GEE for Longitudinal Data - Chapter 8

- GEE: generalized estimating equations (Liang & Zeger, 1986; Zeger & Liang, 1986)
- extension of GLM to longitudinal data analysis using quasi-likelihood estimation
- method is semi-parametric
  - estimating equations are derived without full specification of the joint distribution of a subject's obs (*i.e.*, $\boldsymbol{y}_i$)
- instead, specification of
  - likelihood for the (univariate) marginal distributions of y_{ij}
  - "working" correlation matrix for the vector of repeated observations from each subject

- Ballinger G.A. (2004). Using generalized estimating equations for longitudinal data analysis, *Organizational Research Methods*, 7:127-150.
- Diggle P.J., Heagerty P., Liang K.-Y., Zeger S.L. (2002). *Analysis of Longitudinal Data, 2nd edition*, New York: Oxford University Press.
- Dunlop D.D. (1994). Regression for longitudinal data: a bridge from least squares regression, *The American Statistician*, 48:299-303.
- Hardin J.W., Hilbe J.M. (2003). *Generalized Estimating Equations*, New York: Chapman and Hall.
- Hu F.B., Goldberg J., Hedeker D., Flay B.R., Pentz M.A. (1998). A comparison of generalized estimating equation and random-effects approaches to analyzing binary outcomes from longitudinal studies: illustrations from a smoking prevention study, *American Journal of Epidemiology*, 147:694-703.
  available at: http://www.uic.edu/classes/bstt/bstt513/pubs.html
- Norton E.C., Bieler G.S., Ennett S.T., Zarkin G.A. (1996). Analysis of prevention program effectiveness with clustered data using generalized estimating equations, *Journal of Consulting and Clinical Psychology*, 64:919-926.
- Sheu C.-F. (2000). Regression analysis of correlated binary outcomes, *Behavior Research Methods, Instruments, and Computers*, 32:269-273.
- Zorn C.J.W. (2001). Generalized estimating equation models for correlated data: a review with applications, *American Journal of Political Science*, 45:470-490.

GEE Overview

- GEEs have consistent and asymptotically normal solutions, even with mis-specification of the correlation structure
- Avoids need for multivariate distributions by only assuming a functional form for the marginal distribution at each timepoint (*i.e.*, y_{ij})
- The covariance structure is treated as a nuisance
- Relies on the independence across subjects to estimate consistently the variance of the regression coefficients (even when the assumed correlation structure is incorrect)

GEE Method outline

1. Relate the marginal reponse $\mu_{ij} = E(y_{ij})$ to a linear combination of the covariates

$$g(\mu_{ij}) = \boldsymbol{x}'_{ij}\boldsymbol{\beta}$$

- y_{ij} is the response for subject i at time j
- $\boldsymbol{x}_{ij}$ is a $p \times 1$ vector of covariates
- $\boldsymbol{\beta}$ is a $p \times 1$ vector of unknown regression coefficients
- $g(\cdot)$ is the link function

2. Describe the variance of y_{ij} as a function of the mean

$$V(y_{ij}) = v(\mu_{ij})\phi$$

- ϕ is a possibly unknown scale parameter
- $v(\cdot)$ is a known variance function

Link and Variance Functions

- Normally-distributed response

$$\begin{aligned} g(\mu_{ij}) &= \mu_{ij} \quad \text{“Identity link”} \\ v(\mu_{ij}) &= 1 \\ V(y_{ij}) &= \phi \end{aligned}$$

- Binary response (Bernoulli)

$$\begin{aligned} g(\mu_{ij}) &= \log[\mu_{ij}/(1-\mu_{ij})] \quad \text{“Logit link”} \\ v(\mu_{ij}) &= \mu_{ij}(1-\mu_{ij}) \\ \phi &= 1 \end{aligned}$$

- Poisson response

$$\begin{aligned} g(\mu_{ij}) &= \log(\mu_{ij}) \quad \text{“Log link”} \\ v(\mu_{ij}) &= \mu_{ij} \\ \phi &= 1 \end{aligned}$$

Gee Method outline

3. Choose the form of a $n \times n$ "working" correlation matrix $\boldsymbol{R}_i$ for each $\boldsymbol{y}_i$

- the (j, j') element of $\boldsymbol{R}_i$ is the known, hypothesized, or estimated correlation between y_{ij} and $y_{ij'}$
- This working correlation matrix $\boldsymbol{R}_i$ may depend on a vector of unknown parameters $\boldsymbol{\alpha}$, which is assumed to be the same for all subjects
- Although this correlation matrix can differ from subject to subject, we usually use a working correlation matrix $\boldsymbol{R}_i \approx$ average dependence among the repeated observations over subjects

aside: not well-suited to irregular measurements across time because time is treated categorically

Comments on "working" correlation matrix

- should choose form of $\boldsymbol{R}$ to be consistent with empirical correlations
- GEE method yields consistent estimates of regression coefficients $\boldsymbol{\beta}$ and their variances (thus, standard errors), even with mis-specification of the structure of the covariance matrix
- Loss of efficiency from an incorrect choice of $\boldsymbol{R}$ is lessened as the number of subjects gets large

From O'Muircheartaigh & Francis (1981) Statistics: A Dictionary of Terms and Ideas

- "an estimator (of some population parameter) based on a sample of size N will be consistent if its value gets closer and closer to the true value of the parameter as N increases"
- "... the best test procedure (*i.e.*, the efficient test) will be that with the smallest type II error (or largest power)"

Working Correlation Structures

- Exchangeable: $\boldsymbol{R}_{jj'} = \rho$, all of the correlations are equal
- AR(1): $\boldsymbol{R}_{jj'} = \rho^{|j-j'|}$
- Stationary m-dependent (Toeplitz):

$$\boldsymbol{R}_{jj'} = \begin{cases} \rho_{|j-j'|} & \text{if } j - j' \leq m \\ 0 & \text{if } j - j' > m \end{cases}$$

- Unspecified (or unstructured) $\boldsymbol{R}_{jj'} = \rho_{jj'}$
  - estimate all $n(n-1)/2$ correlations of $\boldsymbol{R}$
  - most efficient, but most useful when there are relatively few timepoints (with many timepoints, estimation of the $n(n-1)/2$ correlations is not parsimonious)
  - missing data complicates estimation of $\boldsymbol{R}$

GEE Estimation

- Define $\boldsymbol{A}_i$ $= n \times n$ diagonal matrix with $V(\mu_{ij})$ as the jth diagonal element
- Define $\boldsymbol{R}_i(\boldsymbol{\alpha})$ $= n \times n$ "working" correlation matrix (of the n repeated measures)

Working variance–covariance matrix for $\boldsymbol{y}_i$ equals

$$V(\boldsymbol{\alpha}) = \phi \boldsymbol{A}_i^{1/2} \boldsymbol{R}_i(\boldsymbol{\alpha}) \boldsymbol{A}_i^{1/2}$$

For normally distributed outcomes, $V(\boldsymbol{\alpha}) = \phi \boldsymbol{R}_i(\boldsymbol{\alpha})$

GEE estimator of $\boldsymbol{\beta}$ is the solution of

$$\sum_{i=1}^{N} \boldsymbol{D}_i' \left[V(\hat{\boldsymbol{\alpha}})\right]^{-1} (\boldsymbol{y}_i - \boldsymbol{\mu}_i) \;=\; 0,$$

where $\hat{\boldsymbol{\alpha}}$ is a consistent estimate of $\boldsymbol{\alpha}$ and $\boldsymbol{D}_i = \partial \boldsymbol{\mu}_i / \partial \boldsymbol{\beta}$

e.g., normal case, $\boldsymbol{\mu}_i = \boldsymbol{X}_i \boldsymbol{\beta}$, $\boldsymbol{D}_i = \boldsymbol{X}_i$, and $V(\hat{\boldsymbol{\alpha}}) = \hat{\phi} \boldsymbol{R}_i(\hat{\boldsymbol{\alpha}})$

$$\sum_{i=1}^{N} \boldsymbol{X}_i' \left[\boldsymbol{R}_i(\hat{\boldsymbol{\alpha}})\right]^{-1} (\boldsymbol{y}_i - \boldsymbol{X}_i \boldsymbol{\beta}) \;=\; 0,$$

$$\hat{\boldsymbol{\beta}} = \left[\sum_{i=1}^{N} \boldsymbol{X}_i' \left[\boldsymbol{R}_i(\hat{\boldsymbol{\alpha}})\right]^{-1} \boldsymbol{X}_i\right]^{-1} \left[\sum_{i=1}^{N} \boldsymbol{X}_i' \left[\boldsymbol{R}_i(\hat{\boldsymbol{\alpha}})\right]^{-1} \boldsymbol{y}_i\right]$$

$\Rightarrow$ akin to weighted least-squares (WLS) estimator

$\Rightarrow$ more generally, because solution only depends on the mean and variance of y, these are quasi-likelihood estimates

GEE solution

Iterate between the quasi-likelihood solution for $\boldsymbol{\beta}$ and a robust method for estimating $\boldsymbol{\alpha}$ as a function of $\boldsymbol{\beta}$

1. Given estimates of $\boldsymbol{R}_i(\boldsymbol{\alpha})$ and ϕ, calculate estimates of $\boldsymbol{\beta}$ using iteratively reweighted LS
2. Given estimates of $\boldsymbol{\beta}$, obtain estimates of $\boldsymbol{\alpha}$ and ϕ. For this, calculate Pearson (or standardized) residuals

$$r_{ij} = (y_{ij} - \hat{\mu}_{ij}) / \sqrt{[V(\hat{\boldsymbol{\alpha}})]_{jj}}$$

and use these residuals to consistently estimate $\boldsymbol{\alpha}$ and ϕ (Liang & Zeger, 1986, present estimators for several different working correlation structures)

Inference

$V(\hat{\boldsymbol{\beta}})$: square root of diagonal elements yield std errors for $\hat{\boldsymbol{\beta}}$

GEE provides two versions of these (with $\hat{\boldsymbol{V}}_i$ denoting $V_i(\hat{\boldsymbol{\alpha}})$)

1. Naive or "model-based"

$$V(\hat{\boldsymbol{\beta}}) = \left[\sum_i^N \boldsymbol{D}_i' \hat{\boldsymbol{V}}_i^{-1} \boldsymbol{D}_i\right]^{-1}$$

2. Robust or "empirical"

$$V(\hat{\boldsymbol{\beta}}) = \boldsymbol{M}_0^{-1} \boldsymbol{M}_1 \boldsymbol{M}_0^{-1},$$

$$\boldsymbol{M}_0 = \sum_i^N \boldsymbol{D}_i' \hat{\boldsymbol{V}}_i^{-1} \boldsymbol{D}_i$$

$$\boldsymbol{M}_1 = \sum_i^N \boldsymbol{D}_i' \hat{\boldsymbol{V}}_i^{-1} (\boldsymbol{y}_i - \hat{\boldsymbol{\mu}}_i)(\boldsymbol{y}_i - \hat{\boldsymbol{\mu}}_i)' \hat{\boldsymbol{V}}_i^{-1} \boldsymbol{D}_i$$

- notice, if $\hat{\boldsymbol{V}}_i = (\boldsymbol{y}_i - \hat{\boldsymbol{\mu}}_i)(\boldsymbol{y}_i - \hat{\boldsymbol{\mu}}_i)'$ then the two are equal (this occurs only if the true correlation structure is correctly modeled)
- In the more general case, the robust or "sandwich" estimator provides a consistent estimator of $V(\hat{\boldsymbol{\beta}})$ even if the working correlation structure $\boldsymbol{R}_i(\boldsymbol{\alpha})$ is not the true correlation of $\boldsymbol{y}_i$

GEE vs MRM

- GEE not concerned with $V(\boldsymbol{y}_i)$
- GEE yields both robust and model-based std errors for $\hat{\boldsymbol{\beta}}$; MRM, in common use, only provides model-based
- GEE solution for all kinds of outcomes; MRM needs to be derived for each
- For non-normal outcomes, GEE provides population-averaged (or marginal) estimates of $\boldsymbol{\beta}$, whereas MRM yields subject-specific (or conditional) estimates
- GEE assumption regarding missing data is more stringent (MCAR) than MRM (which assumes MAR)

Example 8.1: Using the NIMH Schizophrenia dataset, this handout has PROC GENMOD code and output from several GEE analyses varying the working correlation structure. (SAS code and output)

http://tigger.uic.edu/ hedeker/schizgee.txt

GEE Example: Smoking Cessation across Time

Gruder, Mermelstein *et al.*, (1993) JCCP

- 489 subjects measured across 4 timepoints following an intervention designed to help them quit smoking
- Subjects were randomized to one of three conditions
  - control, self-help manuals
  - tx1, manuals plus group meetings (*i.e.*, discussion)
  - tx2, manuals plus enhanced group meetings (*i.e.*, social support)
- Some subjects randomized to tx1 or tx2 never showed up to any meetings following the phone call informing them of where the meetings would take place
- dependent variable: smoking status at particular timepoint was assessed via phone interviews

In Gruder *et al.*, , four groups were formed for the analysis:

1. Control: randomized to the control condition
2. No-show: randomized to receive a group treatment, but never showed up to the group meetings
3. tx1: randomized to and received group meetings
4. tx2: randomized to and received enhanced group meetings

and these four groups were compared using Helmert contrasts:

Group	$H1$	$H2$	$H3$
Control	-1	0	0
No-show	1/3	-1	0
tx1	1/3	1/2	-1
tx2	1/3	1/2	1

Interpretation of Helmert Contrasts

$H1$: test of whether randomization to group versus control influenced subsequent cessation.

$H2$: test of whether showing up to the group meetings influenced subsequent cessation.

$H3$: test of whether the type of meeting influenced cessation.

note: $H1$ is an experimental comparison, but $H2$ and $H3$ are quasi-experimental

Examination of possible confounders: baseline analysis revealed that groups differed in terms of race (w vs nw), so race was included in subsequent analyses involving group

Table 8.1 Point Prevalence Rates (N) of Abstinence over Time by Group

Group	End-of-Program (T1)	6 months (T2)	12 months (T3)	24 months (T4)
No Contact Control	17.4	7.2	18.5	18.2
	(109)	(97)	(92)	(77)
No Shows	26.8	18.9	18.6	18.7
	(190)	(175)	(161)	(139)
Discussion	33.7	14.6	16.3	22.9
	(86)	(82)	(80)	(70)
Social Support	49.0	20.0	24.0	25.6
	(104)	(100)	(96)	(86)

Table 8.2 Correlation of Smoking Abstinence (y/n) Across Time

	T1	T2	T3	T4
T1	1.00	0.33	0.29	0.26
T2	0.33	1.00	0.48	0.34
T3	0.29	0.48	1.00	0.49
T4	0.26	0.34	0.49	1.00

Working Correlation choice:

- exchangeable does not appear like a good choice since the correlations are not approximately equal
- neither the AR(1) nor the m-dependent structures appear reasonable because the correlations within a time lag vary
- unspecified appears to be the most reasonable choice

GEE models - binary outcome, logit, $\boldsymbol{R}$ = UN, $T = 0, 1, 2, 4$

Model 1

$$\eta_{ij} = \beta_0 + \beta_1 T_j + \beta_2 T_j^2 + \beta_3 H1_i + \beta_4 H2_i + \beta_5 H3_i + \beta_6 Race_i$$

Model 2

$$\begin{aligned}\eta_{ij} = {} & \beta_0 + \beta_1 T_j + \beta_2 T_j^2 + \beta_3 H1_i + \beta_4 H2_i + \beta_5 H3_i + \beta_6 Race_i \\ & + \beta_7 (H1_i \times T_j) + \beta_8 (H2_i \times T_j) + \beta_9 (H3_i \times T_j)\end{aligned}$$

Model 3

$$\begin{aligned}\eta_{ij} = {} & \beta_0 + \beta_1 T_j + \beta_2 T_j^2 + \beta_3 H1_i + \beta_4 H2_i + \beta_5 H3_i + \beta_6 Race_i \\ & + \beta_7 (H1_i \times T_j) + \beta_8 (H2_i \times T_j) + \beta_9 (H3_i \times T_j) \\ & + \beta_{10} (H1_i \times T_j^2) + \beta_{11} (H2_i \times T_j^2) + \beta_{12} (H3_i \times T_j^2)\end{aligned}$$

Table 8.3 Smoking Status (0, Smoking; 1, Not Smoking) Across Time ($N = 489$) — GEE Logistic Parameter Estimates (Est.), Standard Errors (SE), and p-Values

	Model 1			Model 2			Model 3		
Parameter	Est.	SE	$p <$	Est.	SE	$p <$	Est.	SE	$p <$
Intercept β_0	$-.999$	.112	.001	-1.015	.116	.001	-1.010	.117	.001
T β_1	$-.633$	.126	.001	$-.619$	.127	.001	$-.631$	.131	.001
T^2 β_2	.132	.029	.001	.132	.029	.001	.135	.030	.001
$H1$ β_3	.583	.170	.001	.765	.207	.001	.869	.226	.001
$H2$ β_4	.288	.121	.018	.334	.138	.012	.435	.151	.004
$H3$ β_5	.202	.119	.091	.269	.138	.051	.274	.149	.066
$Race$ β_6	.358	.200	.074	.353	.200	.078	.354	.200	.077
$H1 \times T$ β_7				$-.142$	.072	.048	$-.509$	.236	.031
$H2 \times T$ β_8				$-.035$	.051	.495	$-.389$	.187	.037
$H3 \times T$ β_9				$-.050$	.053	.346	$-.051$	.200	.800
$H1 \times T^2$ β_{10}							.087	.052	.096
$H2 \times T^2$ β_{11}							.086	.043	.044
$H3 \times T^2$ β_{12}							.000	.046	.995

Single- and Multi-Parameter Wald Tests

1. Single-parameter test, *e.g.*, $H_0 : \beta_1 = 0$

$$z = \hat{\beta}_1 / \hat{se}(\hat{\beta}_1) \quad \text{or} \quad X_1^2 = \hat{\beta}_1^{\,2} / \hat{V}(\hat{\beta}_1)$$

2. Linear combination of parameters, *e.g.*, $H_0 : \beta_1 + \beta_2 = 0$

   for this, suppose $\boldsymbol{\beta}' = [\hat{\beta}_0 \;\; \hat{\beta}_1 \;\; \hat{\beta}_2]$ and define $\boldsymbol{c} = [0 \;\; 1 \;\; 1]$

$$X_1^2 = \left(\boldsymbol{c}\hat{\boldsymbol{\beta}}\right)' \left[\boldsymbol{c}\, \hat{V}(\hat{\boldsymbol{\beta}})\, \boldsymbol{c}'\right]^{-1} \left(\boldsymbol{c}\hat{\boldsymbol{\beta}}\right)$$

   Notice, 1. $(H_0 : \beta_1 = 0)$ is a special case where $\boldsymbol{c} = [0 \;\; 1 \;\; 0]$

3. Multi-parameter test, *e.g.*, $H_0 : \beta_1 = \beta_2 = 0$

$$\boldsymbol{C} = \begin{bmatrix} 0 & 1 & 0 \\ 0 & 0 & 1 \end{bmatrix} \qquad X_2^2 = \left(\boldsymbol{C}\hat{\boldsymbol{\beta}}\right)' \left[\boldsymbol{C}\, \hat{V}(\hat{\boldsymbol{\beta}})\, \boldsymbol{C}'\right]^{-1} \left(\boldsymbol{C}\hat{\boldsymbol{\beta}}\right)$$

Comparing models 1 and 3, models with and without the group by time effects, the null hypothesis is

$$H_0 = \beta_7 = \beta_8 = \beta_9 = \beta_{10} = \beta_{11} = \beta_{12} = 0$$

$$\boldsymbol{C} = \begin{bmatrix} 0 & 0 & 0 & 0 & 0 & 0 & 0 & 1 & 0 & 0 & 0 & 0 & 0 \\ 0 & 0 & 0 & 0 & 0 & 0 & 0 & 0 & 1 & 0 & 0 & 0 & 0 \\ 0 & 0 & 0 & 0 & 0 & 0 & 0 & 0 & 0 & 1 & 0 & 0 & 0 \\ 0 & 0 & 0 & 0 & 0 & 0 & 0 & 0 & 0 & 0 & 1 & 0 & 0 \\ 0 & 0 & 0 & 0 & 0 & 0 & 0 & 0 & 0 & 0 & 0 & 1 & 0 \\ 0 & 0 & 0 & 0 & 0 & 0 & 0 & 0 & 0 & 0 & 0 & 0 & 1 \end{bmatrix}$$

- $X_6^2 = 10.98$, $p = .09$
- Also, several of the individual group by time parameter tests are significant
- observed abstinence rates indicate large post-intervention group differences that are not maintained over time

$\Rightarrow$ model 3 is preferred to model 1

Comparing models 2 and 3, models with and without the group by quadratic time effects, the null hypothesis is

$$H_0 = \beta_{10} = \beta_{11} = \beta_{12} = 0$$

$$\boldsymbol{C} = \begin{bmatrix} 0 & 0 & 0 & 0 & 0 & 0 & 0 & 0 & 0 & 0 & 1 & 0 & 0 \\ 0 & 0 & 0 & 0 & 0 & 0 & 0 & 0 & 0 & 0 & 0 & 1 & 0 \\ 0 & 0 & 0 & 0 & 0 & 0 & 0 & 0 & 0 & 0 & 0 & 0 & 1 \end{bmatrix}$$

- $X^2_3 = 5.91$, $p = .12$
- but, individual $H1 \times T^2$ interaction ($\hat{\beta}_{10} = .0870, p < .096$) and individual $H2 \times T^2$ interaction ($\hat{\beta}_{11} = .0855, p < .044$)
- some evidence for model 3, though, strictly speaking, not quite at the .05 level in terms of the multi-parameter Wald test

Interpretations based on Model 3

H1

- randomization to group increases abstinence at post-intervention ($\hat{\beta}_3 = .869, p < .001$)
- this benefit goes away across time ($\hat{\beta}_7 = -.509, p < .031$, $\hat{\beta}_{10} = .087, p < .096$)

Estimated odds ratio at post-intervention

$$OR = \exp[4/3(.869)] = 3.19$$

(multiply by 4/3 because this equals the difference between the control and treatment groups in the coding of the H1 contrast)

Asymptotic 95% confidence interval for this odds ratio

$$\exp[4/3(.869) \pm 1.96 \times 4/3(.226)] \;=\; (1.76, 5.75)$$

H2

- going to groups increases abstinence at post-intervention ($\hat{\beta}_4 = .435, p < .004$)
- this benefit goes away across time ($\hat{\beta}_8 = -.389, p < .037$, $\hat{\beta}_{11} = .086, p < .044$)

Estimated odds ratio at post-intervention

$$OR = \exp[3/2(.435)] = 1.92$$

(multiply by 3/2 because this equals the difference between those not attending and those attending groups in the coding of the H2 contrast)

Asymptotic 95% confidence interval for this odds ratio

$$\exp[3/2(.435) \pm 1.96 \times 3/2(.151)] \ = \ (1.23, 2.99)$$

H3

- marginally significant benefit of enhanced groups at post-intervention ($\hat{\beta}_5 = .274, p < .066$)
- this does not significantly vary across time ($\hat{\beta}_9 = -.051, p < .80,\ \ \hat{\beta}_{12} = .0003, p < .95$)

Estimated odds ratio at post-intervention

$$OR = \exp[2(.274)] = 1.73$$

(multiply by 2 because this equals the difference between the enhanced and regular groups in the coding of the H3 contrast)

Asymptotic 95% confidence interval for this odds ratio

$$\exp[2(.274) \pm 1.96 \times 2(.149)] \ = \ (.96, 3.10)$$

Determination of group difference at any timepoint

Model 3

$$\begin{aligned}\eta_{ij} = \; & \beta_0 + \beta_1 T_j + \beta_2 T_j^2 + \beta_3 H1_i + \beta_4 H2_i + \beta_5 H3_i + \beta_6 Race_i \\ & + \beta_7 (H1_i \times T_j) + \beta_8 (H2_i \times T_j) + \beta_9 (H3_i \times T_j) \\ & + \beta_{10} (H1_i \times T_j^2) + \beta_{11} (H2_i \times T_j^2) + \beta_{12} (H3_i \times T_j^2)\end{aligned}$$

$$\hat{H}1 = \hat{\beta}_3 + (T \times \hat{\beta}_7) + (T^2 \times \hat{\beta}_{10})$$

e.g., $T = 4$,

$$\hat{H}1 = .869 + (4 \times -.509) + (16 \times .087) = .227$$

is this a signficant difference?

$$H_0 : \beta_3 + (4 \times \beta_7) + (16 \times \beta_{10}) \; = \; 0$$

$\Rightarrow$ Wald test for linear combination of parameters

$$\boldsymbol{c} = \left[\begin{array}{cccccccccccc} 0 & 0 & 0 & 1 & 0 & 0 & 0 & 4 & 0 & 0 & 16 & 0 & 0 \end{array} \right]$$

$X_1^2 = .90$ for this $H1$ contrast at the final timepoint

Similarly, $X_1^2 = 1.79$ and .17, respectively for $H2$ and $H3$ contrasts at last timepoint

$\Rightarrow$ No significant group differences by the end of the study

Model 3 - Estimated Abstinence Rates

Group	End-of-Program (T1)	6 months (T2)	12 months (T3)	24 months (T4)
No Contact Control	.146	.137	.140	.186
No Shows	.263	.204	.176	.194
Discussion	.319	.184	.140	.227
Social Support	.456	.266	.192	.260

obtained as group by time averages of $\hat{p}_{ij} = \frac{1}{1+\exp(-\hat{\eta}_{ij})}$ where

$$\begin{aligned}\hat{\eta}_{ij} &= \hat{\beta}_0 + \hat{\beta}_1 T_j + \hat{\beta}_2 T_j^2 + \hat{\beta}_3 H1_i + \hat{\beta}_4 H2_i + \hat{\beta}_5 H3_i + \hat{\beta}_6 Race_i \\ &\quad + \hat{\beta}_7 (H1_i \times T_j) + \hat{\beta}_8 (H2_i \times T_j) + \hat{\beta}_9 (H3_i \times T_j) \\ &\quad + \hat{\beta}_{10} (H1_i \times T_j^2) + \hat{\beta}_{11} (H2_i \times T_j^2) + \hat{\beta}_{12} (H3_i \times T_j^2)\end{aligned}$$

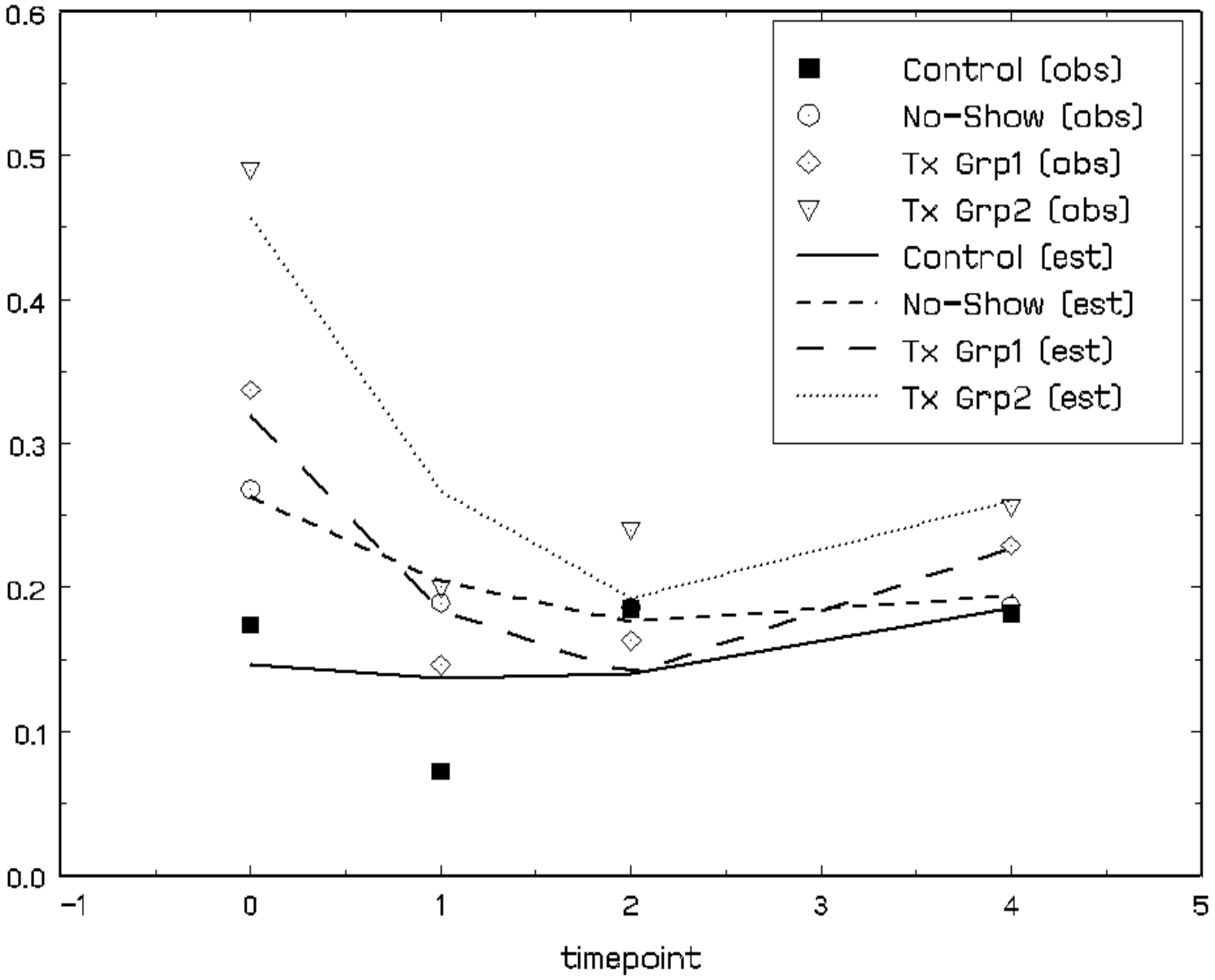


Figure 8.1 Observed point prevalence abstinence rates and estimated probabilities of abstinence across time

Example 8.2: PROC GENMOD code and output from analysis of Robin Mermelstein's smoking cessation study dataset. This handout illustrates GEE modeling of a dichotomous outcome. Includes CONTRAST statements to perform linear combination and multi-parameter Wald tests, and OBSTATS to yield estimated probabilities for each observation (SAS code and output)

http://www.uic.edu/classes/bstt/bstt513/robingeb_Ctime.txt

Another GEE example and comparisons with MRM

chapters 8 and 9

Consider the Reisby data and the question of drug plasma levels and clinical response to depression; define

$Response = 0$ (HDRS > 15) or 1 (HDRS ≤ 15)

$DMI = 0$ (ln dmi below median) or 1 (ln dmi above median)

	Response	
DMI	0	1
0	73	52
1	43	82

$\Rightarrow$ OR $= 2.68$

Reisby data - analysis of dichotomized HDRS

1. Logistic regression (inappropriate model; for comparison)

$$\log\left[\frac{P(Resp_{ij}=1)}{1-P(Resp_{ij}=1)}\right]=\beta_0+\beta_1 DMI_{ij}$$

2. GEE logistic regression with exchangeable structure

$$\log\left[\frac{P(Resp_{ij}=1)}{1-P(Resp_{ij}=1)}\right]=\beta_0+\beta_1 DMI_{ij}$$

3. Random-intercepts logistic regression

$$\log\left[\frac{P(Resp_{ij}=1)}{1-P(Resp_{ij}=1)}\right]=\beta_0+\beta_1 DMI_{ij}+\sigma_\upsilon\theta_i$$

$i=1,\ldots,66$ subjects; $j=1,\ldots,n_i$ observations per subject (max $n_i=4$)

Logistic Regression of dichotomized HDRS - ML ests (std errors)

model term	ordinary LR	GEE exchange	Random Int
intercept β_0	-.339	-.397	-.661
	(.182)	(.231)	(.407)
$\exp(\beta_0)$	.712	.672	.516
$DMI\beta_1$	.985	1.092	1.842
	(.262)	(.319)	(.508)
$\exp(\beta_1)$	2.68	2.98	6.31
subject sd σ_υ			2.004
			(.415)
ICC			.55
$2 \log L$	330.66		293.85

Marginal Models for Longitudinal Data

- Regression of response on $\boldsymbol{x}$ is modeled separately from within-subject correlation
- Model the marginal expectation: $E(y_{ij}) = fn(\boldsymbol{x})$
- Marginal expectation = average response over the sub-population that shares a commone value of $\boldsymbol{x}$
- Marginal expectation is what is modeled in a cross-sectional study

Assumptions of Marginal Model for Longitudinal Data

1. Marginal expectation of the response $E(y_{ij}) = \mu_{ij}$ depends on $\boldsymbol{x}_{ij}$ through link function $g(\mu_{ij})$

   e.g., logit link for binary responses

2. Marginal variance depends on marginal mean:
   $V(y_{ij}) = V(\mu_{ij})\phi$, with V as a known variance function (*e.g.*, $\mu_{ij}(1-\mu_{ij})$ for binary) and ϕ is a scale parameter

3. Correlation between y_{ij} and $y_{ij'}$ is a function of the marginal means and/or parameters $\boldsymbol{\alpha}$

$\Rightarrow$ Marginal regression coefficients have the same interpretation as coefficients from a cross-sectional analysis

Logistic GEE as marginal model - Reisby example

1. Marginal expectation specification: logit link

$$\log\left[\frac{\mu_{ij}}{1-\mu_{ij}}\right] = \log\left[\frac{P(Resp_{ij}=1)}{1-P(Resp_{ij}=1)}\right] = \beta_0 + \beta_1 DMI_{ij}$$

2. Variance specification for binary data: $V(y_{ij}) = \mu_{ij}(1-\mu_{ij})$ and $\phi = 1$ (in usual case)

3. Correlation between y_{ij} and $y_{ij'}$ is exchangeable, AR(1), m-dependent, UN

- $\exp \beta_0$ = ratio of the frequencies of response to non-response (*i.e.*, odds of response) among the sub-population (of observations) with below average DMI
- $\exp \beta_1$ = odds of response among above average DMI observations divided by the odds among below average DMI observations

$\exp \beta_1$ = ratio of population frequencies $\Rightarrow$
"population-averaged"

Random-intercepts logistic regression

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

or

$$g[\Pr(Y_{ij}=1 \mid \theta_i)] = \boldsymbol{x}'_{ij}\boldsymbol{\beta} + \sigma_\upsilon\theta_i$$

which yields

$$\Pr(Y_{ij}=1 \mid \theta_i) = g^{-1}[\boldsymbol{x}'_{ij}\boldsymbol{\beta} + \sigma_\upsilon\theta_i]$$

where g is the logit link function and g^{-1} is its inverse function (*i.e.*, logistic cdf)

Taking the expectation, $\quad E(Y_{ij} \mid \theta_i) = g^{-1}[\boldsymbol{x}'_{ij}\boldsymbol{\beta} + \sigma_\upsilon \theta_i]$

so $\quad \mu_{ij} = E(Y_{ij}) = E[E(Y_{ij} \mid \theta_i)] = \int_\theta g^{-1}[\boldsymbol{x}'_{ij}\boldsymbol{\beta} + \sigma_\upsilon \theta_i] f(\theta)\, d\theta$

When g is a nonlinear function, like logit, and if we assume that

$$g(\mu_{ij}) = \boldsymbol{x}'_{ij}\boldsymbol{\beta} + \sigma_\upsilon \theta_i$$

it is usually not true that $\quad g(\mu_{ij}) = \boldsymbol{x}'_{ij}\boldsymbol{\beta}$

unless $\theta_i = 0$ for all i subjects, or g is the identity link (*i.e.*, the normal regression model for y)

$\Rightarrow$ same reason why the log of the mean of a series of values does not, in general, equal the mean of the log of those values (*i.e.*, the log is a nonlinear function)

Random-intercepts Model - Reisby example

- every subject has their own propensity for response (θ_i)
- the effect of DMI is the same for every subject (β_1)
- covariance among the repeated obs is explicity modeled

- β_0 = log odds of response for a typical subject with $DMI = 0$ and $\theta_i = 0$
- β_1 = log odds ratio of response when a subject is high on DMI relative to when that same subject is not
  - On average, how a subject's resp prob depends on DMI
  - Strictly speaking, it's not really the "same subject," but "subjects with the same value of θ_i"
- σ_υ represents the degree of heterogeneity across subjects in the probability of response, not attributable to DMI

- Most useful when the objective is to make inference about subjects rather than the population average
- Interest in heterogeneity of subjects

Random-intercepts model with time-invariant covariate

$$\log\left[\frac{\Pr(Y_{ij}=1 \mid \theta_i)}{1-\Pr(Y_{ij}=1 \mid \theta_i)}\right] = \beta_0 + \beta_1 x_i + \sigma_\upsilon \theta_i$$

where, say, $x_i = 0$ for controls and $x_i = 1$ for treated patients

- β_0 = log odds of response for a control subject with $\theta_i = 0$
- β_1 = log odds ratio of response when a subject is "treated" relative to when that same subject (or more precisely, subjects with the same θ_i) is "control"

In some sense, interpretation of β_1 goes beyond the observed data

$\Rightarrow$ marginal interpretation is often preferred for time-invariant covariates

Interpretation of regression coefficients

mixed models $\boldsymbol{\beta}$ represent the effects of the explanatory variables on a subject's chance of response (subject-specific)

marginal models $\boldsymbol{\beta}$ represent the effects of the explanatory variables on the population average (population-averaged)

Odds Ratio

mixed models describes the ratio of a subject's odds

marginal models describes the ratio of the population odds

Neuhaus *et al.*, 1991

- if $\sigma_v^2 > 0 \;\; \Rightarrow \;\; |\beta_{ss}| > |\beta_{pa}|$
- discrepancy increases as σ_v^2 increases (unless, in trivial case, $\beta_{ss} = 0$, then $\beta_{pa} = 0$)

Marginal and Random-int LR in terms of latent y

Marginal Logistic Regression

$$y_{ij} = \boldsymbol{x}'_{ij}\boldsymbol{\beta}_{pa} + \varepsilon_{ij} \qquad \varepsilon_{ij} \sim L(0, \pi^2/3) \;\; \rightarrow \;\; V(y_{ij}) = \pi^2/3$$

Random-intercepts Logistic Regression

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

$$\upsilon_i \sim N(0, \sigma^2_\upsilon) \quad \varepsilon_{ij} \sim L(0, \pi^2/3) \quad \rightarrow \; V(y_{ij}) = \pi^2/3 + \sigma^2_\upsilon$$

$\Rightarrow$ suggests that to equate

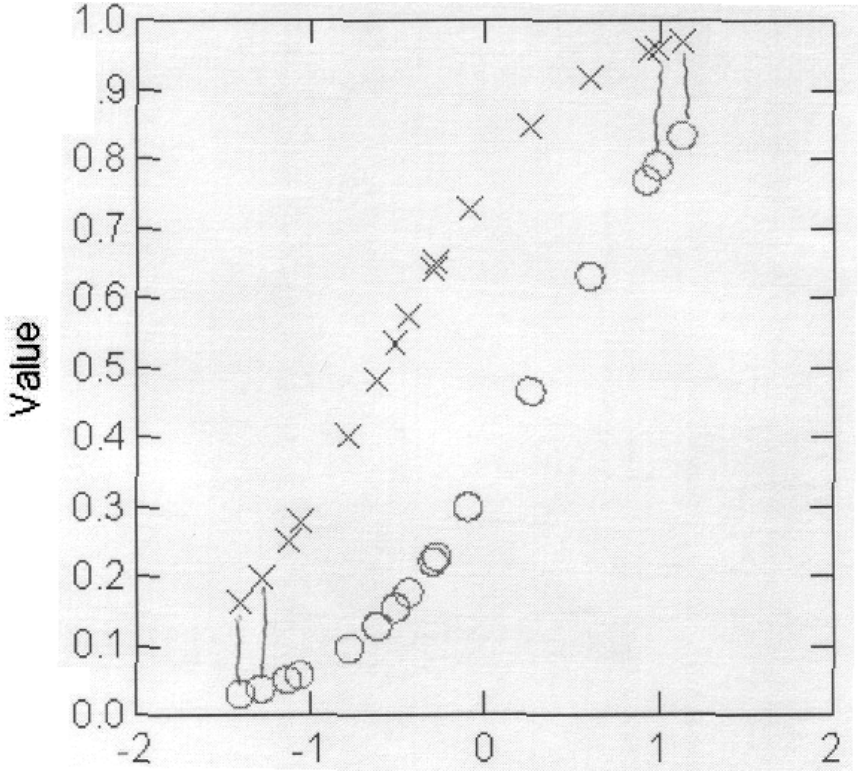
$$\boldsymbol{\beta}_{pa} \approx \boldsymbol{\beta}_{ss} / \sqrt{\frac{\pi^2/3 + \sigma^2_\upsilon}{\pi^2/3}} \;=\; \boldsymbol{\beta}_{ss} / \sqrt{\frac{3}{\pi^2}\sigma^2_\upsilon + 1}$$

Zeger *et al.*, 1988 suggests a slightly larger denominator

$$\boldsymbol{\beta}_{pa} \approx \boldsymbol{\beta}_{ss} / \sqrt{\left(\frac{16}{15}\right)^2 \frac{3}{\pi^2}\sigma^2_\upsilon + 1}$$

HDRS response probability by DMI median cut

subjects with varying DMI values over time ($N = 20$)



$\circ$ DMI=0 $\times$ DMI=1

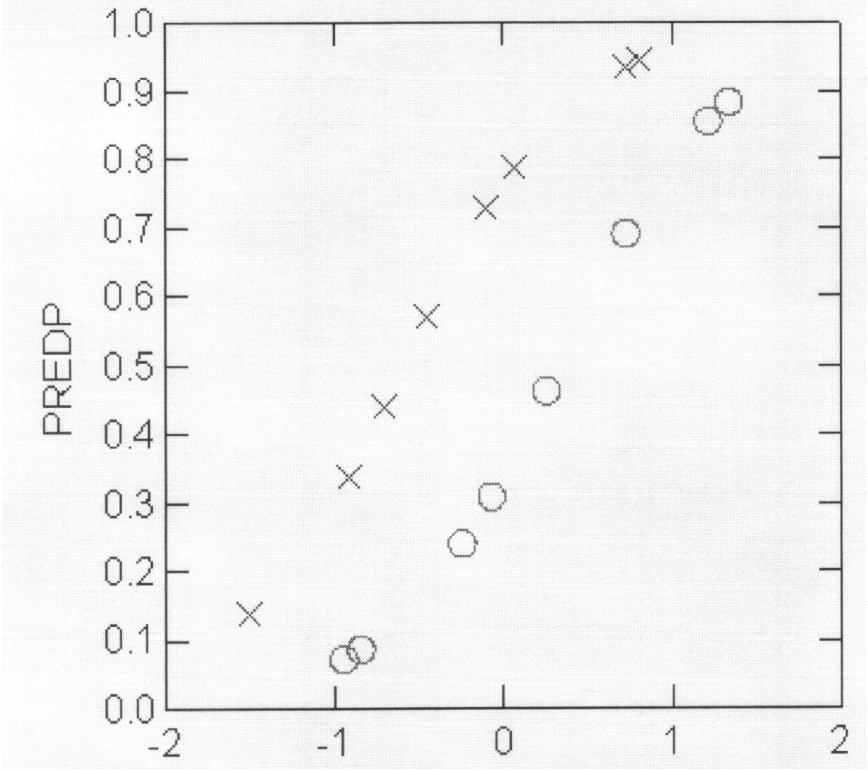
$$P(Resp_{ij} = 1|\theta_i) = 1/[1 + \exp(-(-.66 + 1.84DMI_{ij} + 2.00\theta_i))]$$

From GEE: $\hat{P}_{DMI=0} = .40$ $\hat{P}_{DMI=1} = .67$

HDRS response probability by DMI median cut

subjects with consistent DMI values over time

(low DMI $N = 24$; high DMI $N = 22$)



θ $\quad\circ$ DMI=0 $\quad\times$ DMI=1

$$P(Resp_{ij} = 1|\theta_i) = 1/[1 + \exp(-(-.66 + 1.84DMI_{ij} + 2.00\theta_i))]$$

From GEE: $\hat{P}_{DMI=0} = .40 \qquad \hat{P}_{DMI=1} = .67$

Example 8.3: PROC IML code and output showing how to get the marginalized probability estimates from GEE and NLMIXED analysis for a random-intercepts model, including using quadrature for the latter (SAS code and output)

http://www.uic.edu/classes/bstt/bstt513/ReisGEEfit.txt