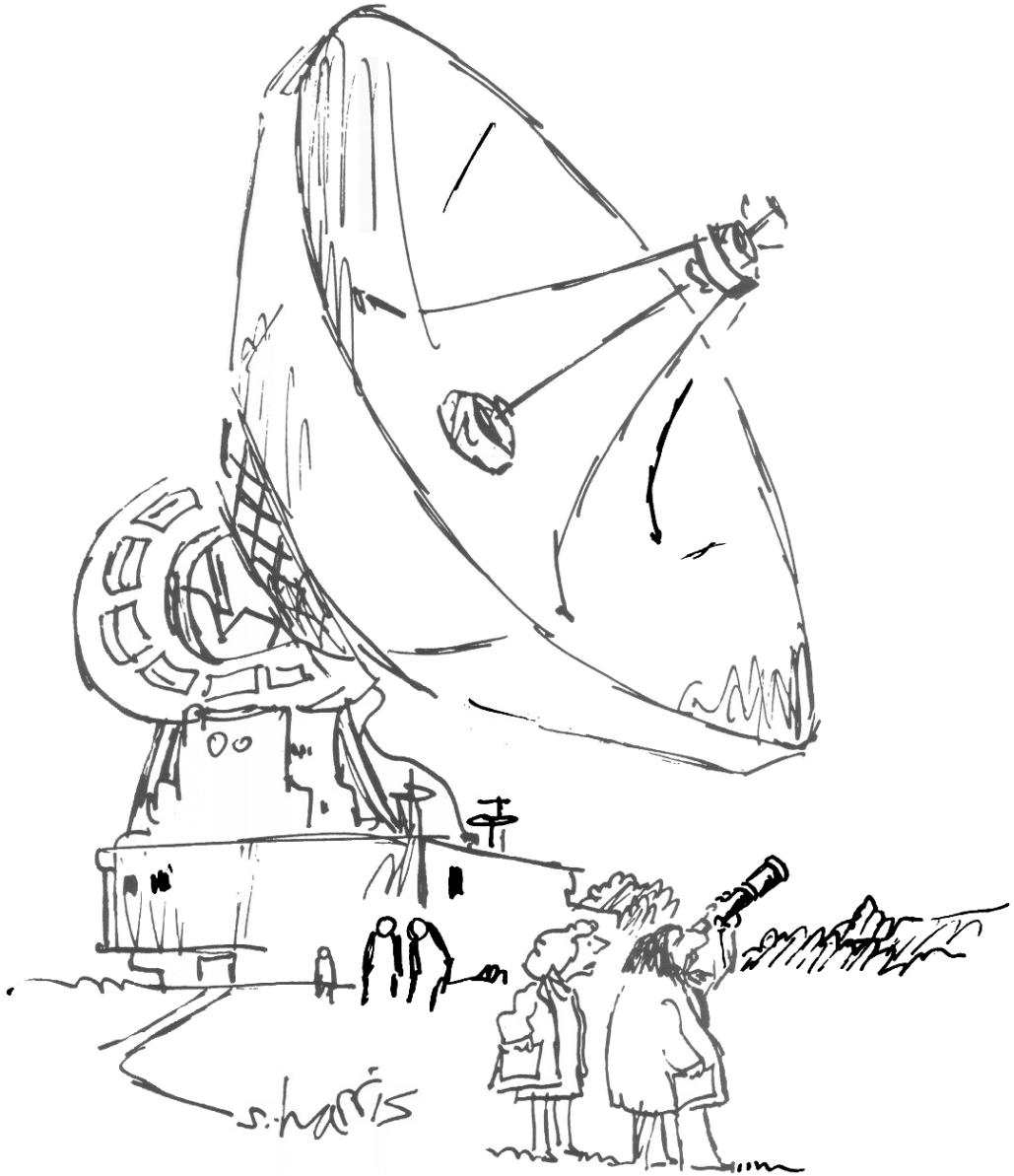


NIMH Schizophrenia Study - Severity of Illness (N = 437)
Logistic Regression ML Estimates - *Fixed effects model*

	estimates	se	z	$p <$
intercept	3.702	0.441	8.39	.001
Drug (0 = plc; 1 = drug)	-0.405	0.483	-0.84	.41
Time (sqrt week)	-1.112	0.233	-4.78	.001
Drug by Time	-0.418	0.256	-1.64	.11

$-2 \log L = 1362.06$

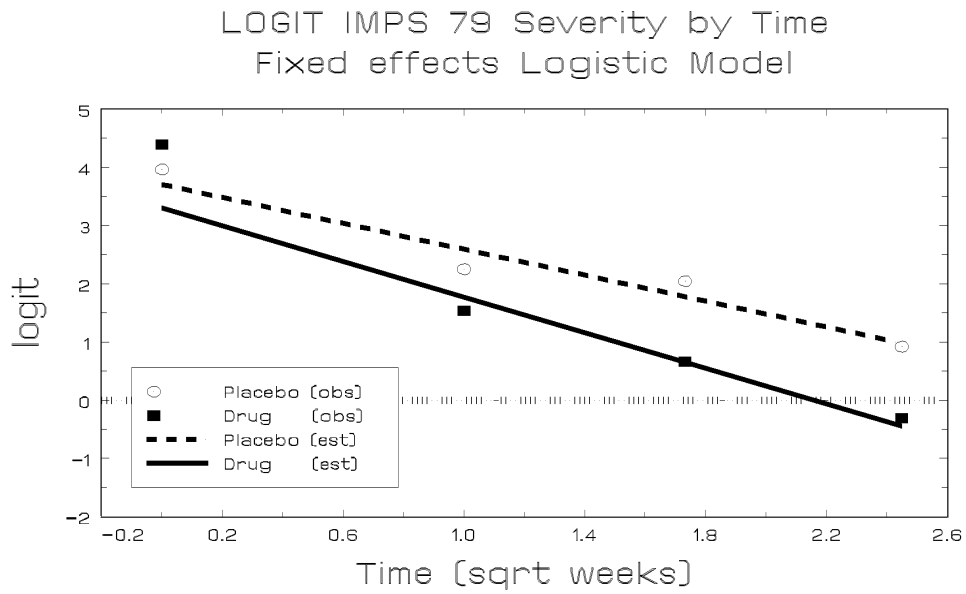
ok if data were cross-sectional longitudinal or if $\sigma_{\upsilon} = 0$



"JUST CHECKING."

Fitted Logits across Time by Condition

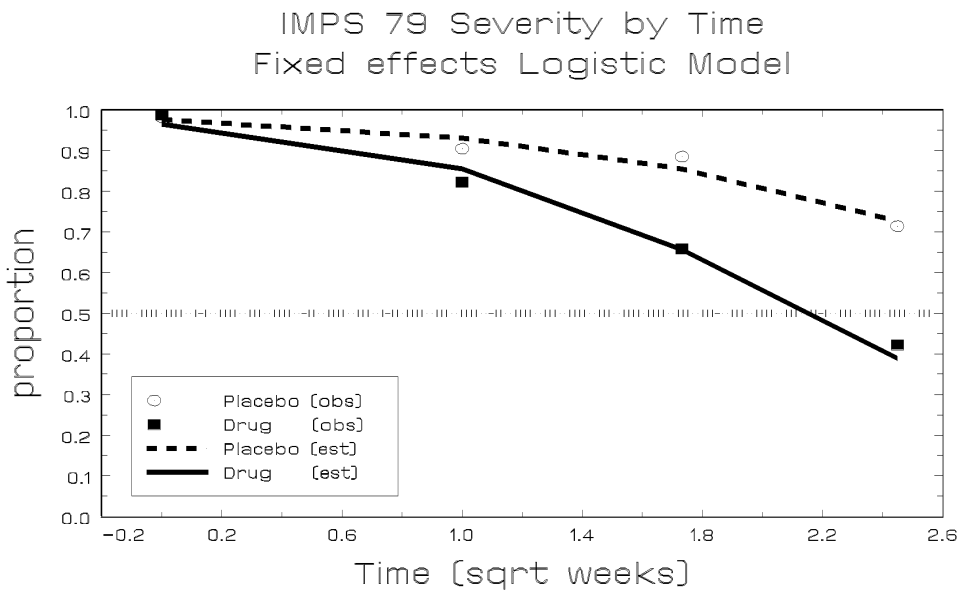
fixed-effects logistic regression model



$$\log\left[\frac{Pr(Y_{ij}=1)}{1-Pr(Y_{ij}=1)}\right] = 3.70 - .41\,D_i - 1.11\,T_j - .42\,(D_i \times T_j)$$

Fitted Proportions across Time by Condition

fixed-effects logistic regression model



$$Pr(Y_{ij} = 1) \;=\; \frac{1}{1 + \exp\left[-\left(3.70 - .41\,D_i - 1.11\,T_j - .42\,D_i T_j\right)\right]}$$

Within-Subjects / Between-Subjects components

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

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

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

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

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

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

NIMH Schizophrenia Study - Severity of Illness (N = 437)

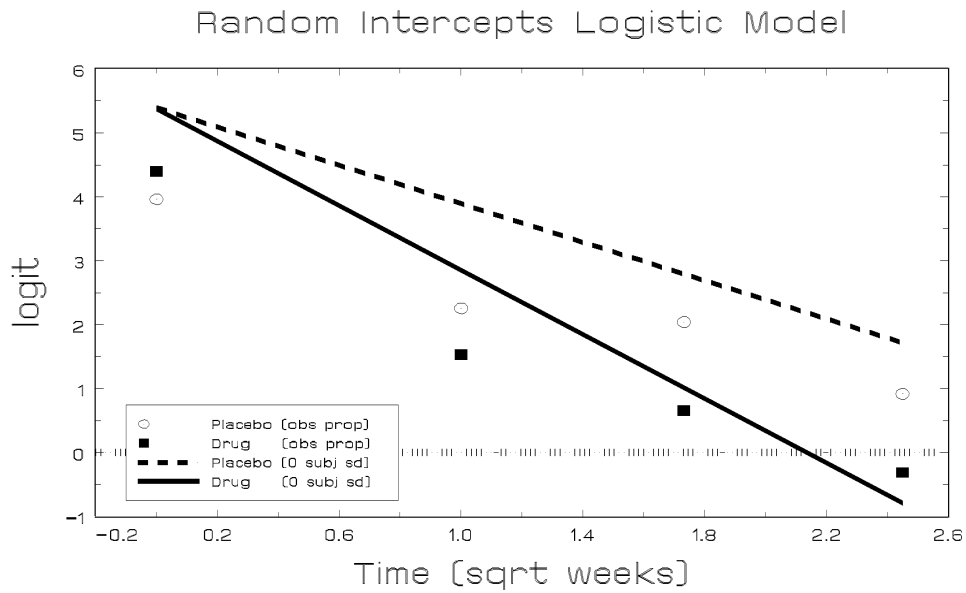
Logistic ML Estimates (se) - *random-intercepts model*

	estimates	se	z	$p <$
intercept	5.387	0.631	8.54	.001
Drug (0 = plc; 1 = drug)	-0.025	0.654	-0.04	.97
Time (sqrt week)	-1.500	0.291	-5.16	.001
Drug by Time	-1.015	0.334	-3.04	.0024
Intercept variance	4.478	0.947		

Intra-person correlation $= 4.478/(4.478 + \pi^2/3) = .58$

$-2 \log L = 1249.73 \qquad \chi^2_1 = 112.33$

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



$$\log\left[\frac{Pr(Y_{ij}=1)}{1-Pr(Y_{ij}=1)}\right] = 5.39 - .03\,D_i - 1.50\,T_j - 1.01\,(D_i \times T_j) + \upsilon_{0i}$$

$$\upsilon_{0i} \sim \mathcal{NID}(0, \hat{\sigma}_\upsilon^2 = 4.48)$$

$\hat{\boldsymbol{\beta}}$ assesses change in (conditional) logit due to $\boldsymbol{x}$ for subjects with the same value of υ_{0i}

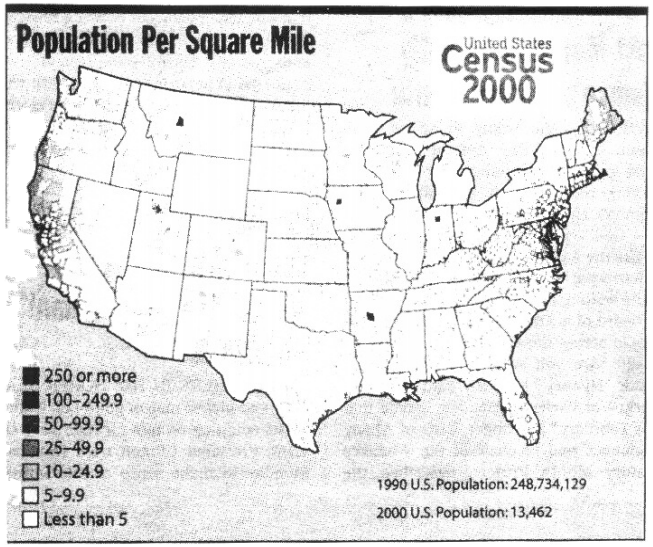
the ONION®

★★★

VOLUME 36 ISSUE 12 AMERICA'S FINEST NEWS SOURCE™ 6–12 APRIL 2000

U.S. Population At 13,462

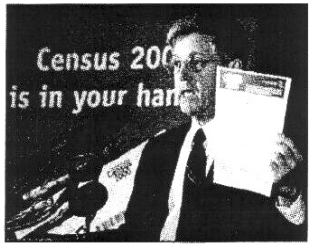
'We Don't Think Everybody Sent In Their Census Forms,' Say Officials



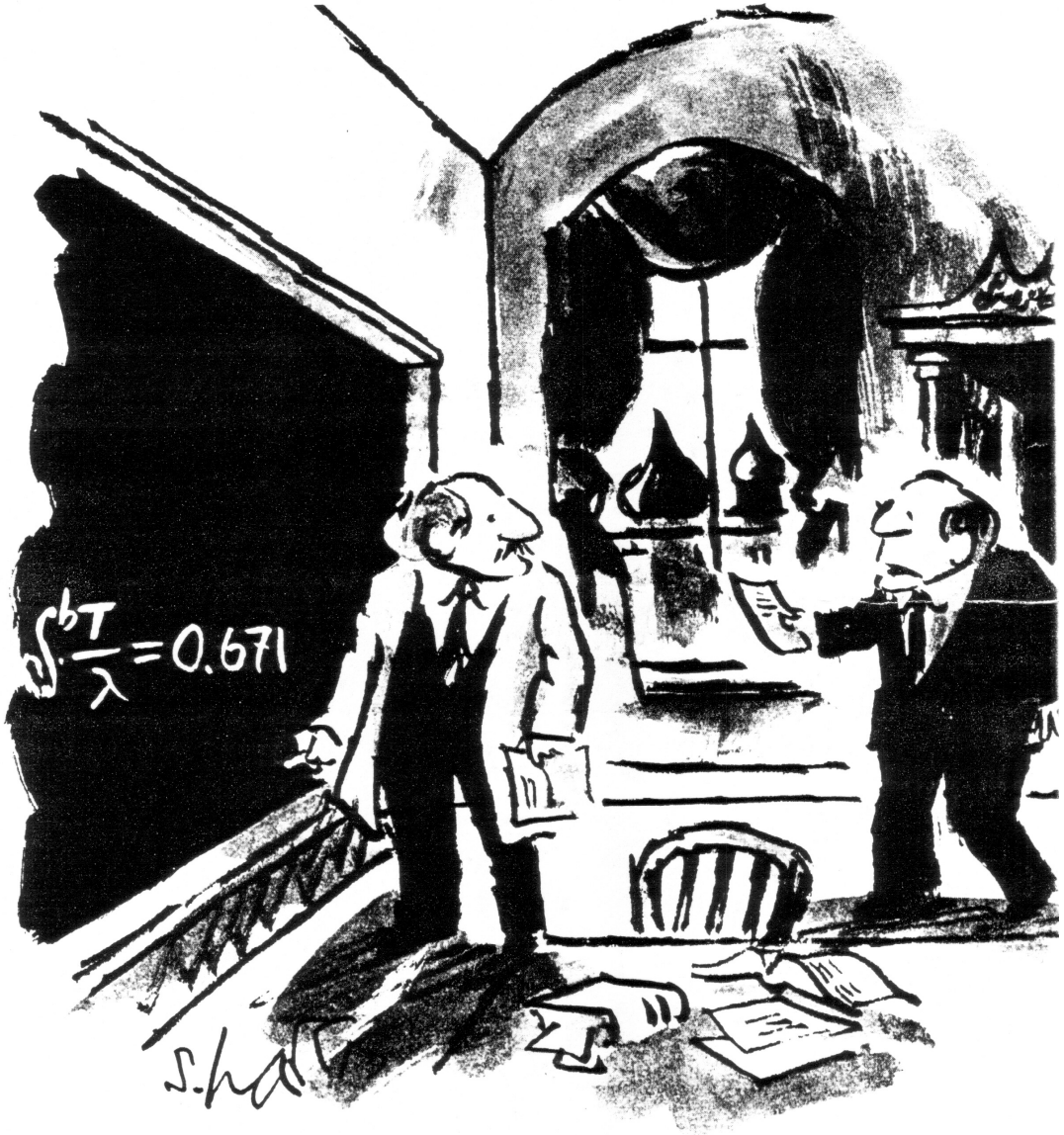
WASHINGTON, DC—With the April 1 deadline for returning Census 2000 forms finally passed, the Bureau of the Census announced Monday that the U.S. population stands at 13,462.

"We at the Census Bureau are shocked by the incredible decrease in the population that apparently took place in the 10 years since the last Census in 1990," Census Bureau director Kenneth Prewitt said. "A 1999 projection estimated the U.S. population at 274 mil-

see CENSUS page 9



Above: Census Bureau director Kenneth Prewitt.



"COMRADE — THE COMMISSAR OF MATHEMATICS WANTS IT TO EQUAL 29.86."

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}

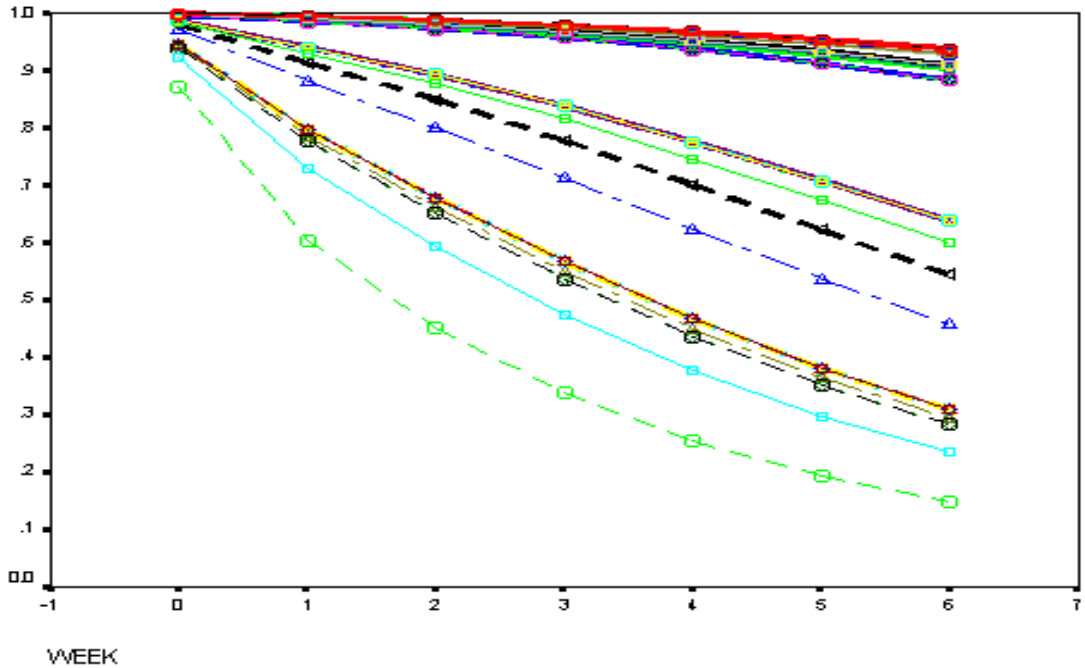
- referred to as "subject-specific"
- how a *subject's* response probability depends on $\boldsymbol{x}$

σ_υ^2 = 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) probabilities across time

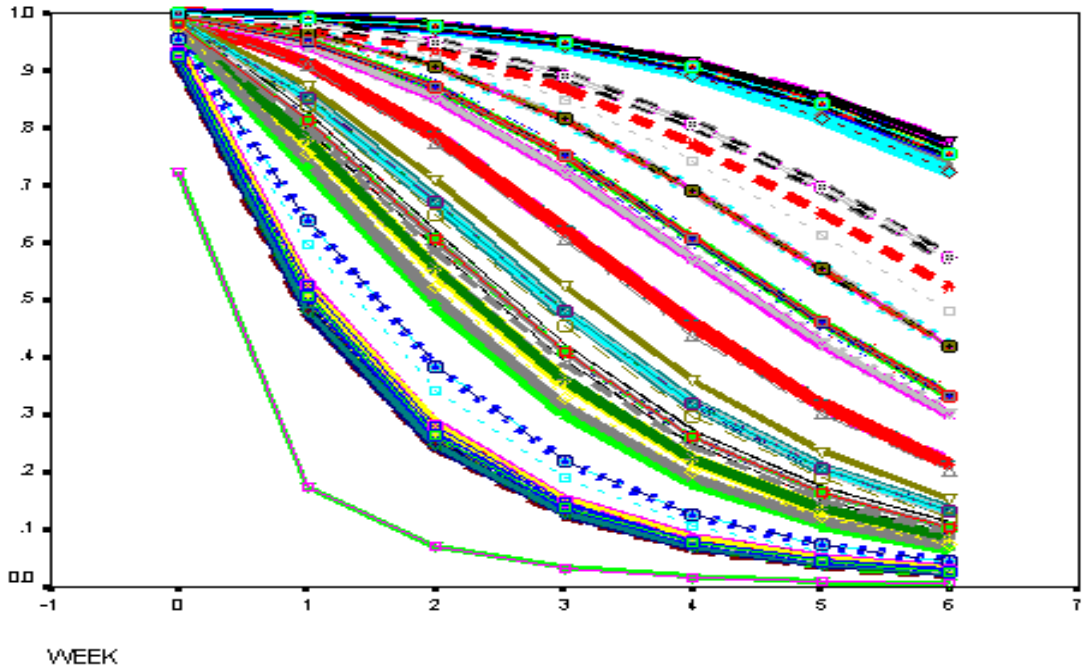
Random intercepts model - placebo group



$$P(Y_{ij} = 1) = \frac{1}{1 + \exp[-(5.39 - .03\, D_i - 1.50\, T_j - 1.01\, D_i T_j + \hat{v}_{0i})]}$$

Estimated (subject-specific) probabilities across time

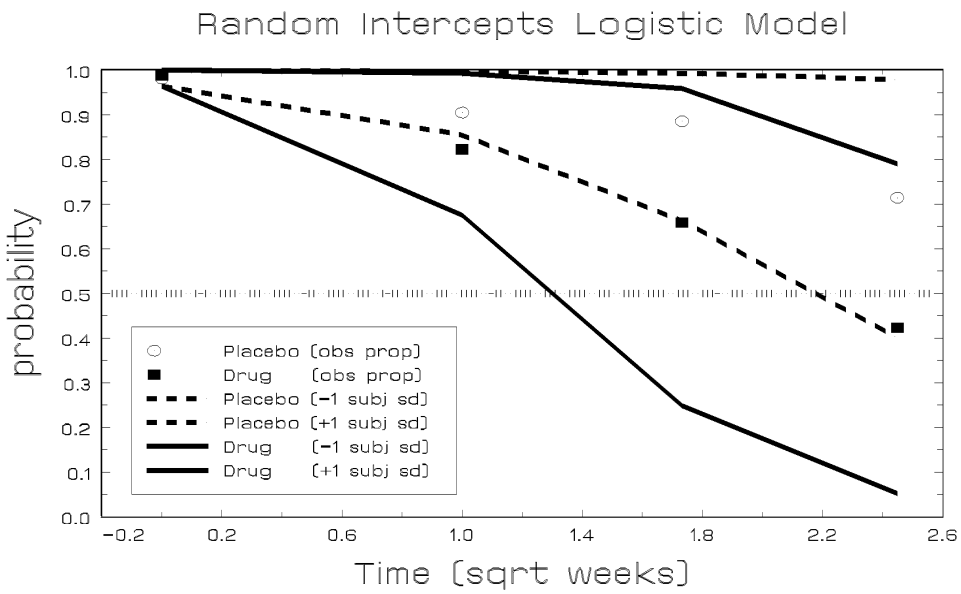
Random intercepts model - drug group



$$P(Y_{ij} = 1) = \frac{1}{1 + \exp[-(5.39 - .03\, D_i - 1.50\, T_j - 1.01\, D_i T_j + \hat{v}_{0i})]}$$

Estimated Subject-Specific Probabilites

random-intercepts logistic regression model



$$Pr(Y_{ij} = 1) = \frac{1}{1 + \exp\left[-\left(5.39 - .03\, D_i - 1.50\, T_j - 1.01\, D_i T_j + \upsilon_{0i}\right)\right]}$$

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

Model fit of observed marginal proportions

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

2. calculate marginalization factor

$$\hat{s} \;=\; \sqrt{\hat{d}} \;=\; \sqrt{(\hat{\sigma}_v^2 + \sigma^2)/\sigma^2} \;=\; \sqrt{\hat{\sigma}_v^2/\sigma^2 + 1}$$

- $\sigma = 1$ for probit or $\sigma = \pi/\sqrt{3}$ for logistic
- $\hat{d}$ is the design effect in the sampling literature

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

4. $\hat{\boldsymbol{p}}_i = \Phi(\hat{\mathbf{z}}_i)$ for probit and $\hat{\boldsymbol{p}}_i = \Psi(\hat{\mathbf{z}}_i)$ for logistic, Φ represents the normal cdf and Ψ the logistic cdf, *i.e.*, $1/[1 + \exp(-\hat{\mathbf{z}}_i)]$

notes:

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

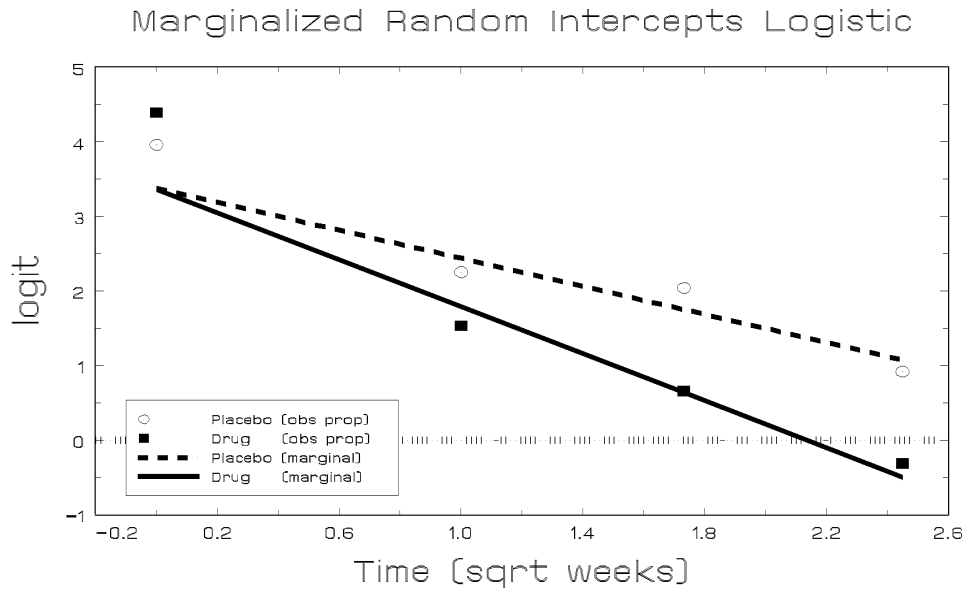
$$\hat{\boldsymbol{s}} = \frac{1}{\sigma}\left[\mathrm{Diag}(\hat{V}(\boldsymbol{y}_i))\right]^{1/2}$$

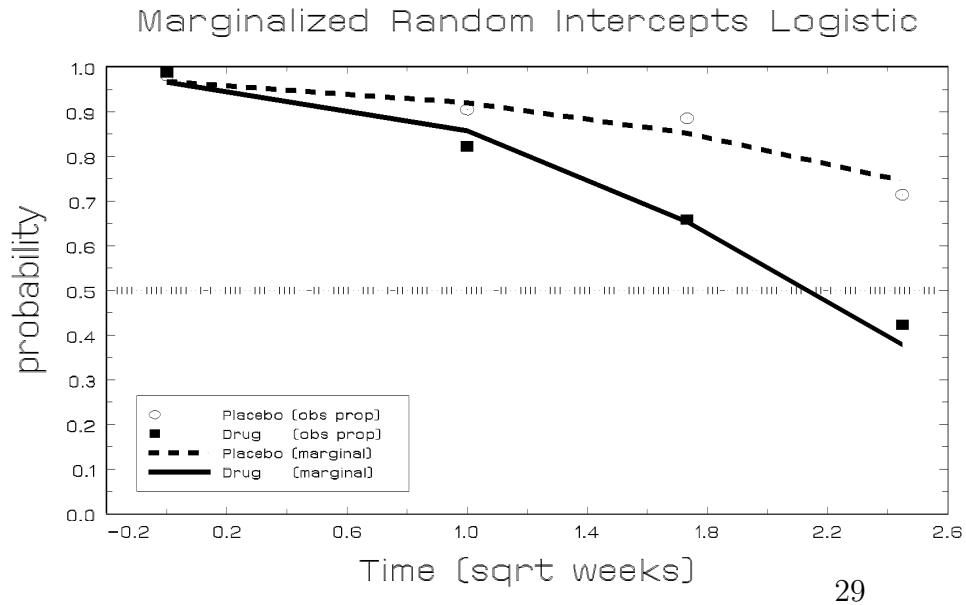
  - $\hat{V}(\boldsymbol{y}_i) = \boldsymbol{Z}_i\hat{\boldsymbol{\Sigma}}_{\upsilon}\boldsymbol{Z}_i' + \sigma^2\boldsymbol{I}_i$
  - $\boldsymbol{Z}_i$ = design matrix for random effects

  and perform element-wise division

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

Estimated Marginal Logits and Probabilities





SAS NLMIXED code: SCHZBINL.SAS

```
DATA one; INFILE 'c:\mixdemo\schizx1.dat';
INPUT id imps79 imps79b imps79o int tx week sweek txswk ;

/* get rid of observations with missing values */
IF imps79 > -9;

PROC FORMAT;
VALUE imps79b 0 = 'le mild' 1 = 'ge moderate';
VALUE tx 0 = 'placebo' 1 = 'drug';

/* fixed-effects logistic regression model */
PROC LOGISTIC DESCENDING;
MODEL imps79b = tx sweek tx*sweek;
RUN;

/* random intercept logistic regression via GLIMMIX */
PROC GLIMMIX DATA=one METHOD=QUAD(QPOINTS=21) NOCLPRINT;
CLASS id;
MODEL imps79b(DESC) = tx sweek tx*sweek / SOLUTION DIST=BINARY LINK=LOGIT;
RANDOM INTERCEPT / SUBJECT=id;
RUN;
```

```
/* random intercept logistic regression via NLMIXED */
PROC NLMIXED DATA=one QPOINTS=21;
PARMS b0=3.70 b1=-.40 b2=-1.11 b3=-.42 varu=1;
z = b0 + b1*tx + b2*sweek + b3*tx*sweek + u;
IF (imps79b=1) THEN
p = 1 / (1 + EXP(-z));
ELSE
p = 1 - (1 / (1 + EXP(-z)));
ll = LOG(p);
MODEL imps79b ~ GENERAL(ll);
RANDOM u ~ NORMAL(0,varu) SUBJECT=id;
ESTIMATE 'icc' varu/((((ATAN(1)*4)**2)/3)+varu);
RUN;
```

SAS IML code: SCHZBFIT1.SAS

```
TITLE1 'nimh schizophrenia data - estimated marginal probabilities';
PROC IML;
/* Results from nlmixed analysis:  random intercept model */;

/* covariate matrices for placebo and drug groups */;
 x0  =  {  1 0 0.00000 0,
           1 0 1.00000 0,
           1 0 1.73205 0,
           1 0 2.44949 0};
 x1  =  {  1 1 0.00000 0.00000,
           1 1 1.00000 1.00000,
           1 1 1.73205 1.73205,
           1 1 2.44949 2.44949};

/* nlmixed estimates of covariate effects and random effect variance */;
 beta  =  {5.387, -0.025, -1.500, -1.015};
 varu  =  {4.478};

/* marginalization of person-specific estimates */;
 pi    =  ATAN(1)*4;
 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 * varu * T(zvec) + emat;


/* marginalization factor */;
sdy = SQRT(VECDIAG(vary) / VECDIAG(emat));

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

grp0 = 1 / ( 1 + EXP(0 - z0));
grp1 = 1 / ( 1 + EXP(0 - z1));

print 'random intercept model';
print 'marginalization of person-specific estimates';
print 'marginal prob for group 0 - response' grp0 [FORMAT=8.4];
print 'marginal prob for group 1 - response' grp1 [FORMAT=8.4];
```

Random intercept and trend model

within-subjects / between-subjects components

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

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

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

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

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

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

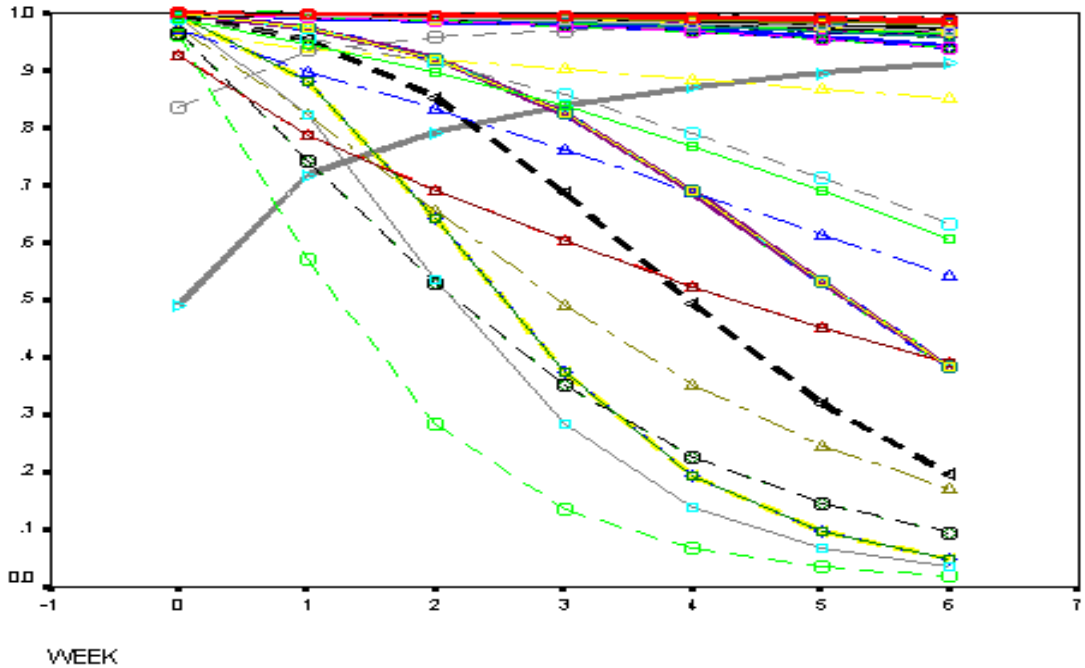
Logistic ML Estimates (se) - random intercept and trend model

	estimates	se	z	$p <$
intercept	5.928	0.948	6.25	.001
Drug (0 = plc; 1 = drug)	0.287	0.742	0.39	.70
Time (sqrt week)	-1.399	0.476	-2.94	.004
Drug by Time	-1.615	0.481	-3.36	.001
Variance-covariance terms				
Intercept var	6.975	2.908		
Int-Time covar	-2.111	1.210	$(r_{\upsilon_0\upsilon_1} = -.45)$	
Time var	3.096	1.161		

$-2 \log L = 1227.38, \quad \chi^2_2 = 21.95, p < .001$

Estimated (subject-specific) probabilities across time

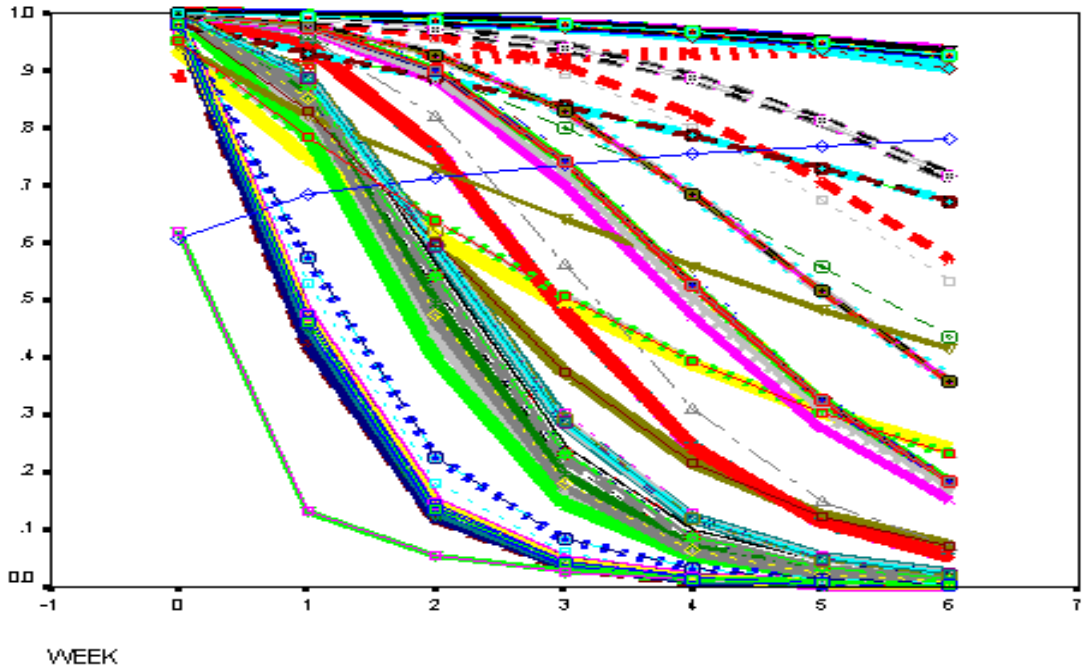
Random intercepts and trends model - placebo group



$$P(Y_{ij} = 1) = \frac{1}{1 + \exp[-(5.93 + .29\, D_i - 1.40\, T_j - 1.62\, D_i T_j + \hat{v}_{0i} + \hat{v}_{1i}\, T_j)]}$$

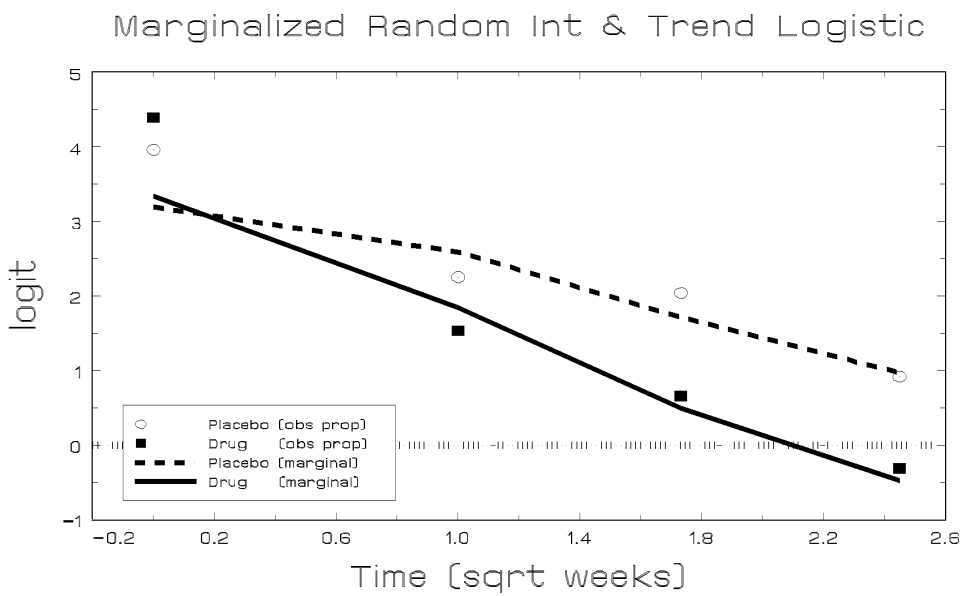
Estimated (subject-specific) probabilities across time

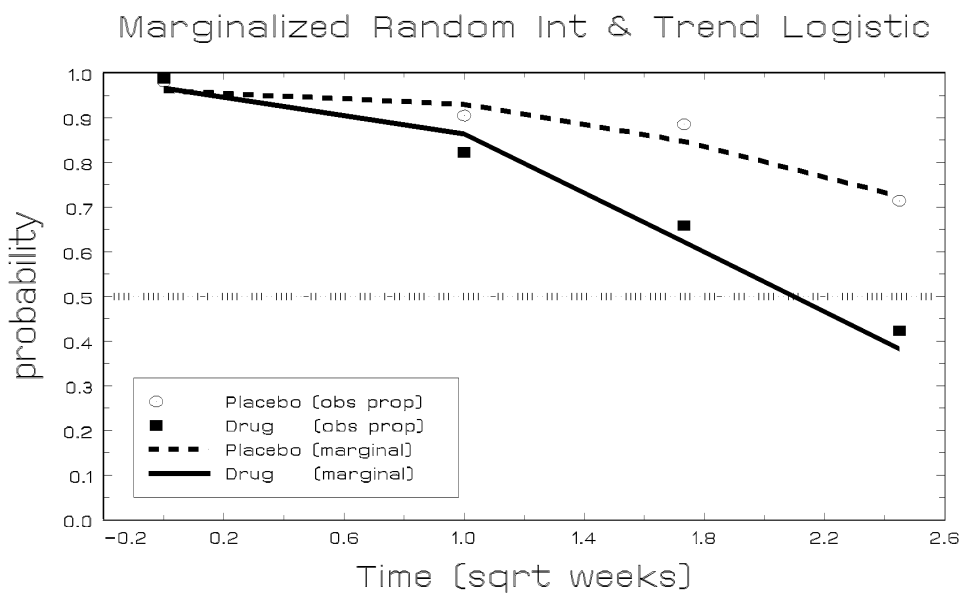
Random intercepts and trends model - drug group



$$P(Y_{ij} = 1) = \frac{1}{1 + \exp[-(5.93 + .29\, D_i - 1.40\, T_j - 1.62\, D_i T_j + \hat{v}_{0i} + \hat{v}_{1i}\, T_j)]}$$

Estimated Marginal Logits and Probabilities





SAS NLMIXED code: random-trend model, in SCHZBINL.SAS

```
/* random trend logistic regression via GLIMMIX */
PROC GLIMMIX DATA=one METHOD=QUAD(QPOINTS=11) NOCLPRINT;
CLASS id;
MODEL imps79b(DESC) = tx sweek tx*sweek / SOLUTION DIST=BINARY LINK=LOGIT;
RANDOM INTERCEPT sweek / SUBJECT=id TYPE=UN GCORR SOLUTION;
ODS LISTING EXCLUDE SOLUTIONR; ODS OUTPUT SOLUTIONR=ebest2;
RUN;

/* logistic random-trend model via NLMIXED */
PROC NLMIXED DATA=one QPOINTS=11;
PARMS b0=5.39 b1=-0.03 b2=-1.50 b3=-1.02 v0=4.48 c01=0 v1=1;
z = b0 + b1*tx + b2*sweek + b3*tx*sweek + u0 + u1*sweek;
IF (imps79b=1) THEN
p = 1 / (1 + EXP(-z));
ELSE
p = 1 - (1 / (1 + EXP(-z)));
ll = LOG(p);
MODEL imps79b ~ GENERAL(ll);
RANDOM u0 u1 ~ NORMAL([0,0], [v0,c01,v1]) SUBJECT=id OUT=ebest2b;
ESTIMATE 're corr' c01/SQRT(v0*v1);
RUN;
```

SAS IML code: SCHZBFIT2.SAS

```
TITLE1 'nimh schizophrenia Data - estimated marginal probabilities';
PROC IML;
/* results from nlmixed analysis:  random intercept & trend model */;

/* covariate matrices for placebo and drug groups */;
 x0  =  {  1 0 0.00000 0,
           1 0 1.00000 0,
           1 0 1.73205 0,
           1 0 2.44949 0};
 x1  =  {  1 1 0.00000 0.00000,
           1 1 1.00000 1.00000,
           1 1 1.73205 1.73205,
           1 1 2.44949 2.44949};

/* nlmixed estimates of covariate effects and random effect variance-covariance matrix */;
 beta  =  { 5.928, 0.287, -1.399, -1.615};
 varu  =  {6.975 -2.111,
            -2.111 3.096};

/* marginalization of person-specific estimates */;
 pi    =  ATAN(1)*4;
 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 * varu * T(zmat) + emat;


/* marginalization factor */;
sdy = SQRT(VECDIAG(vary) / VECDIAG(emat));


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

grp0 = 1 / ( 1 + EXP(0 - z0));
grp1 = 1 / ( 1 + EXP(0 - z1));

print 'random intercept and trend model';
print 'marginalization of person-specific estimates';
print 'marginal response probability for group 0' grp0 [FORMAT=8.4];
print 'marginal response probability for group 1' grp1 [FORMAT=8.4];
```

Logistic GEE as marginal model

$$logit_{ij} \;=\; \boldsymbol{x}'_{ij}\,\boldsymbol{\beta}$$

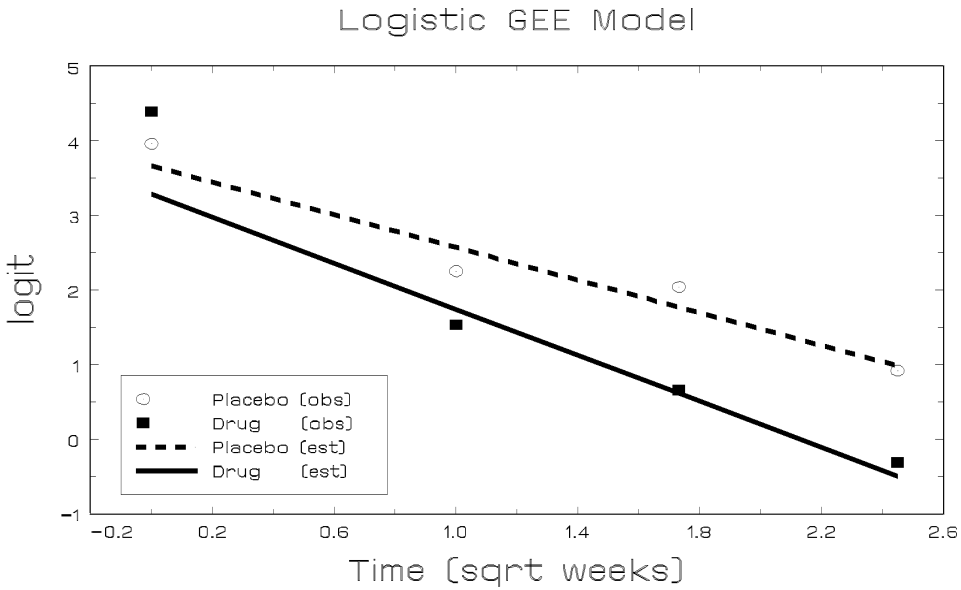
- Working correlation of repeated observations
  exchangeable (all are equal), AR(1), banded (m-dependent), unstructured
- robust standard errors
- does not include any subject-specific (random) effects, does not focus on heterogeneity
  β_0 = log odds of response among sub-population with $\boldsymbol{x} = 0$
  $\boldsymbol{\beta}$ = log odds ratio for response associated with unit changes in $\boldsymbol{x}$ in the population of subjects
- $\exp(\boldsymbol{\beta})$ = ratio of population frequencies
  - referred to as "population-averaged"

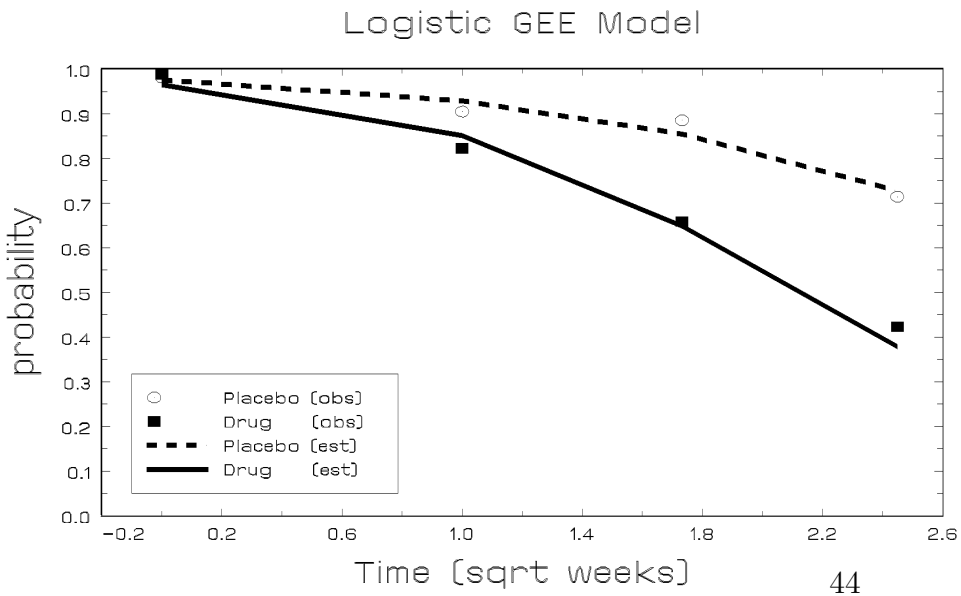
NIMH Schizophrenia Study - Severity of Illness (N = 437)
Logistic Regression GEE - exchangeable correlation structure

	GEE estimates	se	z	$p <$
intercept	3.661	0.485	7.54	.001
Drug (0 = plc; 1 = drug)	-0.381	0.521	-0.73	.46
Time (sqrt week)	-1.094	0.252	-4.35	.001
Drug by Time	-0.449	0.269	-1.67	.10

- non-significant drug by time interaction
- working corr based on data from 7 timepts (weeks 0 to 6)
- several have little data (wks 2, 4, 5) & wk 0 is near-constant
- very poorly estimated working correlation matrix
- analysis of 4 primary timepts and UN working corr yields significant interaction ($p < .047$)

Estimated Marginal Logits and Probabilities





SAS GENMOD code: GEE logistic regression - SCHZGEE.SAS

```
DATA one; INFILE 'c:\mixdemo\schizx1.dat';
INPUT id imps79 imps79b imps79o int tx week sweek txswk;

/* get rid of observations with missing values */
IF imps79 > -9;

/* get rid of weeks with very few observations */
IF week EQ 0 or week EQ 1 OR week EQ 3 OR week EQ 6;

PROC FORMAT;
VALUE imps79b 0 = 'le mild' 1 = 'ge moderate';
VALUE tx 0 = 'placebo' 1 = 'drug';

/* gee logistic regression model:  unstructured */
PROC GENMOD DESCENDING;
CLASS id week;
MODEL imps79b = tx sweek txswk / LINK=LOGIT DIST=BIN;
REPEATED SUBJECT=id / WITHIN=week CORRW TYPE=UN;
RUN;
```

Conclusions - mixed-effects logistic regression models useful for incomplete longitudinal dichotomous data

- can handle subjects measured incompletely or at different timepoints (missing data assumed MAR)
- degree of within-subjects variation on dichotomous outcome is important to consider (might have 3-timepoint study where 90% of subjects have same response across timepoints)
- subject-specific (or conditional) interpretation of regression coefficients
- generalizations to other categorical outcomes
  - ordinal outcomes - mixed-effects ordinal logistic regression
    - proportional odds model
    - partial or non-proportional odds model
  - nominal outcomes - mixed-effects nominal logistic regression