Truncated Longitudinal Outcomes with Nonignorable Missing Data

Sujuan Gao
Indiana University School of Medicine
ADC Data Core Meeting
October 6, 2007
Cognitive Decline

- Related to AD pathology
- Characterize the disease process as a continuum in the cognitive spectrum with a gradual onset
- More feasible to measure a large study, resulting in a greater number of participants and greater power
Statistical Issues in Studying Cognitive Decline

- Reliability --- what is the real change in cognitive function
- Ceiling or floor effect
- Missing data

References:
A mixture of two populations --- a cross-sectional look

normal variation

disease pathology
A longitudinal look --- what contribute to decline?
Illustration with simple linear regression

\[ y_{ij} = 20 - x_{ij} + e_{ij} \]
With truncation

\[ y_{ij}^* = 16.36 - 0.76x_{ij} + e_{ij} \]
Attenuation toward the null
Truncated Outcomes in Linear Regressions

- Ordinary least square estimator is biased and inconsistent (Goldberger, 1981, J. Econometrics)
- Methods of correction
  --- EM algorithm with multiple truncation points (Tsui et al, 1988, JASA, Haselblad et al, 1980, JASA)
- Tobin Model
  --- a single truncation point with normal data (Amemiya, 1984, J Econometrics).
Longitudinal Settings

Hughes, 1999, Biometrics
- General longitudinal model for normal distribution
- EM based maximum likelihood approach

Specific models (for viral load in HIV and laboratory data):
- Marta et al, 2000
- Lyles et al 2000, *JRSS (C)*
- Moulton 2002, *Statistical Methods in Medical Research*
- Wu 2002, *JASA*
Truncated Longitudinal Outcome with Missing Data

Notations:

$y_{ij}$: the $j$th measurement from the $i$th subject.

$y_i$: the vector containing measurement from the $i$th subject.

$X_{ij}$ and $W_{ij}$ : vectors of fixed effect covariates.

$Z_{ij}$: vector of random effect covariates.

$(T_{ij}, Q_{ij})$ observed:

$$Q_{ij} = \begin{cases} 
  y_{ij}, & \text{if } T_{ij} = 0 \\
  \leq y_{ij}, & \text{if } T_{ij} = 1 
\end{cases}$$
I. Model for Longitudinal Outcomes

\[ y_{ij} = X_{ij} \beta + Z_{ij} \gamma_i + e_{ij} \]

- Fixed Effect parameter
- Random Effect
  \[ \gamma_i \sim N(0, D) \]
- \[ e_{ij} \sim N(0, \sigma^2 I) \]

\[ y \sim N(X\beta, V) \text{ where } V = Z'DZ + \sigma^2 I \]
2. The Drop Out Model

Let $R_{ij}$ be an indicator variable for missing

Let $p_{ij} = \text{Prob}(R_{ij} = 1)$

We assume the following model for the missing data mechanism:

$$\eta[\text{Prob}(R_{ij} = 1 | R_{i, j-1} = 0)] = W_{ij} \alpha + \delta U_{ij} \gamma_i$$
Maximum Likelihood Approach

\[
L = \prod_{i=1}^{n} \int_{-\infty}^{\infty} \left( \prod_{j=1}^{m_i} (1 - p_{ij}) f(y_{ij} \mid X_{ij}, Z_{ij}, \gamma_i) \right) \left(1 - p_{ij} \right) \left[ \int_{T}^{\infty} f(y_{ij} \mid X_{ij}, Z_{ij}, \gamma_i) dy_{ij} \right] \prod_{R_{ij}=1}^{p_{in_{i+1}}} \gamma_i \right) d \gamma_i
\]

L_1

L_2

L_3
Specifying the models

- Logit link function for the drop out model
- Recursive relationship for the marginal probability $R$

$$p_{ij} = \eta (1 - p_{ij-1}) + p_{ij-1}, \ j = 2, \Lambda , n_i, \ p_{i1=0}$$

- Numerical integration techniques in SAS proc nlmixed
The Indianapolis-Ibadan Dementia Project

- A longitudinal study of dementia/AD and cognitive decline.
- Study participants (one of two cohorts)
  - African American residents of Indianapolis (age $\geq 65$ at baseline).
- Baseline cognitive evaluation.
- Follow-up evaluations at 2, 5, 8 and 11 years after baseline.
- Identify risk factors for cognitive decline.
The Indianapolis Ibadan Dementia Project
Cognitive Measures

- **CSID**: community screening instrument for dementia.
- Measuring multiple domains including memory, executive function, language, spatial orientation.
- A summary score:
  - sum of correct answers to 33 questions
  - range 0-40.
The Indianapolis Ibadan Dementia Project

• Baseline cognitive scores:
  20.4% at ceiling.

Difference in mean scores

<table>
<thead>
<tr>
<th></th>
<th>Low Education</th>
<th>High Education</th>
<th>P-value</th>
</tr>
</thead>
<tbody>
<tr>
<td>Cognitive score (SD)</td>
<td>35.0 (3.7)</td>
<td>37.1 (3.0)</td>
<td>&lt;0.0001</td>
</tr>
<tr>
<td>% at ceiling</td>
<td>5.6</td>
<td>22.8</td>
<td>&lt;0.0001</td>
</tr>
</tbody>
</table>
The Indianapolis Ibadan Dementia Project

Follow up status

<table>
<thead>
<tr>
<th>Year 2</th>
<th>Year 5</th>
<th>Year 8</th>
<th>Year 11</th>
<th># subjects</th>
</tr>
</thead>
<tbody>
<tr>
<td>Obs.</td>
<td>Obs.</td>
<td>Obs.</td>
<td>Obs.</td>
<td>425</td>
</tr>
<tr>
<td>Obs.</td>
<td>Obs.</td>
<td>Obs.</td>
<td>*</td>
<td>267</td>
</tr>
<tr>
<td>Obs.</td>
<td>Obs.</td>
<td>*</td>
<td>*</td>
<td>497</td>
</tr>
<tr>
<td>Obs.</td>
<td>*</td>
<td>*</td>
<td>*</td>
<td>469</td>
</tr>
<tr>
<td>*</td>
<td>*</td>
<td>*</td>
<td>*</td>
<td>370</td>
</tr>
</tbody>
</table>

Obs: cognitive scores observed
*
*: cognitive scores missing
The Indianapolis Ibadan Dementia Project
Education and Missing Data

<table>
<thead>
<tr>
<th>% missing</th>
<th>Low education</th>
<th>High education</th>
</tr>
</thead>
<tbody>
<tr>
<td>Year 2</td>
<td>22.5</td>
<td>17.6</td>
</tr>
<tr>
<td>Year 5</td>
<td>34.4</td>
<td>27.4</td>
</tr>
<tr>
<td>Year 8</td>
<td>61.4</td>
<td>39.1</td>
</tr>
<tr>
<td>Year 11</td>
<td>64.3</td>
<td>36.3</td>
</tr>
</tbody>
</table>
The Indianapolis Ibadan Dementia Project

• Education seems to be related to:
  – The truncation process
  – The missing data mechanism

Our interest:
  Is education related to cognitive decline?
Simulations

- Based on data structure from the IIDP.
- Predictors for longitudinal scores:
  - age group (75+ vs 75-)
  - Education (yr education <=6)
  - time of follow-up
  - interaction between education and time
  - Subject specific random effect

- Cumulative incidence model for drop out
  - age group (75+ vs 75-)
  - education (yr education <=6)
  - time of follow-up
  - Subject specific random effect (in the opposite direction)
Discussion

Assumptions non-verifiable
  – multivariate normal for the longitudinal outcomes
  – model for the missing data mechanism and the dependence on the random effects

Sensitivity analyses needed
  – Sensitivity to the normal distribution assumption
  – Sensitivity to the missing data model.

Extensions:
  – non-normal distributions
  – intermittent missing data structure
Discussion

Practical Issues

– Whether instruments designed for dementia screening are ideal for identifying factors related to cognitive health?
– With severe ceiling effect, it would be difficult to define cognitive health.
Acknowledgement

• Indiana Alzheimer’s Disease Center

• Indianapolis-Ibadan Dementia Project

• Collaborative work with Rodelphe Thiebaut (Bordeaux, France)