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

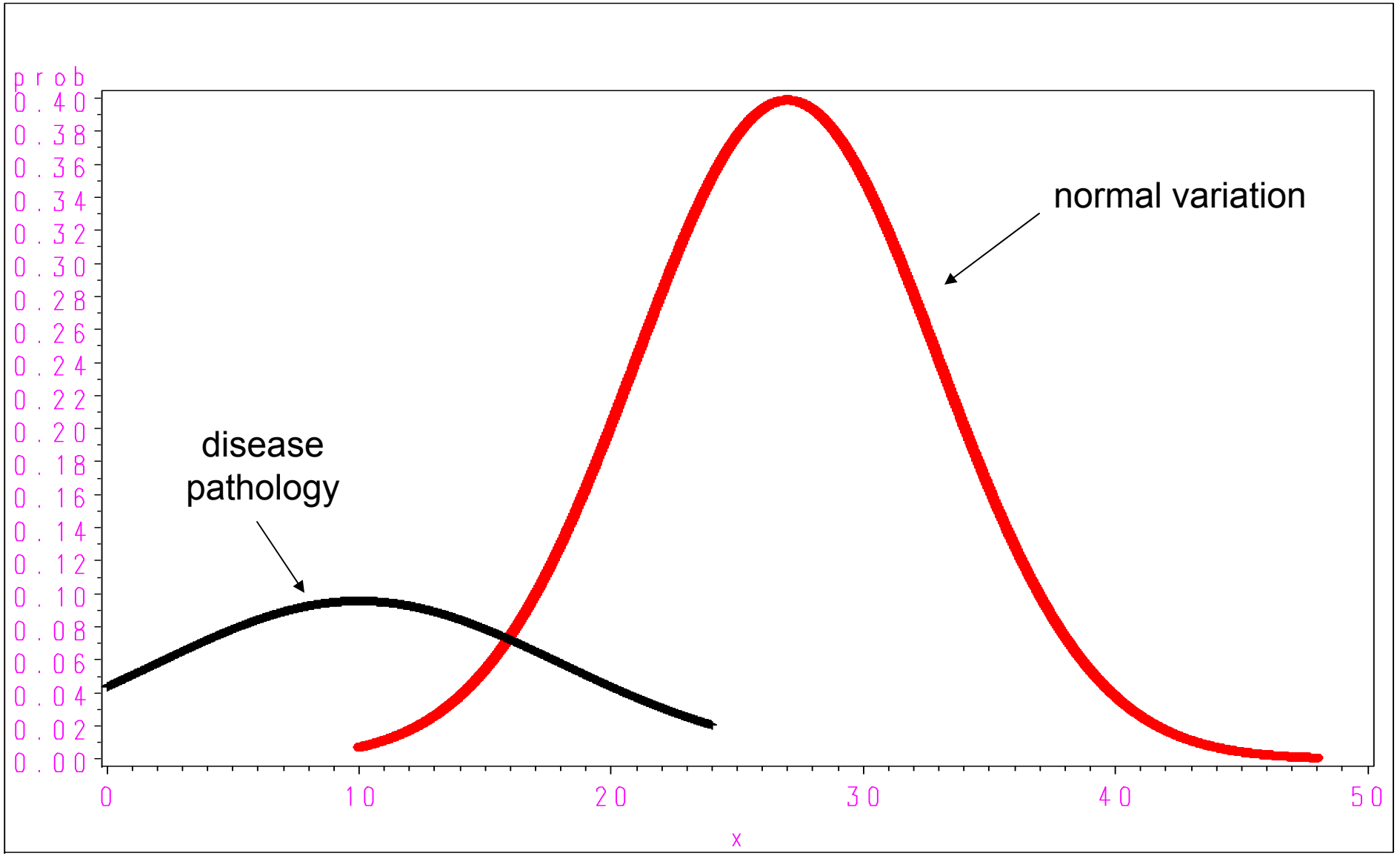
Statistical Issues in Studying Cognitive Decline

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

References:

1. Van Belle G, Uhlmann RF, Hughes JP, Larson EB. Reliability of estimates of changes in mental status test performance in senile dementia of the Alzheimer's type. *Journal of Clinical Epidemiology*. 1990; 43:589-595.
2. Kukull WA. Problems in measuring and interpreting cognitive decline [editorial; comment]. *Journal of the American Geriatrics Society.* 1998; 46(12):1578-9
3. Morris MC, Evans DA, Hebert LE, Bienias JL. Methodological issues in the study of cognitive decline. Am J Epidemiol 1999; 149:789-93.

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



A longitudinal look --- what contribute to decline?

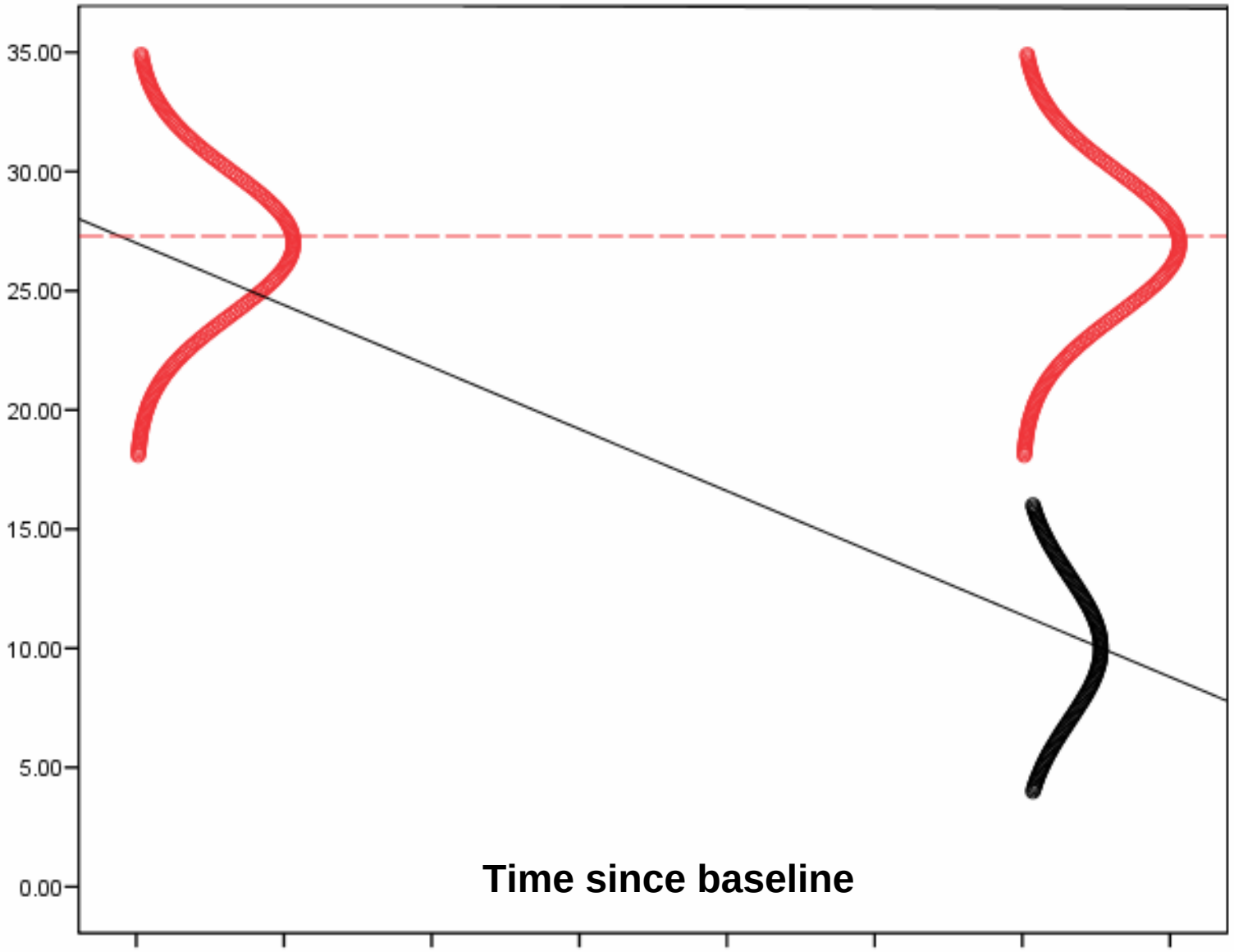
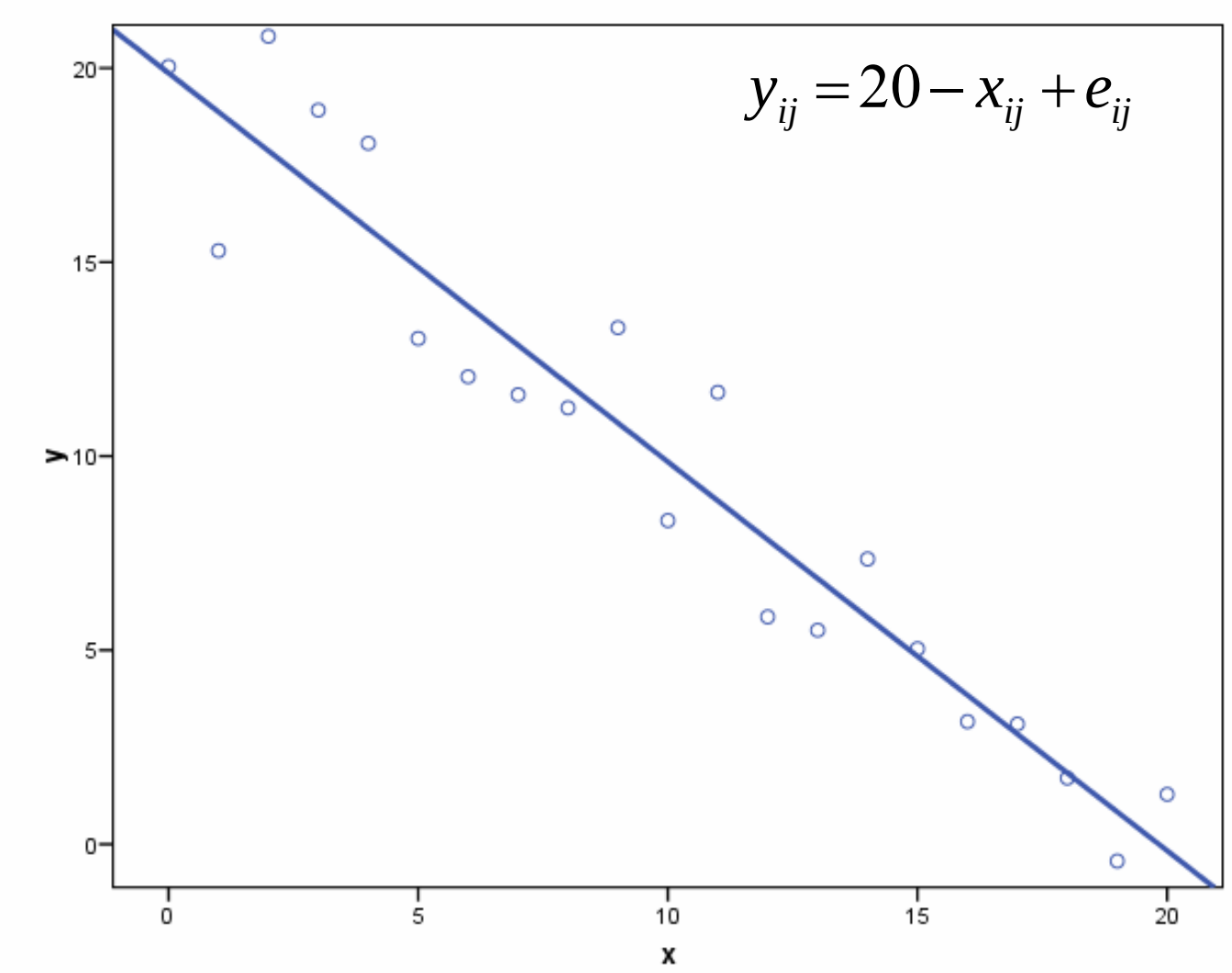
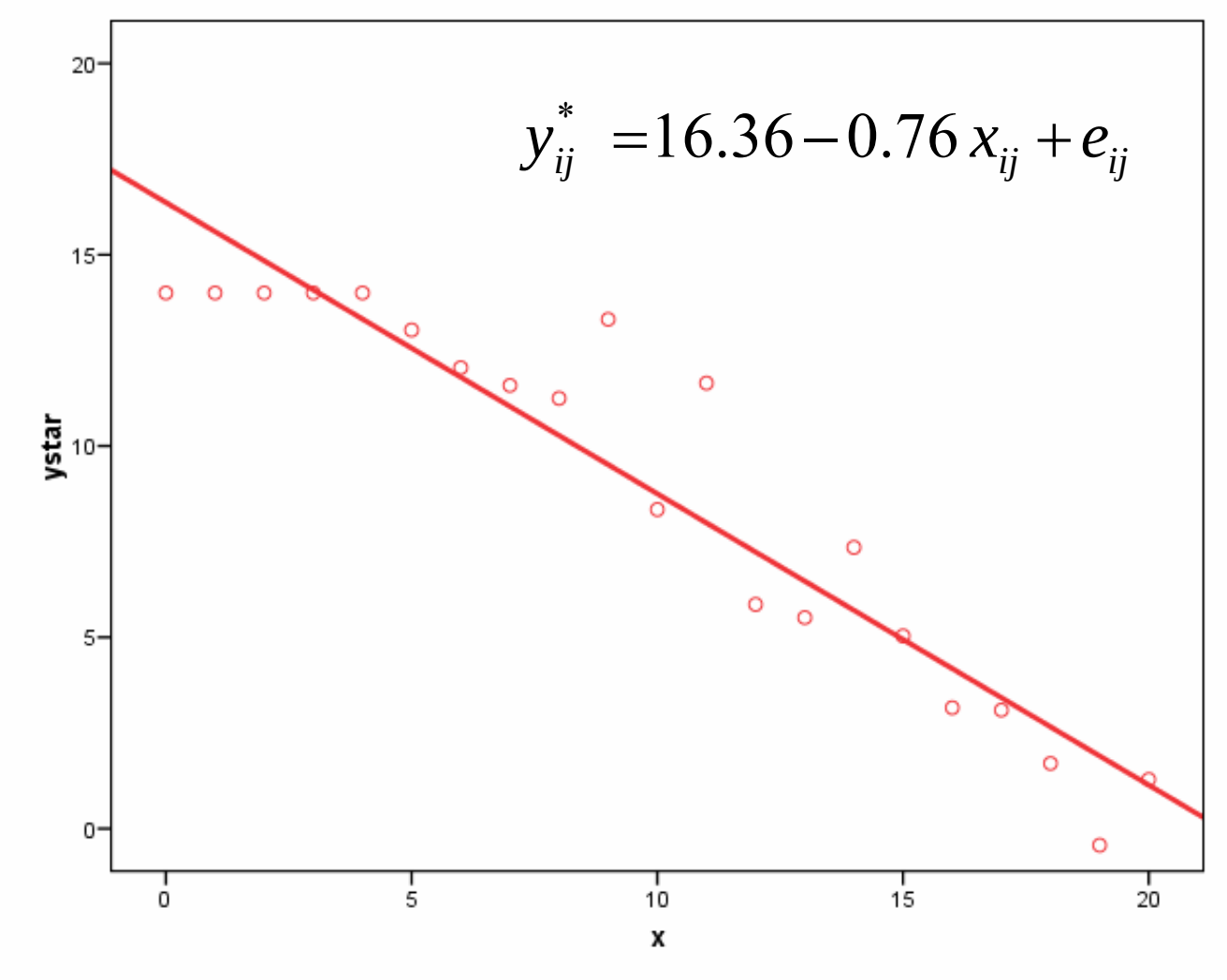


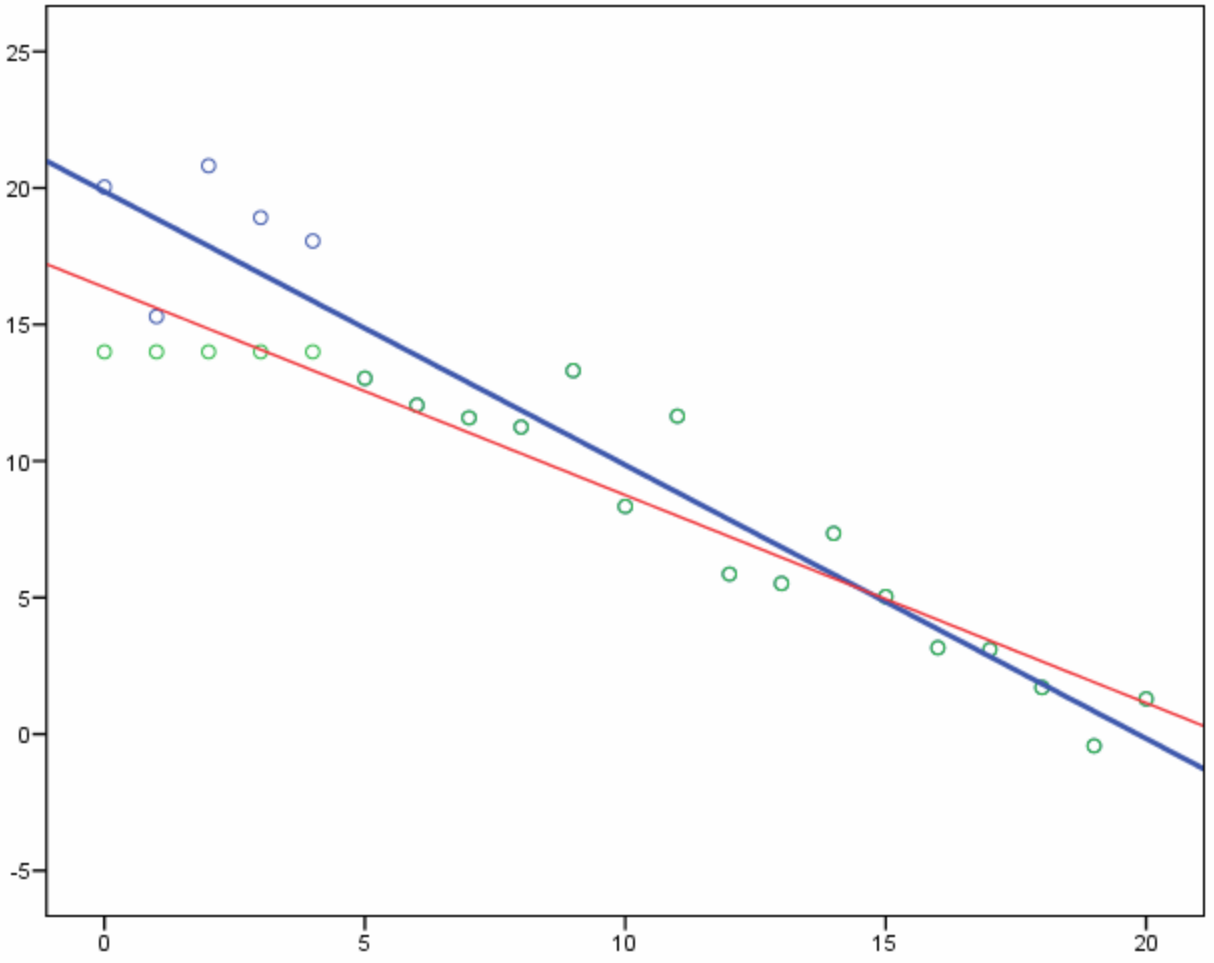
Illustration with simple linear regression



With truncation



Attenuation toward the null



Truncated Outcomes in Linear Regressions

- Ordinary least square estimator is biased and inconsistent

  (Gldberger, 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)*
- Fitzgerald et al 2002, *Stat Med*
- Moulton 2002, *Statistical Methods in Medical Research*
- Wu 2002, *JASA*

Truncated Longitudinal Outcome with Missing Data

Notations:

y_{ij}: the jth measurement from the ith subject.

$\boldsymbol{y}_i$: the vector containing measurement from the ith subject.

$\boldsymbol{X}_{ij}$ and $\boldsymbol{W}_{ij}$: vectors of fixed effect covariates.

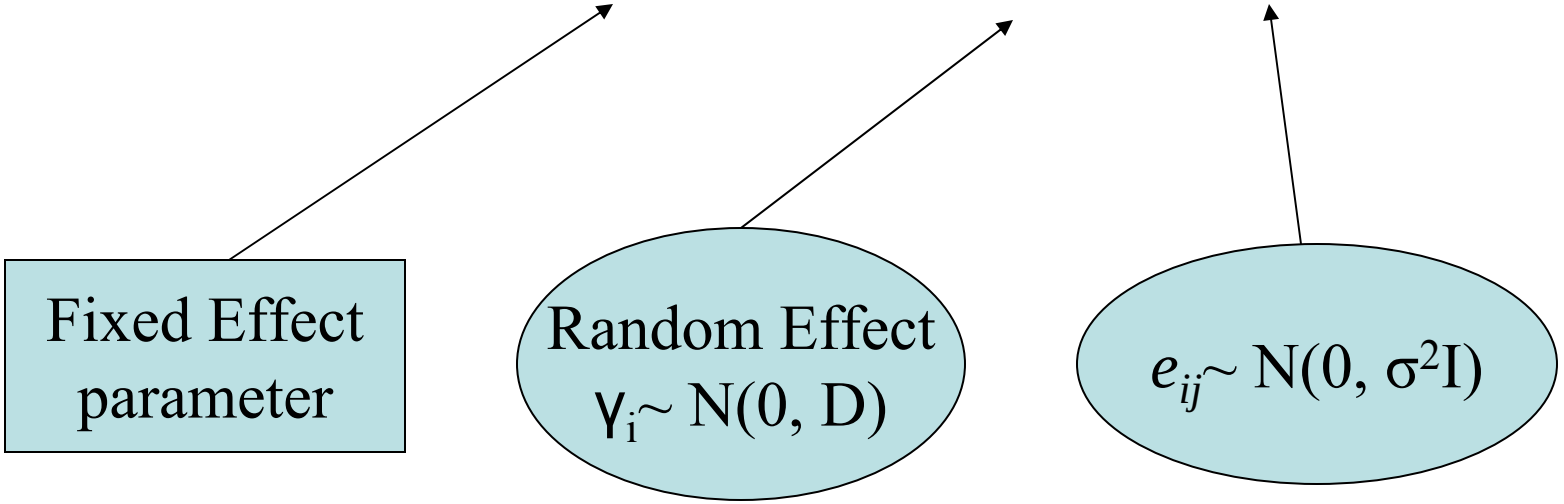
$\boldsymbol{Z}_{ij}$: vector of random effect covariates.

(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

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



$y \sim N(X\beta, V)$ where $V = Z'DZ + \sigma^2 I$

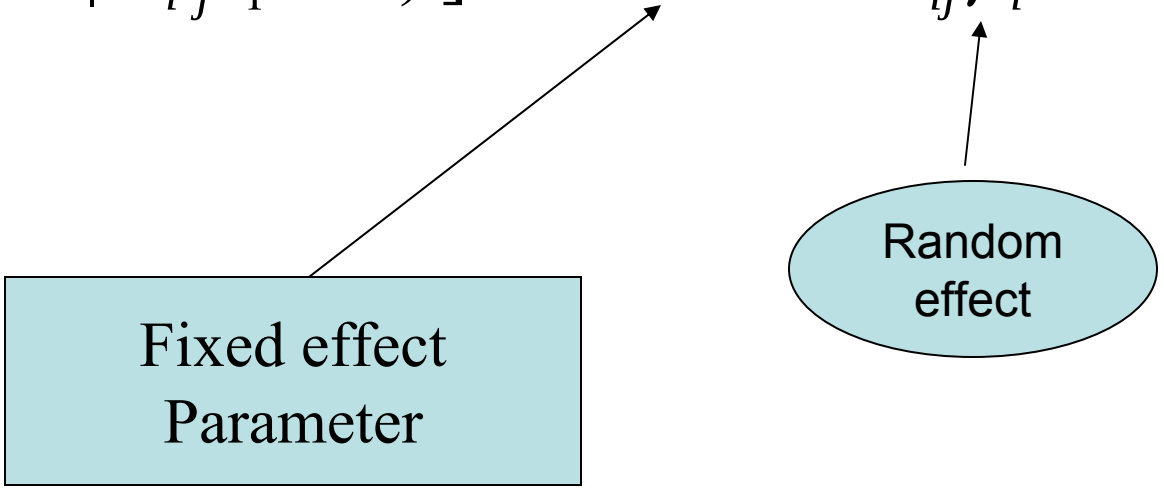
2. The Drop Out Model

Let R_{ij} be an indicator variable for missing

Let p_{ij}=Prob(R_{ij}=*1)*

We assume the following model for the missing data mechanism:

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



Maximum Likelihood Approach

$$L = \prod_{i=1}^{n} \int_{-\infty}^{\infty} \left(\prod_{j=1}^{m_i} (1 - p_{ij}) f(y_{ij} \mid X_{ij}, Z_{ij}, \gamma_i) \right) \quad L_1$$

$$\prod_{j=m_i+1} (1 - p_{ij}) [\int_T^{\infty} f(y_{ij} \mid X_{ij}, Z_{ij}, \gamma_i) dy_{ij}] \quad L_2$$

$$\prod_{R_{ij}=1} p_{in_i+1} \gamma_i \Big) d\gamma_i \quad L_3$$

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 >= 65 at baseline).
- Baseline cognitive evaluation.
- Follow-up evaluations at 2, 5, 8 and 11 years after baseline.
- Identify risk factors for cognitive decline.

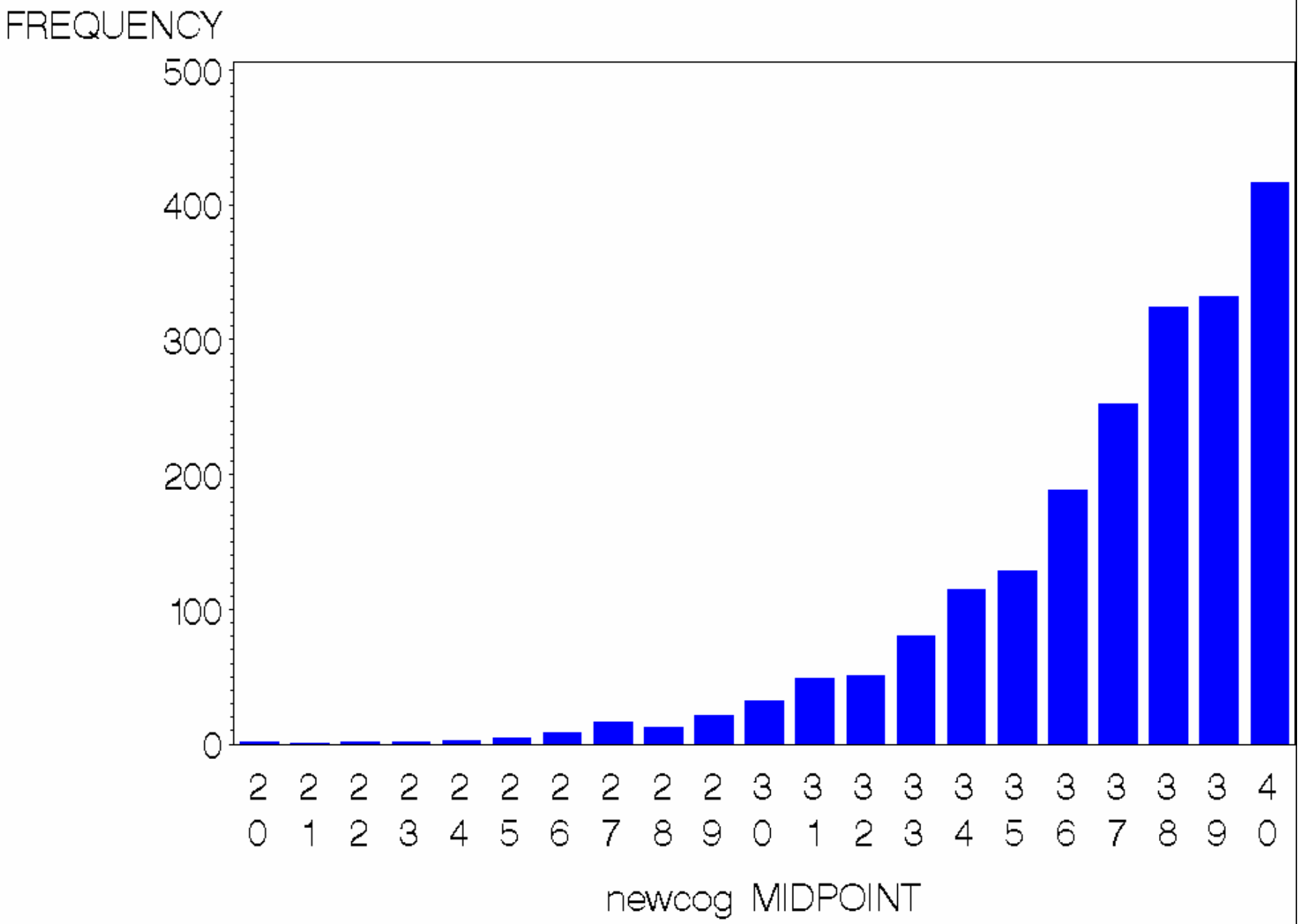
The Indianapolis Ibadan Dementia Project
Cognitive Measures

- 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.

Distribution of Baseline Scores



The Indianapolis Ibadan Dementia Project

- Baseline cognitive scores:
  20.4% at ceiling.

  Difference in mean scores

	Low Education	High Education	P-value
Cognitive score (SD)	35.0 (3.7)	37.1 (3.0)	<0.0001
% at ceiling	5.6	22.8	<0.0001

The Indianapolis Ibadan Dementia Project

Follow up status

Year 2	Year 5	Year 8	Year 11	# subjects
Obs.	Obs.	Obs.	Obs.	425
Obs.	Obs.	Obs	*	267
Obs.	Obs.	*	*	497
Obs.	*	*	*	469
*	*	*	*	370

Obs: cognitive scores observed

*: cognitive scores missing

The Indianapolis Ibadan Dementia Project
Education and Missing Data

% missing	Low education	High education
Year 2	22.5	17.6
Year 5	34.4	27.4
Year 8	61.4	39.1
Year 11	64.3	36.3

The Indianapolis Ibadan Dementia Project

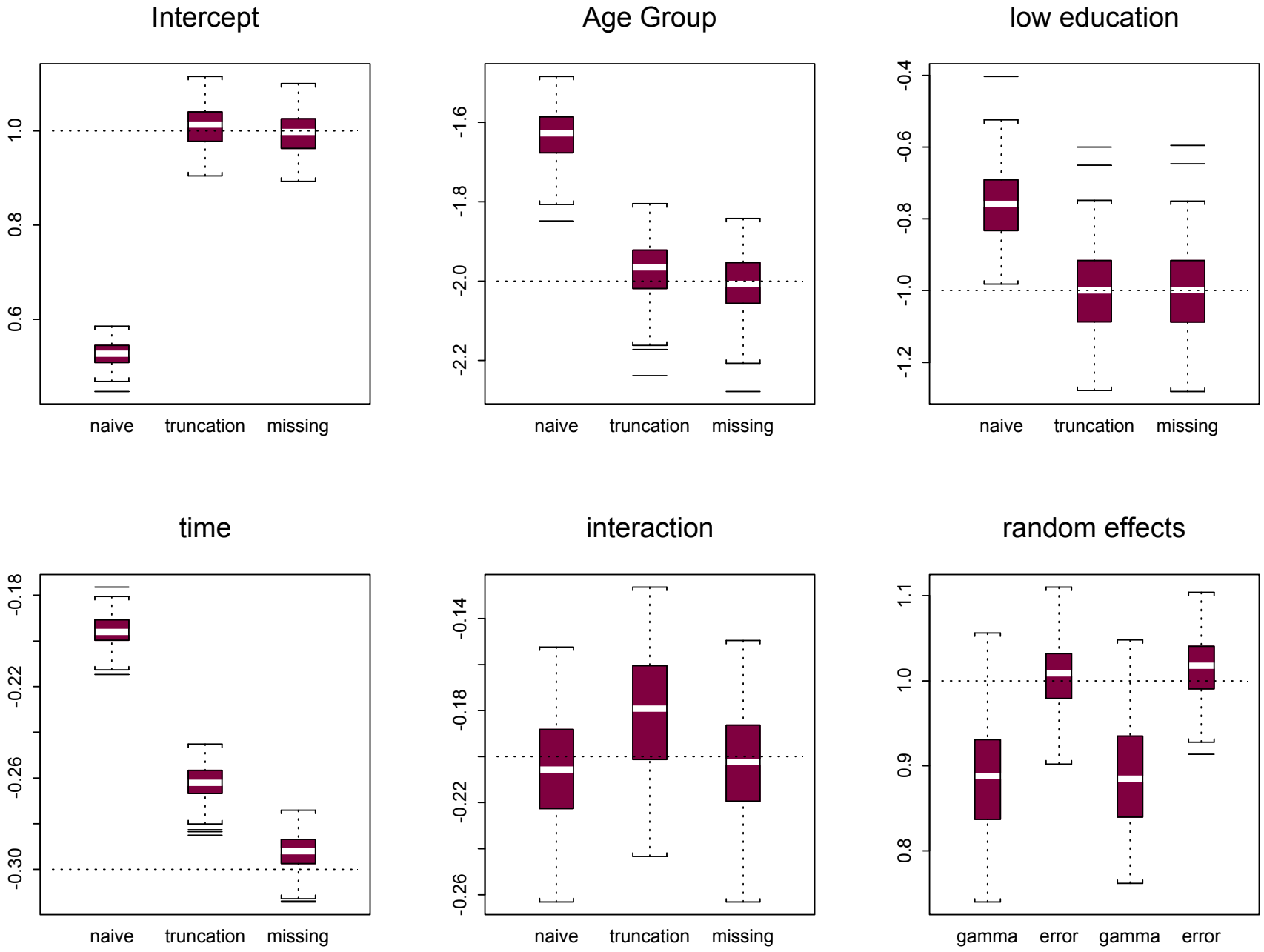
- Education seems to be related to:
  - The truncation process
  - The missing data mechanism

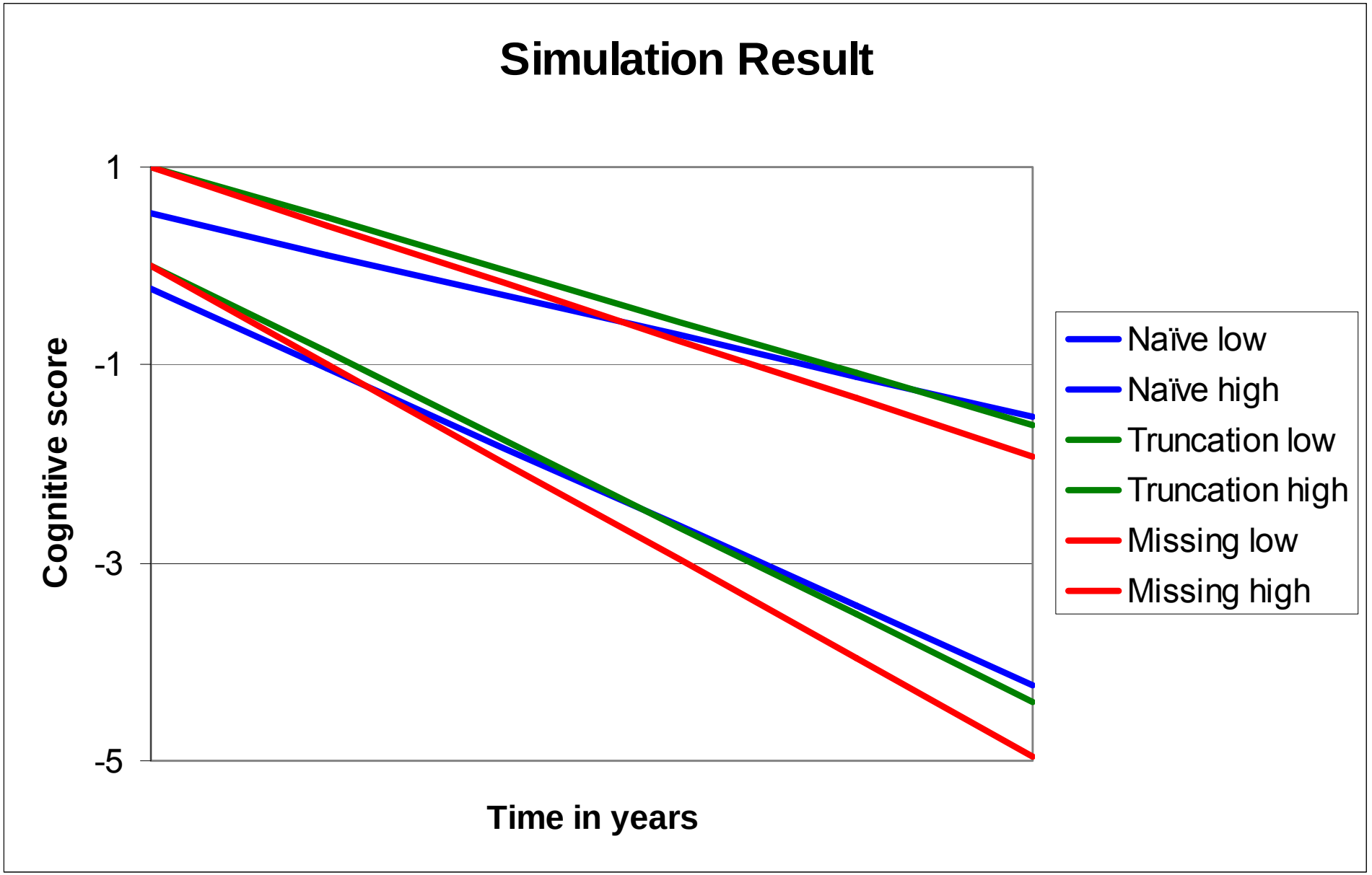
  Our interest:

  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)


Intercept
1.0
0.8
0.6
naive
truncation
missing
Age Group
-1.6
-1.8
-2.0
-2.2
naive
truncation
missing
low education
-0.4
-0.6
-0.8
-1.0
-1.2
naive
truncation
missing
time
-0.18
-0.22
-0.26
-0.30
naive
truncation
missing
interaction
-0.14
-0.18
-0.22
-0.26
naive
truncation
missing
random effects
1.1
1.0
0.9
0.8
gamma
error
gamma
error


Simulation Result
1
-1
-3
-5
Cognitive score
Time in years
Naïve low
Naïve high
Truncation low
Truncation high
Missing low
Missing high

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)