

*Measurement, Design, and Analytic Techniques in Mental  
Health and Behavioral Sciences*

*Lecture 8 (Feb 6, 2007): Hierarchical  
(multilevel) modeling for clustered data, I*

XH Andrew Zhou

azhou@u.washington.edu

Professor, Department of Biostatistics, University of Washington

## Three models for longitudinal data

---

### Four models for longitudinal data:

- Marginal models: the relationship of the expected response on covariates is modeled separately from within-subject correlation.
- Random effects models
- Fixed effects models
- Transitional models

## Notation

---

- $Y_{ij}$ : the value of the dependent variable observed at the time  $t_{ij}$  for observations  $j = 1, \dots, n_i$  on subject  $i$ .
- $x_{ij}$ : the vector of  $p$  covariates observed at time  $t_{ij}$  for observations  $j = 1, \dots, n_i$  on subject  $i$ .
- $\mu_{ij} = E(Y_{ij})$  and  $Var(Y_{ij}) = \nu_{ij}$
- Repeated outcomes for subject  $i$  are put into an  $n_i$  vector,  $\mathbf{Y}_i = (Y_{i1}, \dots, Y_{in_i})$  with mean  $E(\mathbf{Y}_i) = \boldsymbol{\mu}_i$  and  $n_i \times n_i$  covariance matrix  $Var(\mathbf{Y}_i) = \mathbf{V}_i$ .

# Marginal models

- We model the marginal expectation,  $E(Y_{ij} | X_{ij} = x)$ , the average response over the sub-population that shares a common value of  $x$ . A cross-sectional study models the marginal expectation.
- $E(Y_{ij} | X_{ij}) = \mu_{ij}$  depends on covariates  $x_{ij}$  by  $h(\mu_{ij}) = x'_{ij}\beta$ , where  $h$  is a known link function.
- The marginal variance depends on the marginal mean:

$$\text{Var}(Y_{ij}) = \nu(\mu_{ij})\phi,$$

where  $\nu$  is a known variance function, and  $\phi$  is a possibly unknown scale parameter.

- The correlation between  $Y_{ij}$  and  $Y_{ik}$  is a function of the marginal means and possibly additional parameters  $\alpha$ :

$$\text{Corr}(Y_{ij}, Y_{ik}) = \rho(\mu_{ij}, \mu_{ik}, \alpha),$$

where  $\rho(\cdot)$  is a known function.

## Interpretation of a marginal model

---

- Marginal regression coefficients,  $\beta$ , have the same interpretation as the ones from cross-sectional analysis.
- Marginal models are natural analogues for correlated data of GLM for independent data.

## Estimation method - GEE

---

- Assume a working correlation matrix of  $Y_i$ ,  $R_i(\alpha)$ . Then, the covariance matrix of  $Y_i$  is  $V_i = B_i^{1/2} R_i(\alpha) B_i^{1/2} \phi$ , where  $B_i = \text{diag}(\nu(\mu_{11}, \dots, \mu_{in_i}))$ .
- Estimation of  $\beta$  and  $\alpha$  can be obtained by the following generalized estimation equation (GEE):

$$S(\beta, \alpha, \phi) = \sum_{i=1}^n \frac{\partial \mu_i'}{\beta} V_i^{-1} (Y_i - \mu_i) = 0.$$

## An iterative algorithm

---

- Obtain an estimate of  $\beta$  by solving the above GEE, given  $\alpha$  and  $\phi$ .
- Use the following Pearson residuals  $r_{ij}^P$  to  $\alpha$  and  $\phi$ , given  $\beta$ :

$$r_{ij}^P = \frac{y_{ij} - \mu_{ij}(\beta)}{\sqrt{\nu(\mu_{ij}(\beta))}}$$

- A moment estimator of the overdispersion parameter,  $\phi$ :

$$\hat{\phi} = \frac{1}{n} \sum_{i=1}^n \frac{1}{n_i} (r_{ij}^P)^2.$$

## AN iterative algorithm, cont

The following moment estimator of  $\alpha$  for the different working correlation structure:

- Independence: correlation structure,  $Corr(Y_{ij}, Y_{ij'}) = 0, j \neq j'$ .
- Exchangeable: correlation structure,  $Corr(Y_{ij}, Y_{ij'}) = \alpha, j \neq j'$ ; estimator:

$$\hat{\alpha} = \frac{1}{n} \frac{1}{n_i(n_i - 1)} \sum_{j \neq j'} r_{ij}^P r_{ij'}^P.$$

- AR(1): correlation structure,

$$Corr(Y_{ij}, Y_{i(j+t)}) = \alpha^t, t = 0, 1, \dots, n_i - j;$$

estimator,

$$\hat{\alpha} = \frac{1}{n} \sum_{i=1}^n \frac{1}{n_i - 1} \sum_{j=1}^{n_i-1} r_{ij}^P r_{i(j+1)}^P.$$

- Unstructured: correlation structure:

$$Corr(Y_{ij}, Y_{ij'}) = \alpha_{jj'}, j \neq j';$$

estimator:



## Asymptotic Properties

---

- The estimator,  $\hat{\beta}$ , the marginal regression parameter vector,  $\beta$ , is asymptotically normal and consistent as the number of clustered increases to  $\infty$ .
- $\hat{\beta}$  is consistent of  $\beta$ , assuming that the mean structure is correctly specified, even if the correlation structure is misspecified. A consistent covariance matrix estimator can be obtained by means of the sandwich estimator

# Robust sandwich covariance estimators

---

- Assuming  $\alpha$  and  $\phi$  are known, we have the following  $p$  equations:

$$S(\hat{\beta})_{\alpha, \phi} = \sum_{i=1}^n S_i(\hat{\beta}) = 0,$$

where

$$S_i(\beta) = \frac{\partial}{\partial \mu_i} \beta' V_i^{-1} (Y_i - \mu_i).$$

- Then, we can write the covariance matrix of  $S(\hat{\beta})_{\alpha, \phi}$  as follows:

$$\text{Cov}(S(\hat{\beta}_{\alpha, \phi})) = \frac{\partial S(\beta)_{\alpha, \phi}}{\partial \beta} \text{Cov}(\hat{\beta}) \left( \frac{\partial S(\beta)_{\alpha, \phi}}{\partial \beta} \right)'$$

## Robust sandwich covariance estimators, cont

- Therefore, we obtain the following expression for  $Cov(\hat{\beta})$ :

$$Cov(\hat{\beta}) = \left( \frac{\partial S(\beta)_{\alpha, \phi}}{\partial \beta} \right)^{-1} Cov(S(\hat{\beta}_{\alpha, \phi})) \left( \left\{ \frac{\partial S(\beta)_{\alpha, \phi}}{\partial \beta} \right\}' \right)^{-1}.$$

- Since  $S(\hat{\beta})_{\alpha, \phi}$  is a sum of independent vectors with mean 0, we can estimate its covariance matrix by its empirical one:

$$\widehat{Cov}(S(\hat{\beta})_{\alpha, \phi}) = \frac{1}{n} \sum_{i=1}^n S_i(\hat{\beta}) S_i(\hat{\beta})'.$$

## Example on toenail infection treatment

---

- De Backer (1998) et al. reported a study on the relative effectiveness of two treatments on toenail infection.
- 378 patients were randomly assigned to one of the two treatments and evaluated at seven visits, at weeks 0, 4, 8, 12, 24, 36, and 48.
- The outcome variable is a binary variable, indicating degree of separation of the nail plate from the nail bed (0: none or mild; 1: moderate or severe).

## Example on toenail infection treatment, cont

---

- $Y_{ij}$ : outcome of subject  $j$  at visit  $i$
- $X_{2j}$ : treatment group of subject  $j$
- $X_{3ij}$ : exact timing (in month) of visit  $i$  of subject  $j$

## STATA codes

---

- `do C:\Zhou\Teaching\2007\examples-mlmus\ch4_toenail.do`
- `use toenail, clear`
- A useful command for longitudinal data is `xtides`, which describes the participation pattern in the data set.

## STATA codes

```

patient: 1, 2, ..., 383      n = 294
visit: 1, 2, ..., 7        T =          7
      Delta(visit) = 1; (7-1)+1 = 7
(patient*visit uniquely identifies each observation)
Distribution of T_i: min 5% 25% 50% 75% 95% max
                    1   3   7   7   7   7   7

```

Freq.	Percent	Cum.	Pattern
224	76.19	76.19	1111111
21	7.14	83.33	11111.1
10	3.40	86.73	1111.11
6	2.04	88.78	111....
5	1.70	90.48	1.....
5	1.70	92.18	11111..
4	1.36	93.54	1111...
3	1.02	94.56	11.....
3	1.02	95.58	111.111
13	4.42	100.00	(other patterns)
294	100.00		XXXXXXXX

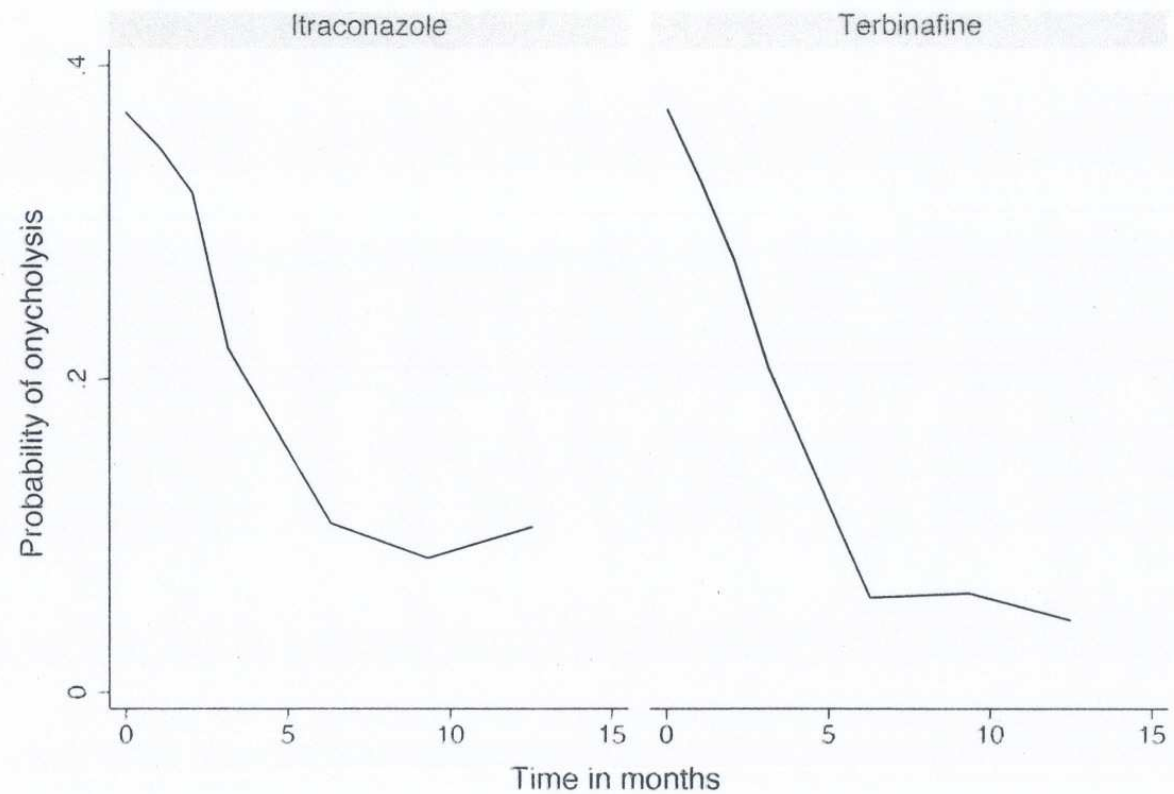
The data set is not balanced since all patients did not attend all

## Graphical display

- A useful graphical display of the data is a line graph, plotting the observed proportions at each visit against time, the average time associated with each visit.



# Graphical display



Graphs by treatment

## Estimation using Generalized Estimation Equation (GEE)

---

```
cd c:\zhou\teaching\2007\examples-mlmus capture log close log using  
toenail, replace text set more off set scheme sj
```

```
use toenail, clear gen trt_month = treatment*month xtgee outcome  
treatment month trt_month, i(patient) link(logit) family(binom)  
robust eform
```

## GEE results

```
GEE population-averaged model      Number of obs      = 1908
Group variable:      patient      Number of groups   = 294
Link: logit Obs per group: min = 1 Family: binomial avg = 6.5
Correlation: exchangeable          max = 7
Wald chi2(3)
= 63.44 Scale parameter:      1      Prob > chi2 = 0.000
(Std. Err. adjusted for clustering on patient)
```

---

outcome	Odds Ratio	Semi-robust Std. Err.	z	P> z	[95% Conf. Interva
treatment	1.007207	.2618022	0.03	0.978	.6051549 1.67
month	.8425856	.0253208	-5.70	0.000	.7943911 .89
trt_month	.9252113	.0501514	-1.43	0.152	.8319576 1.02

---

## GEE, cont

---

Working correlations in STATA:

- Independence
- Exchangeable (the same correlation for all unit - the default in STATA)
- AR(1) (autoregressive lag 1, which only makes sense for longitudinal data)
- Unrestricted (a different correlation for each pair of responses - it should not be used for large clusters)

## Random effects models

---

- There exists a vector of random effects,  $U_i$ . Given  $U_i$ , the responses  $Y_{i1}, \dots, Y_{in_i}$  are mutually independent and follow a GLM with density  $f(y_{ij} | U_i) = \exp\{[y_{ij}\theta_{ij} - \psi(\theta_{ij})]/\phi\}$ .
- The conditional mean,  $\mu_{ij} = E(Y_{ij} | U_i)$  and the conditional variance,  $\nu_{ij} = \text{Var}(Y_{ij} | U_i)$ :

$$h(\mu_{ij}) = x'_{ij}\beta^* + d'_{ij}U_i, \nu_{ij} = \nu(\mu_{ij})\phi,$$

where  $h$  and  $\nu$  are known link and variance functions, respectively, and  $d_{ij}$  is a subset of  $x_{ij}$ .

- Correlation between the random effects and covariates is zero.

## Interpretation of random effects models

---

- Basic idea: there is natural heterogeneity across individuals in their regression coefficients and this heterogeneity can be represented by a probability distribution.
- Correlation among observations for one person arises from their sharing unobservable variables,  $U_i$ .
- The random effects model is most useful when the aim is to make inference about individual rather the population average.

# Estimation

- With a random effects model, the marginal likelihood is obtained for inference.
- The marginal likelihood of subject  $i$  is obtained by integrating out random effects over their distribution.

$$P(y_{i1}, \dots, Y_{in_i} | x_i) = \int \prod_{j=1}^{n_i} P(y_{ij} | x_i, U_i) \Phi(u, 0, \Psi) du,$$

where  $\Phi(u, 0, \Psi)$  is the multivariate normal distribution with mean vector 0 and the covariance matrix  $\Psi$ .

- In general, the marginal likelihood does not have a closed form, due to intractable integration involved.
- There are three ways of approximating this intractable integrals, Laplace approximation, numerical integration using quadrature or adaptive quadrature, and Monte Carlo integration.

## Marginal and generalized quasi-likelihood

---

- Marginal quasi-likelihood (MQL) and penalized quasi-likelihood (PQL) are based on approximating generalized linear models by linear mixed models so that the Iterative Generalized Least Squares (IGLS) algorithm can be applied (hence no longer corresponds to maximum likelihood).
- In the MQL, the Taylor expansion of is done at zero, while in the PQL, the expansion is done at the posterior mode.



## Example 2

---

```
* random intercept model
gllamm outcome treatment month trt_month, i(patient) family(binom)
link(logit) nip(30) adapt gllamm, eform
```

# STATA Output

number of level 1 units = 1908 number of level 2 units = 294

Condition Number = 23.0763

gllamm model

log likelihood = -625.38558

---

outcome	Coef.	Std. Err.	z	P> z	[95% Conf. Interval]
treatment	-.1608751	.5802054	-0.28	0.782	-1.298057 .976306
month	-.3911055	.0443906	-8.81	0.000	-.4781096 -.30410
trt_month	-.136829	.0680213	-2.01	0.044	-.2701484 -.00350
_cons	-1.620364	.4322408	-3.75	0.000	-2.46754 -.773187

---

Variances and covariances of random effects

\*\*\*level 2 (patient)

var(1): 16.084107 (3.0626223)

---

## Interpretations

---

- There appears there is a significant interaction between treatment and time at the 5% level.

## Subject-specific versus population-averaged

- The population-averaged probabilities from the random-intercept model can be obtained by integrating out the random intercept:

$$\begin{aligned} P(Y_{ij} = 1 \mid x_{2i}, x_{3i}) &= \int P(Y_{ij} = 1 \mid x_{2i}, x_{3i}, \xi_i) \phi(\xi_i, 0, \psi) d\xi_i \\ &= \int \frac{\exp(\beta_1 + \beta_2 x_{2i} + \beta_3 x_{3i} + \xi_i)}{1 + \exp(\beta_1 + \beta_2 x_{2i} + \beta_3 x_{3i} + \xi_i)} \phi(\xi_i; 0, \psi) d\xi_i, \end{aligned}$$

$\psi(\xi_i; 0, \psi)$  is the normal density function with mean zero and variance  $\psi$ .

- These probabilities can be obtained using the *gllapred* command with the options *mu* (for the mean response) and *marginal* (for integrating over the random intercept distribution)
- `.gllapred margprob, mu marg`

## Population-averaged probabilities

- We compared predicted population-averaged from the ordinary logit and random-intercept logit models.

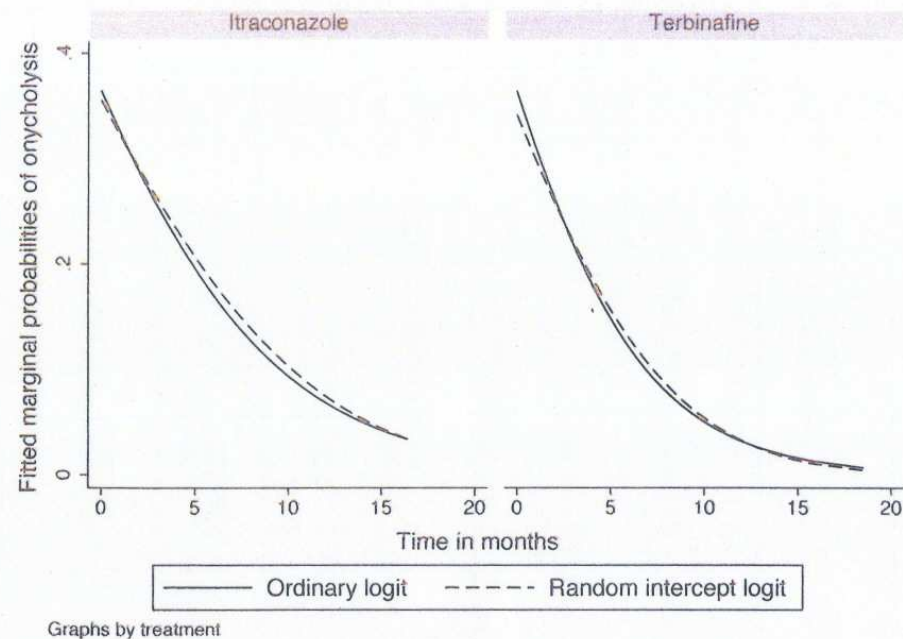


Figure 4.9: Fitted marginal probabilities using ordinary and random-intercept logistic regression

## Subject-specific prediction

---

- Subject-specific predictions for specific values of  $\xi$  can be produced using *gllapred* with the *mu* and *us(varname)*
- \* conditional, subject-specific probabilities  

```
gen zeta1=0 gllapred condprob0, mu us(zeta) gen lower1 = -4 gllapred  
condprobm4, mu us(lower) gen upper1 = 4 gllapred condprob4, mu  
us(upper) replace lower1 = -2 gllapred condprobm2, mu us(lower)  
replace upper1 = 2 gllapred condprob2, mu us(upper)
```

## Subject-specific prediction, cont

---

```
twoway (line prop mn_month, sort) ///
      (line margprob month, sort clpatt(dash)) ///
      (line condprob0 month, sort clpatt(dot) clwidth(medthick)) ///
      (line condprob4 month, sort clpatt(dot) clwidth(medthick)) ///
      (line condprobm4 month, sort clpatt(dot) clwidth(medthick)) ///
      (line condprob2 month, sort clpatt(dot) clwidth(medthick)) ///
      (line condprobm2 month, sort clpatt(dot)clwidth(medthick)), ///
by(treatment) ///
legend(order(1 "Observed proportions" 2 "Marginal probabities" 3 "Cond
xtitle(Time in months) ytitle(Probabilities of onycholysis)
```

## Subject-specific predictions

- The following are conditional and marginal probabilities for the random-intercept logistic regression model

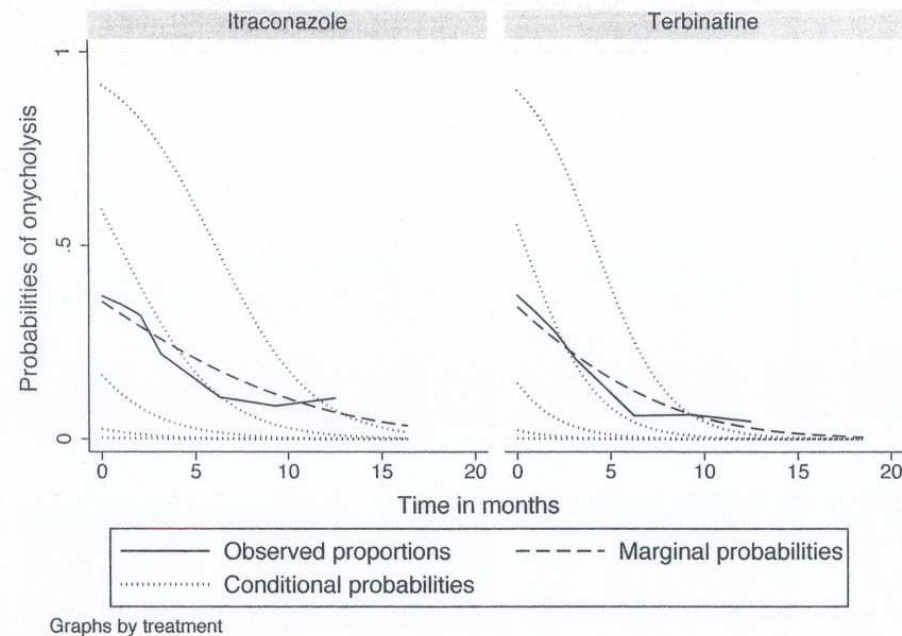


Figure 4.10: Conditional and marginal probabilities for the random-intercept logistic regression model



## Transition (Markov) models

---

- A transition model models correlation among  $Y_{i1}, \dots, Y_{in_i}$  by allowing the past values,  $Y_{i1}, \dots, Y_{i(j-1)}$ , explicitly influence the present observation,  $Y_{ij}$ .
- For a binary outcome, for example:

$$\text{logit}P(Y_{ij} = 1 \mid Y_{i(j-1)}, \dots, Y_{i1}) = x'_{ij}\beta^{**} + \alpha_{i(j-1)}.$$

# Contrasting approaches

---

- Dependent variable:  $Y_{ij}$ : whether the child has respiratory infection (1=yes, 0=no).
- Covariate:  $x_{ij}$ : whether child  $i$  is vitamin A deficient (1=yes, 0=no) at visit  $j$  of child  $i$ .
- Marginal model:

$$\text{logit}(P(Y_{ij} = 1)) = \log \frac{P(Y_{ij} = 1)}{P(Y_{ij} = 0)} = \beta_0 + \beta_1 x_{ij}.$$

- Random effects model:

$$\text{logit}(P(Y_{ij} = 1 | U_i)) = \beta_0^* + \beta_1^* x_{ij} + U_i.$$

- Transition Markov model:

$$\text{logit}(P(Y_{ij} = 1 | Y_{i(j-1)}, \dots, Y_{i1})) = \beta_0^{**} + \beta_1^{**} x_{ij} + \alpha Y_{i(j-1)}.$$

# Interpretation

- Marginal model:
  - $\exp(\beta_0)$ : the ratio of the frequency of infected to uninfected children among the sub-population that is not vitamin A deficient.
  - $\exp(\beta_1)$ : the odds of infection among vitamin A deficient children divided by the odds of infection among children with vitamin A.
- Random effects model:
  - $\beta_0^*$ : log-odds for respiratory infection for a typical child with random effect  $U_i = 0$ .
  - $\beta_1^*$ : log-odds ratio for infection when a child is deficient relative to when that same child is not.
- Transition model:
  - $\beta_1^{**}$ : log-odds ratio for infection among children who are free of infection at the previous visit.
  - $\exp(\alpha)$ : the ratio of the odds of infection among children who did and did not have infection at the prior visit.