UPPSALA
UNIVERSITET

Covariate Modeling in Aggregate Data Meta Analysis: A Simulation Study Evaluating Performance of Model Linearization

Patanjali Ravva, MS
Mats O. Karlsson, PhD
Jonathan L. French, ScD

PAGE 2011, Athens
08 June 2011

Introduction

- Model based meta-analyses have become an important tool in informing decision making in drug development.
- The main objectives of model based meta-analyses include:
  - Quantifying the treatment effect size and variability
  - Characterizing the heterogeneity
  - Others
- The question to ask ourselves when performing these analyses
  - Are we interested in making inferences at a group/summary (AD) level? or
  - Do we want to make inferences at an individual patient level (IPD)?
    - Aggregation Bias

Introduction (cont'd)

- What type of data do we have?
  - Landmark vs. longitudinal
- How do we deal with individual patient level data (IPD) from some studies
  - Reduce to summary measure or some how include IPD.
- Does the model include covariate relationships? If so, how are covariates incorporated in the model?

Introduction (cont'd)

- To address these questions, it helps viewing the summary level data as an aggregation of individual patient data over a grouping variable to derive an aggregate model based on individual patient level model
- This framework allows us to explore/understand
  - ways to incorporate individual level data along with aggregate data for efficient use of all the data
  - sources of bias when a simple aggregate model is fit to the data
  - inclusion of covariates into the model
  - type of data we need to gather from literature when building literature databases

Linearization Method

- Lets consider a meta-analysis describing the dose-response relationship (continuous endpoint and landmark data) for a specific class of compounds.
- Typically, the studies included are parallel group dose-ranging studies (doses: d=placebo, d1, d2, d3, d4)
- Let y_{ikj} be the response for j^{th} subject from k^{th} arm of an i^{th} trial

Linearization Method (cont'd)

Two common IPD models for these pooled data across trials is

$$f(x,z,dose) = y_{ikj} = (e_{0i}) - \left(\frac{emax \cdot \theta_1 \cdot x_{ikj} \pm \theta_2 \cdot z_{ikj} + dose_{ij}}{ed_{50} + dose_{ikj}}\right) + \epsilon_{ikj} \quad (1)$$

$$f(x,z,dose) = y_{ikj} = (e_{0i}) - \left(\frac{emax \cdot \theta_1 \cdot \left(\frac{x_{ikj}}{a}\right)^{\theta_1} \pm \theta_2 \cdot \left(\frac{z_{ikj}}{b}\right)^{\theta_2} + dose_{ikj}}{ed_{50} + dose_{ikj}}\right) + \epsilon_{ikj} \quad (2)$$

$$\epsilon_{ij} = \mathcal{N}(0, \sigma_i^2)$$

Where x and z are two covariates

For aggregate data, we observe $\bar{Y}_{ik}$ an estimate of the

$$E_x(Y|dose,x) = \int_x f(Y|dose,x)p(x|dose)dx$$

Linearization Method (cont'd)

For the IPD model (1)

$$E_{xz}[f(x,z,dose)] = E_{xz}\left[(e_{0i}) - \left(\frac{emax \cdot \theta_1 \cdot x_{ikj} \pm \theta_2 \cdot z_{ikj} + dose_{ikj}}{ed_{50} + dose_{ikj}}\right) + \epsilon_{ikj}\right] \quad (4$$

$$= (e_{0i}) - \left(\frac{emax \cdot \theta_1 \cdot E_x(x_{ik}) \pm \theta_2 \cdot E_z(z_{ik}) + dose_{ikj}}{ed_{50} + dose_{ikj}}\right) + E(\epsilon_{ikj}) \quad (5$$

We can approximate (4) by replacing expected values of the covariates with $\bar{x}_{ik}$ and $\bar{z}_{ik}$

$$= (e_{0i}) - \left(\frac{emax \cdot \theta_1 \cdot (\bar{x}_{ik}) \pm \theta_2 \cdot (\bar{z}_{ik}) + dose_{ikj}}{ed_{50} + dose_{ikj}}\right) \quad (6$$

Linearization Method (cont'd)

For IPD model(2)

$$E_{xz}[f(x,z,dose)] = E\left[e_{0i} - \left(\frac{emax * \left(\frac{x_{ikj}}{a}\right)^{\theta_1} \cdot \left(\frac{z_{ikj}}{b}\right)^{\theta_2} \cdot dose_{ij}}{ed_{50} + dose_{ikj}}\right) + \epsilon_{ikj}\right] \quad (7)$$

$$\neq \left[e_{0i} - \left(\frac{emax * \left(\frac{E_x(x_{ik})}{a}\right)^{\theta_1} \cdot \left(\frac{E_z(z_{ik})}{b}\right)^{\theta_2} \cdot dose_{ij}}{ed_{50} + dose_{ikj}}\right) + E\left(\epsilon_{ikj}\right)\right]$$

Linearization Method (cont'd)

- However, an approximate equation can be derived by linearization of model (eq. 2) using second order Taylor series approximation.
- This approximation for a bivariate function in a generic form is presented below where x and z are the two variables of function f :

Linearization Method (cont'd)

$$f(x,z) \approxeq f(\mu_x,\mu_z) + f'_x(\mu_x,\mu_z)(x-\mu_x) + f'_z(\mu_x,\mu_z)(z-\mu_z) + \frac{1}{2}\cdot f''_x(\mu_x,\mu_z)(x-\mu_x)^2 + \frac{1}{2}\cdot f''_z(\mu_x,\mu_z)(z-\mu_z)^2 + f'_{xz}(\mu_x,\mu_z)(x-\mu_x)(z-\mu_z)$$

$$E[f(x,z)] \approxeq f(\mu_x,\mu_z) + \frac{1}{2}\cdot f''_x(\mu_x,\mu_z)\sigma_x^2 + \frac{1}{2}\cdot f''_z(\mu_x,\mu_z)\sigma_z^2 + f'_{xz}(\mu_x,\mu_z)Cov(x,z)$$

$$\approxeq f(\mu_x,\mu_z) + \frac{1}{2}\cdot f''_x(\mu_x,\mu_z)\sigma_x^2 + \frac{1}{2}\cdot f''_z(\mu_x,\mu_z)\sigma_z^2 + f'_{xz}(\mu_x,\mu_z)\sigma_x\sigma_z\rho_{x,z}$$

Simulation Method

- Simulations to understand the effect of degree of nonlinearity with respect to covariate effects and between-trial to within-trial variability of covariates on estimated model parameters when modeled using individual patient data (IPD), aggregate data (AD), linearized AD and linearized combined AD and IPD (AD_IPD) models

Simulation Method (cont'd)

- No. of drugs: 5
- No. of studies: 18
  - DrgA: 2,
  - DrgB, DrgC, DrgD and DrgE – 4 studies each
- No. of doses per study: 5 (including placebo)
- No. of subjects per dose: 50
- No. of simulations per scenario: 500

Simulation Method (cont'd)

- Analysis Datasets for each simulation:
  - IPD, AD, AD_IPD (IPD - 2 studies from drug A)
    - First IPD data are simulated then reduced to create AD and AD_IPD datasets
- Analysis Models
  - IPD data with IPD model
  - AD data with AD and AD_Lin models
  - AD_IPD with AD_IPD_Lin Model
- Evaluation
  - The bias and precision in parameter estimates under all scenarios were assessed as mean estimation error and relative root mean squared error (RMSE), respectively.

Simulation Method (cont'd)
Simulation Scenarios

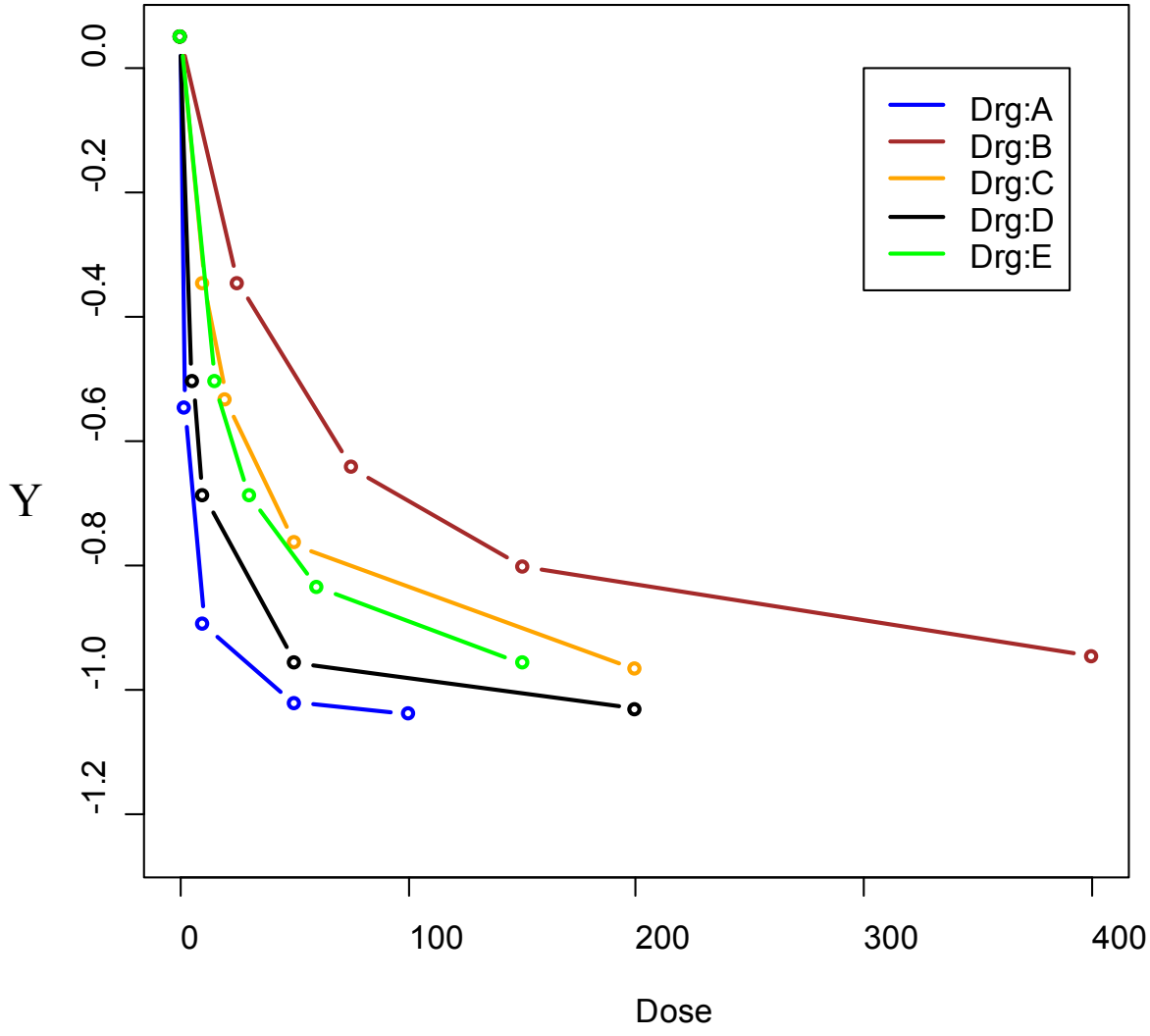
		Degree of Nonlinearity		
Covariate Distribution: Between Study to within Study variability		Small	Moderate	Large
	Small	x	x	x
	Moderate	x	x	x
	Large	x	x	x

Simulation Method (cont'd)

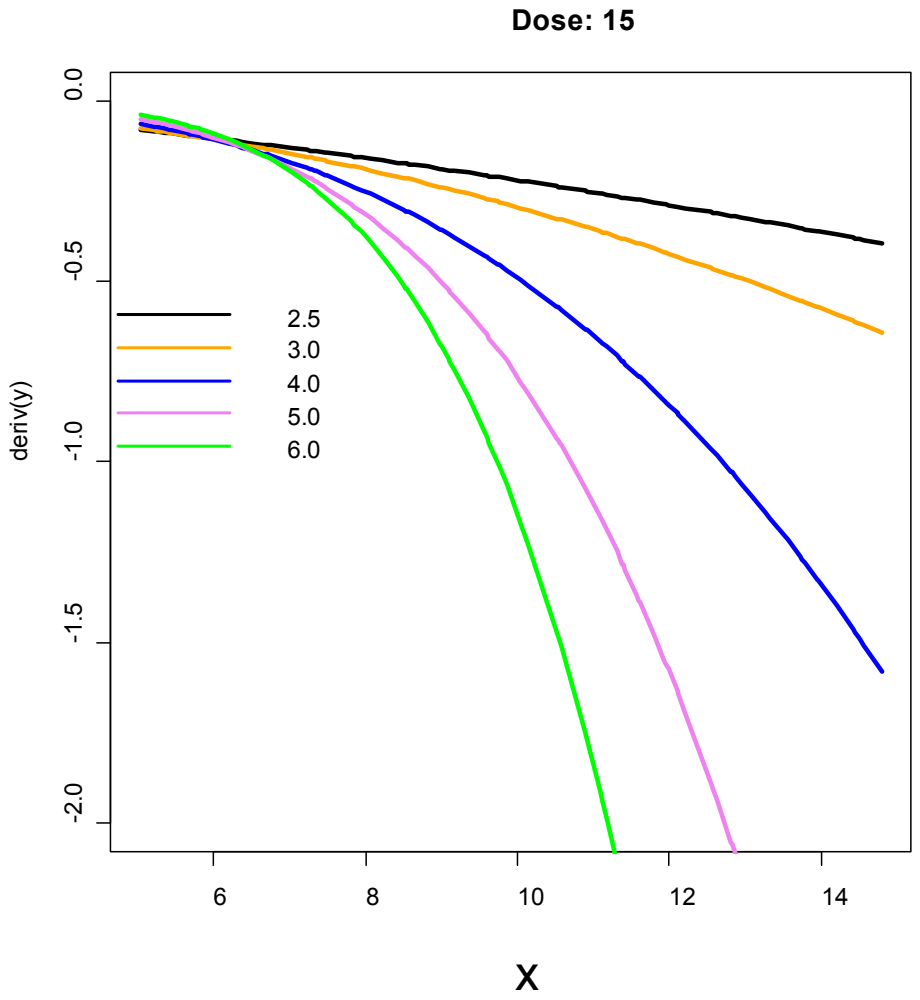
Table 1: Model parameter estimates for simulation

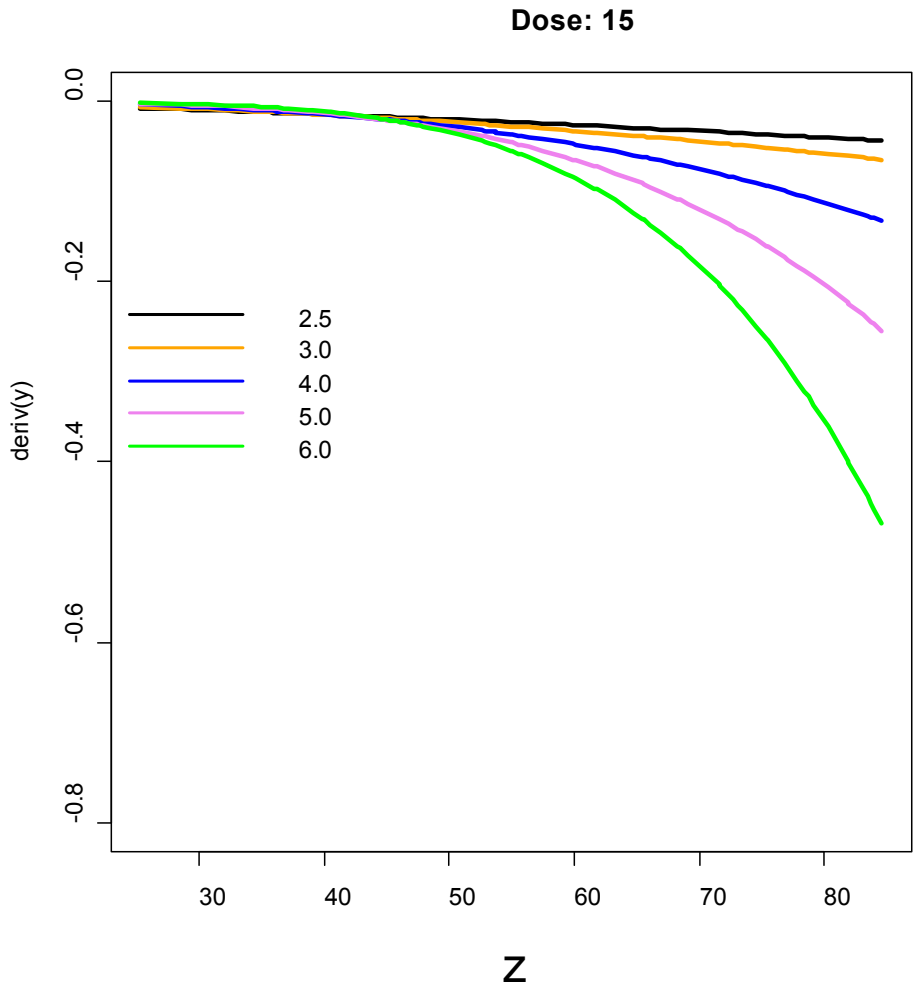
Parameter	Estimates	Parameter	Estimates	Parameter	Estimates
E_0	0.0506	$\theta_{1,S}$	2.5	$\omega^2_{BTV,S}(X); \omega^2_{BSV,S}(X)$	0.227;1.280
Emax	-1.11	$\theta_{1,M}$	4	$\omega^2_{BTV,M}(X); \omega^2_{BSV,M}(X)$	0.454;0.640
$ED_{50.DrgA}$	1.72	$\theta_{1,L}$	6	$\omega^2_{BTV,L}(X); \omega^2_{BSV,L}(X)$	0.909;0.320
$ED_{50.DrgB}$	45	$\theta_{2,S}$	0.889	$\omega^2_{BTV,S}(Z); \omega^2_{BSV,S}(Z)$	6.09;80.60
$ED_{50.DrgC}$	18	$\theta_{2,M}$	4	$\omega^2_{BTV,M}(Z); \omega^2_{BSV,M}(Z)$	12.18;40.30
$ED_{50.DrgD}$	5	$\theta_{2,L}$	6	$\omega^2_{BTV,L}(Z); \omega^2_{BSV,L}(Z)$	24.37;20.15
$ED_{50.DrgE}$	15			$\rho_{BTV}(X,Z)$	0.569
ω^2_{E0}	0.0754			$\rho_{BSV}(X,Z)$	0.137
σ^2	0.515				

Dose-Response Relationship

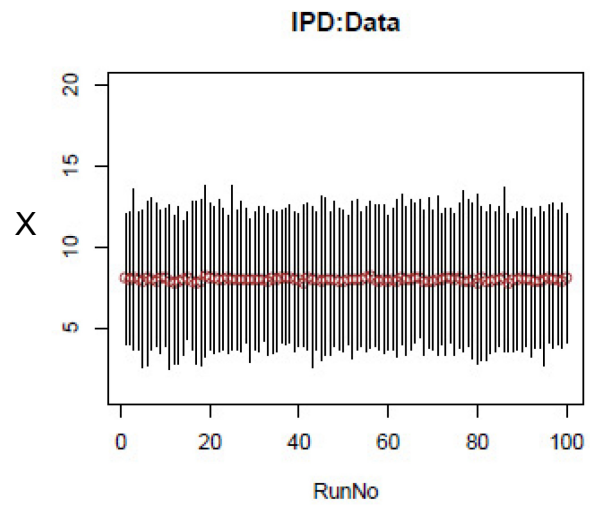


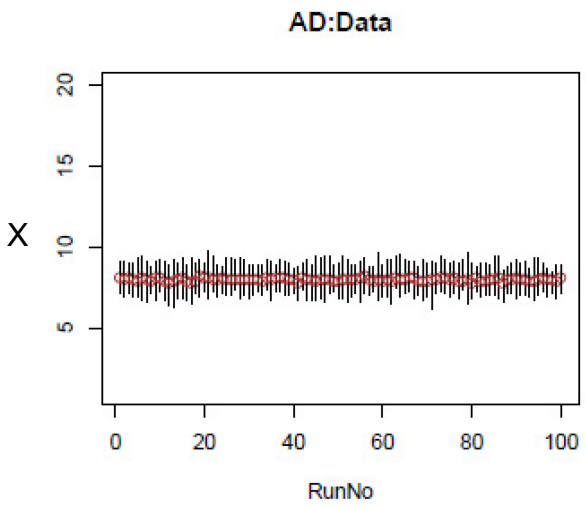
Degree of nonlinearity in covariate-response relationship

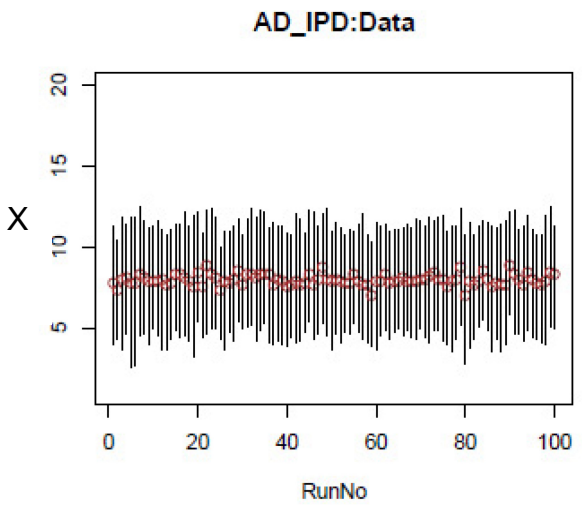


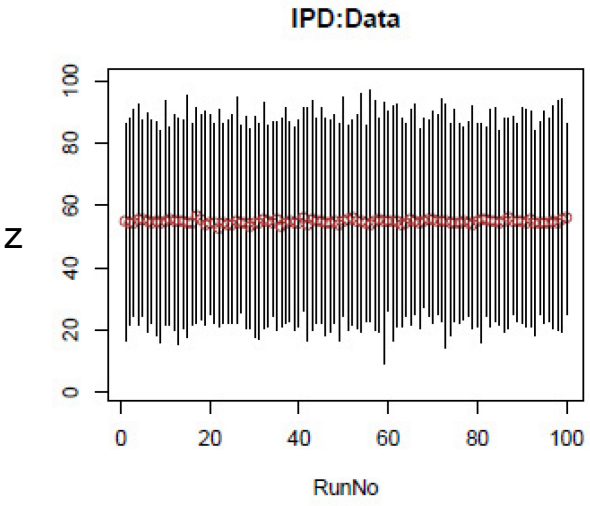


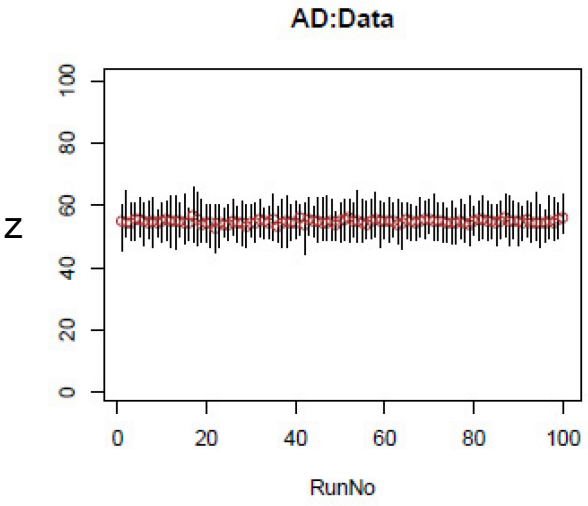
Distribution of Covariates: Between Study vs. Within Study (Scenario 1)

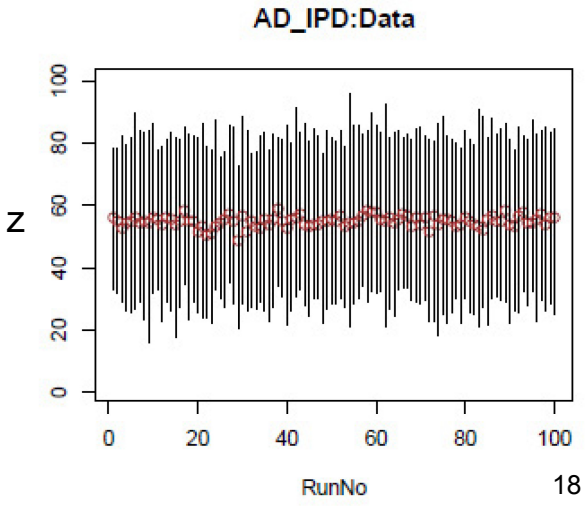




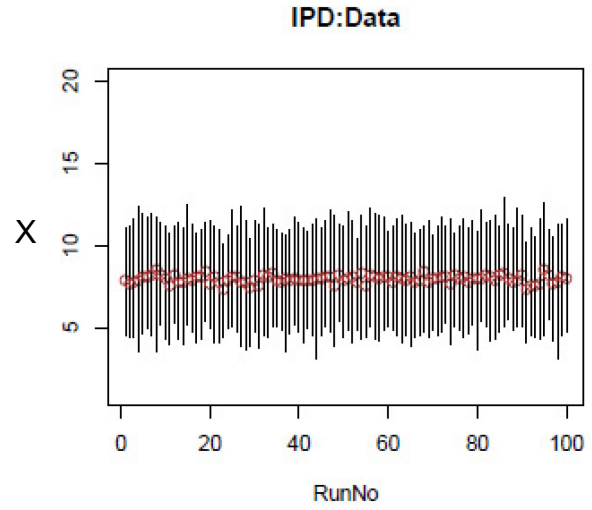


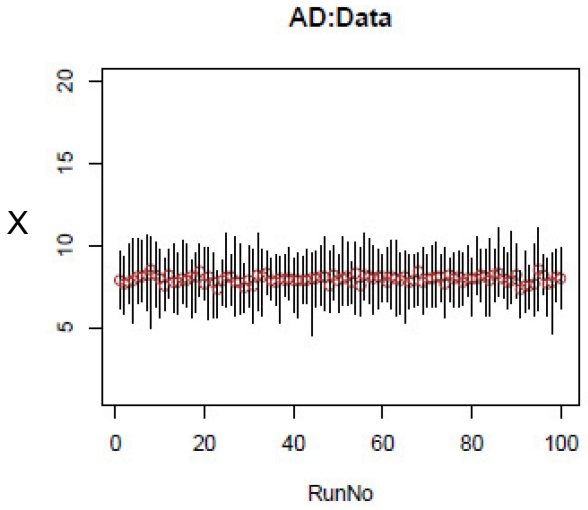


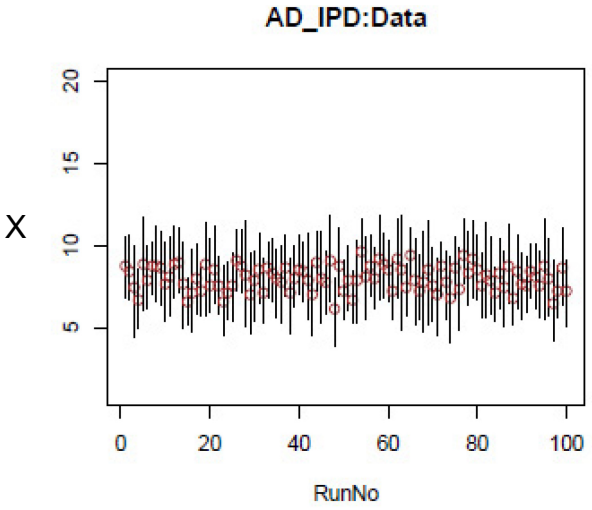


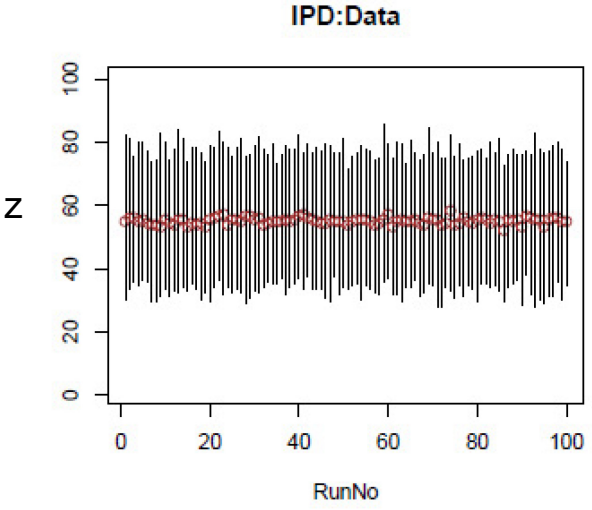


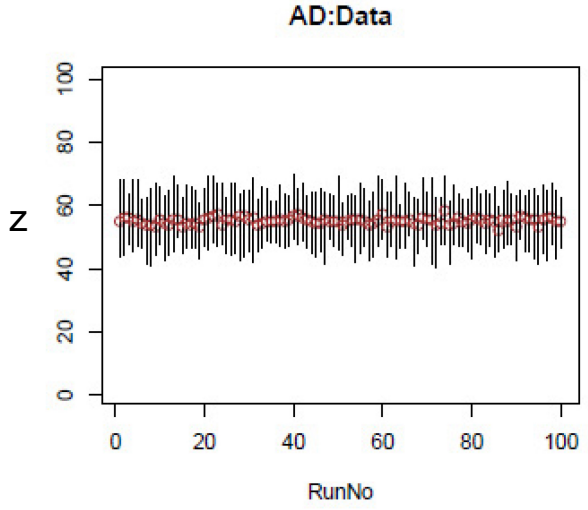
Distribution of Covariates: Between Study vs. Within Study (Scenario 5)

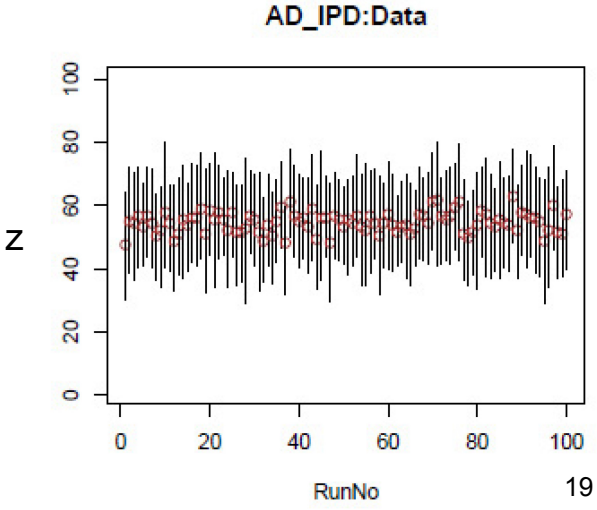






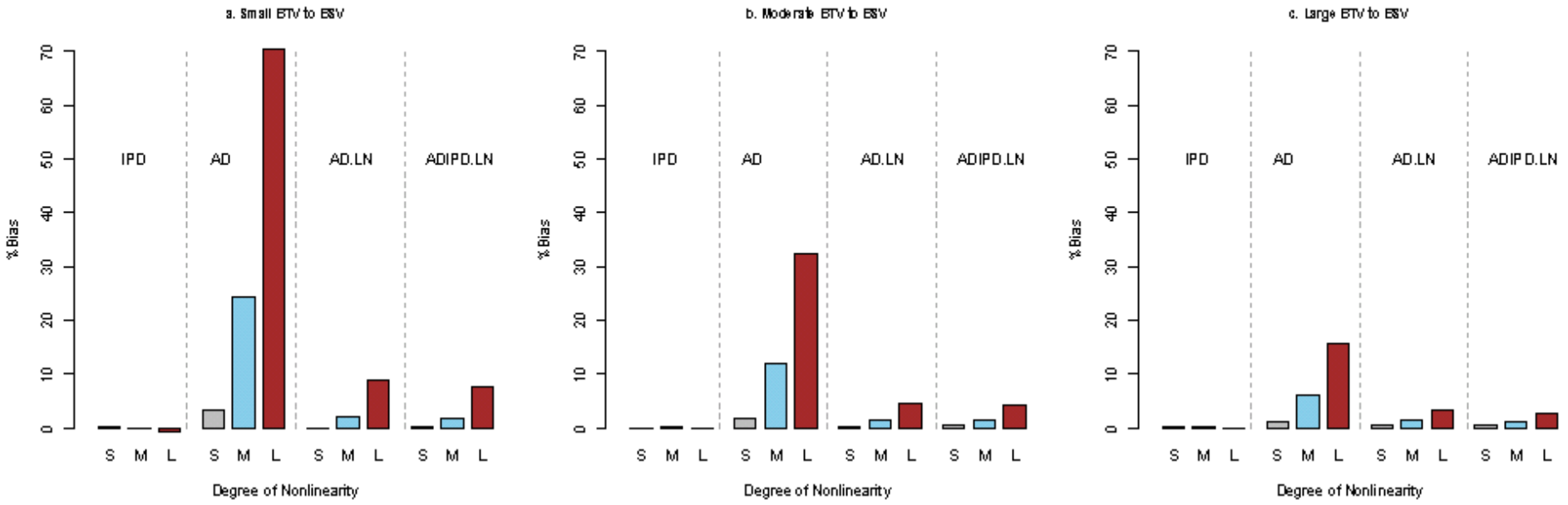






Results: Bias in Emax Parameter

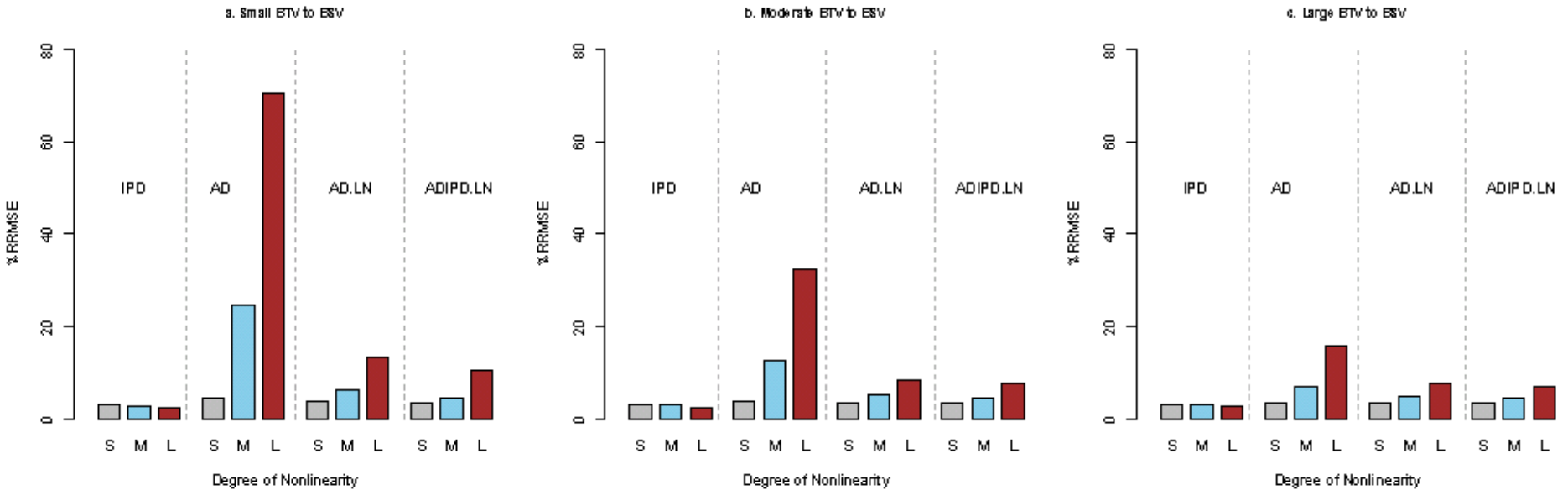
Figure 2: Percent Bias in Emax



- With increasing degree of nonlinearity in the model (with respect to covariates), the bias in the estimates for the emax parameter increased noticeably for AD model.
- Appropriately derived aggregation models (AD Lin & ADIPD Lin) using a linearization approach adequately corrected the bias in the emax parameter.

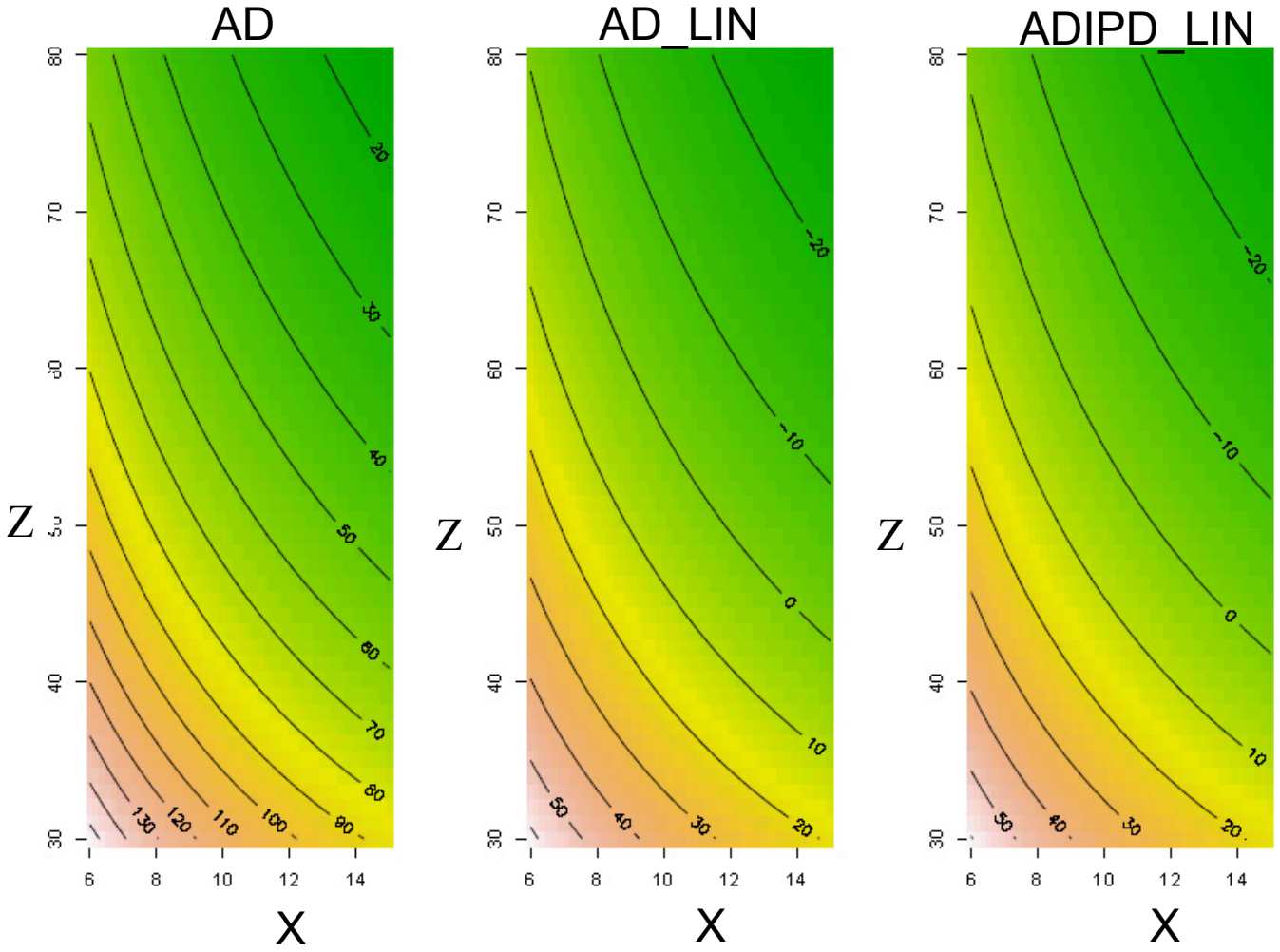
Results: Percent RMSE in Emax Parameter

Figure 3: Percent RMSE in Emax



• RMSE for emax parameter using linearized models were signficantly lower compared to the simpler AD model across varying degrees of nonlinearity
• As the ratio of between-trial to within-trial variability (BTV/BSV) in covariates increased, the bias and RMSE in emax parameter under the AD model decreased.

Results: Large nonlinearity & small BTV to BSV



Conclusion

- The proposed linearization method adequately addressed the issue of aggregation bias when modeling aggregate data using nonlinear models.
- With increasing heterogeneity in covariates (as assessed by the ratio of BTV to BSV), the bias in the model parameter estimates decreases.