Time-varying model for longitudinal data measured with error with application to physical activity and sleep

Victor Kipnis

Biometry
National Cancer Institute

Collaborators:

Heather R. Bowles, NCI
Raymond J. Carroll, Texas A&M University
Laurence S. Freedman, Gertner Institute, Israel
Daniel A. Kipnis, Yale University
James J. McClain, NCI
Douglas Midthune, NCI

Outline

- Effect of physical activity on sleep quality: existing evidence and involved problems
- Longitudinal studies: three different effects and their interpretation
- New joint time -varying model for longitudinal data
- Application to BodyMedia data
- Simulation study
- Discussion

Existing evidence and involved problems

- Inadequate sleep has been linked to many serious diseases, (cancer, diabetes, cardiovascular disease, and obesity)
- There is strong theoretical rationale for expecting physical activity (PA) to be related to sleep improvement
- Supporting evidence from experimental and epidemiologic studies is not compelling
- At least 2 major reasons for discrepancies:
  - limited sample size in trials consisting of good sleepers
  - dependence on self-report in epi studies

Recent longitudinal studies with objective measures

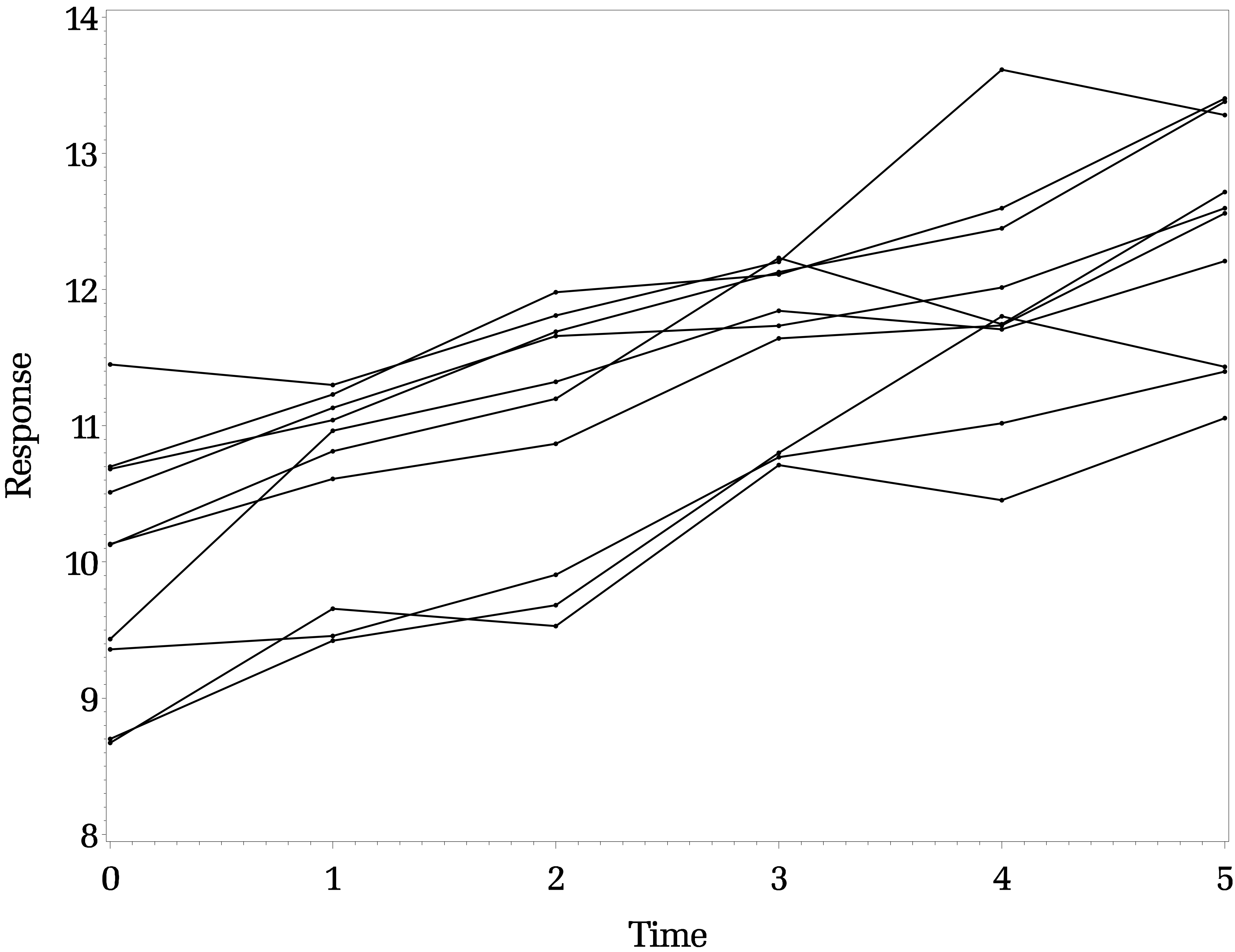
- Three recent studies used ActiGraph accelerometer for seven consecutive days to measure PA and sleep characteristics and to estimate *individual* relationship between PA and sleep
- Results are controversial: for *sleep efficiency* (ratio of sleep minutes to lying down minutes), Pesonen et al. (2011) found *negative* effect, Lambiase et al. (2013) found *no* effect, while Ekstedt et al. (2013) found *positive* effect

Recent longitudinal studies with objective measures

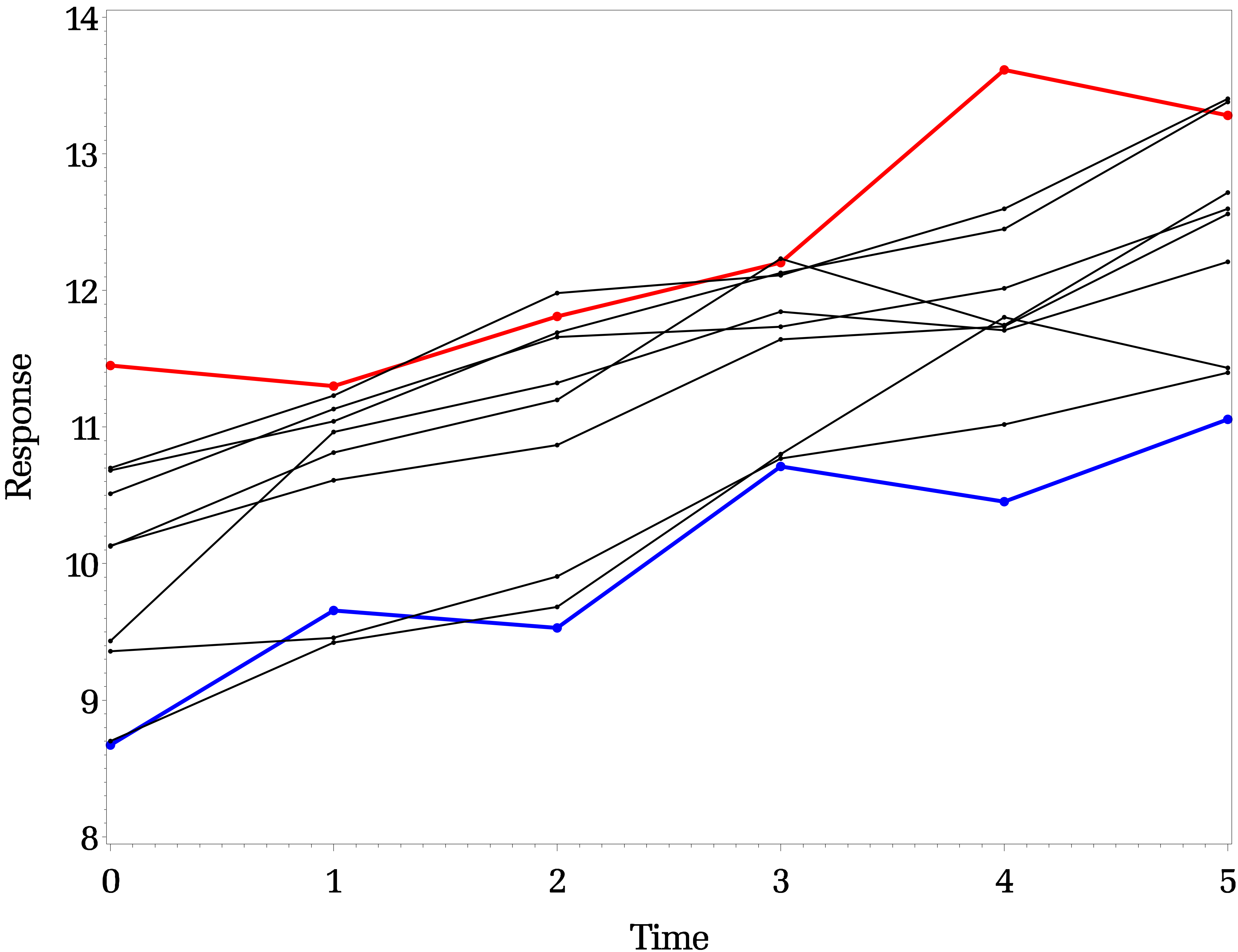
- All three studies used a linear mixed model with a random intercept to estimate *individual-level* (*within-subject*) *effect*
- Pesonen et al. & Lambiase et al. used *measured PA* as their main exposure; Ekstedt et al. used temporal *deviations of PA* measurements from the corresponding within-person means
- Could this lead to estimates with different properties?
- Complication – no study addressed *measurement error* in accelerometry assessment of PA and sleep

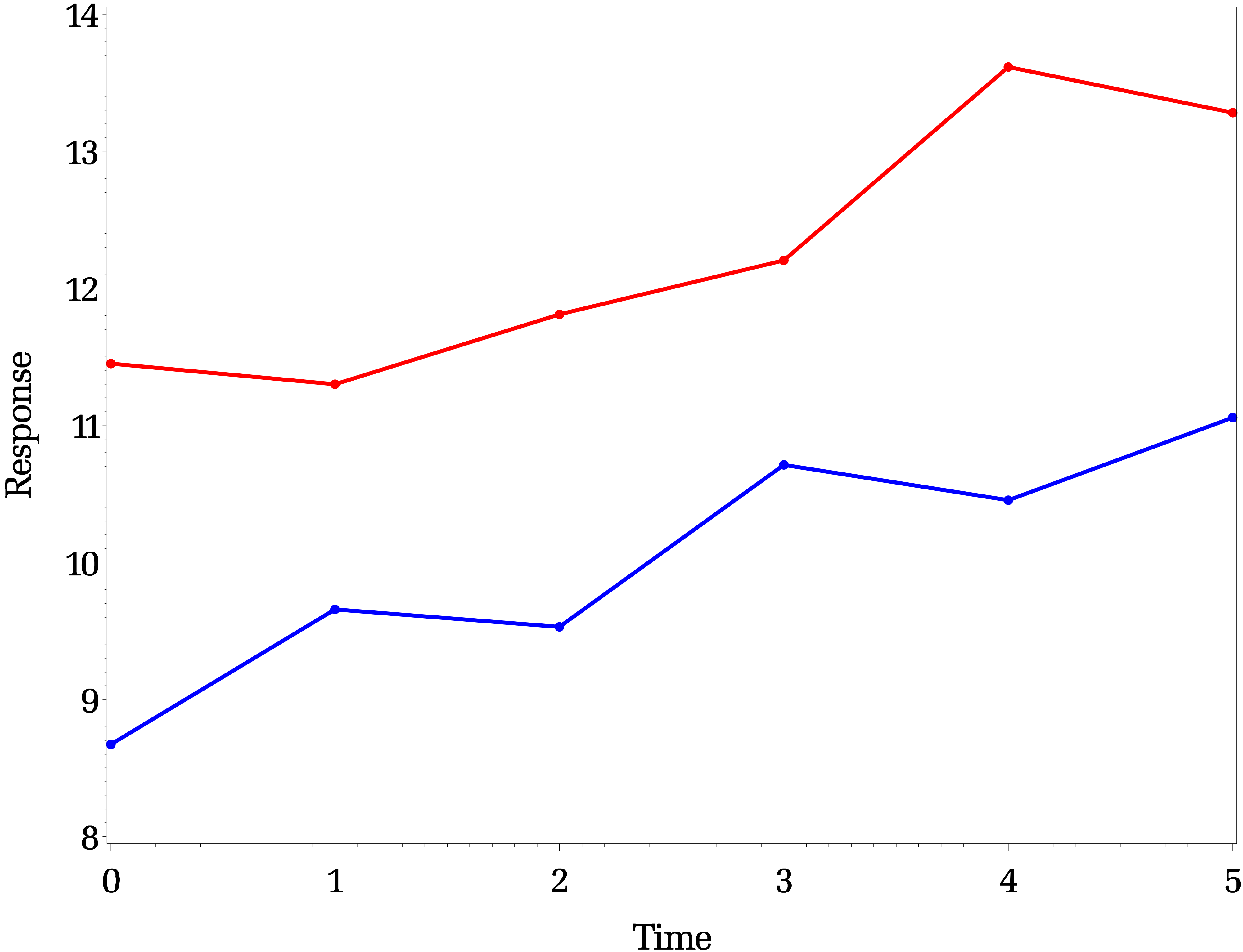
Longitudinal studies

- Defining feature: measurement are taken of the same subjects repeatedly over time
- Primary goal (raison d'être): analysis of within-subject change in health outcome and factors that influence this change over time
- Analyzing within-subject change removes extraneous variation among subjects because they serve as their own controls

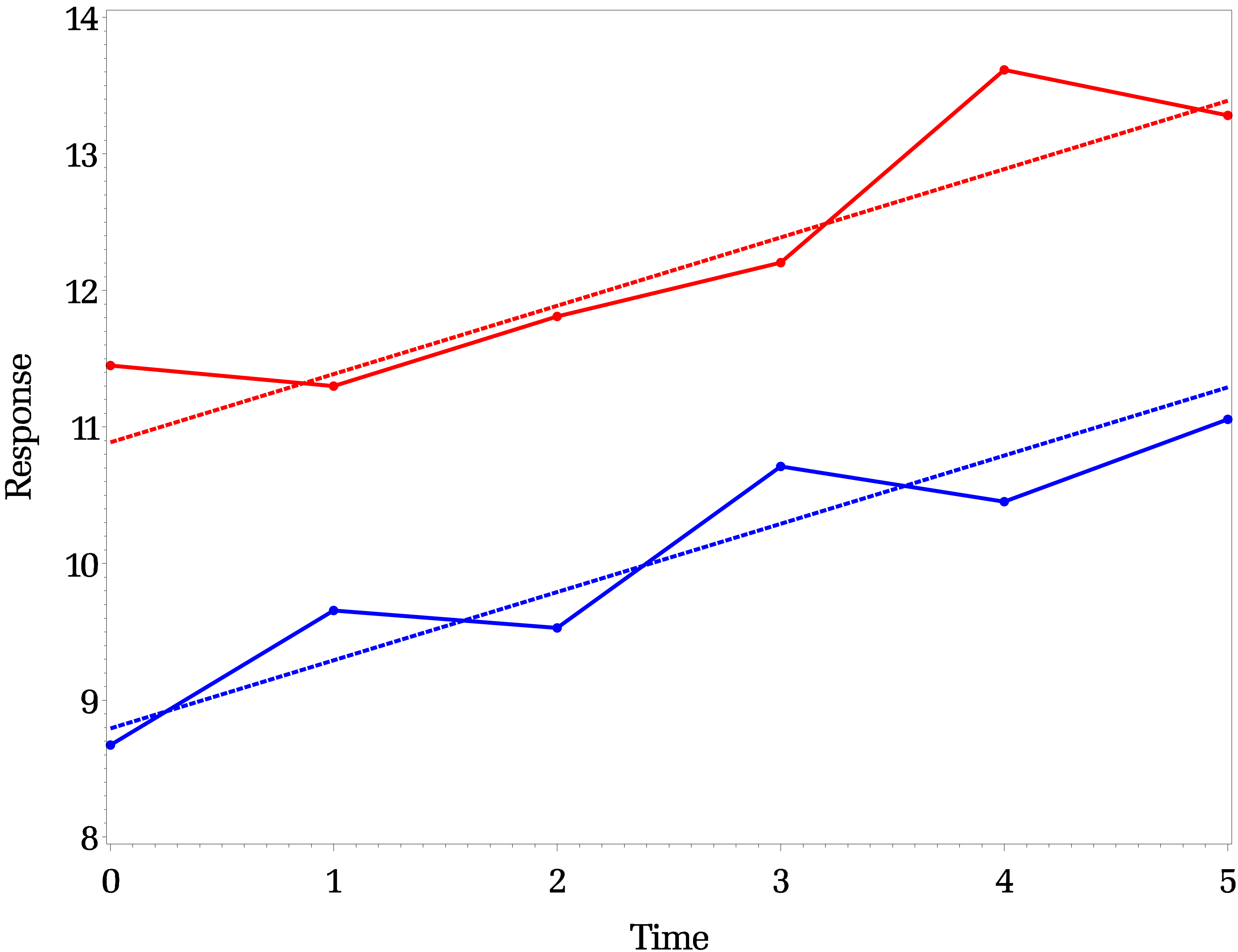


14
13
12
11
10
9
8
Response
0
1
2
3
4
5
Time

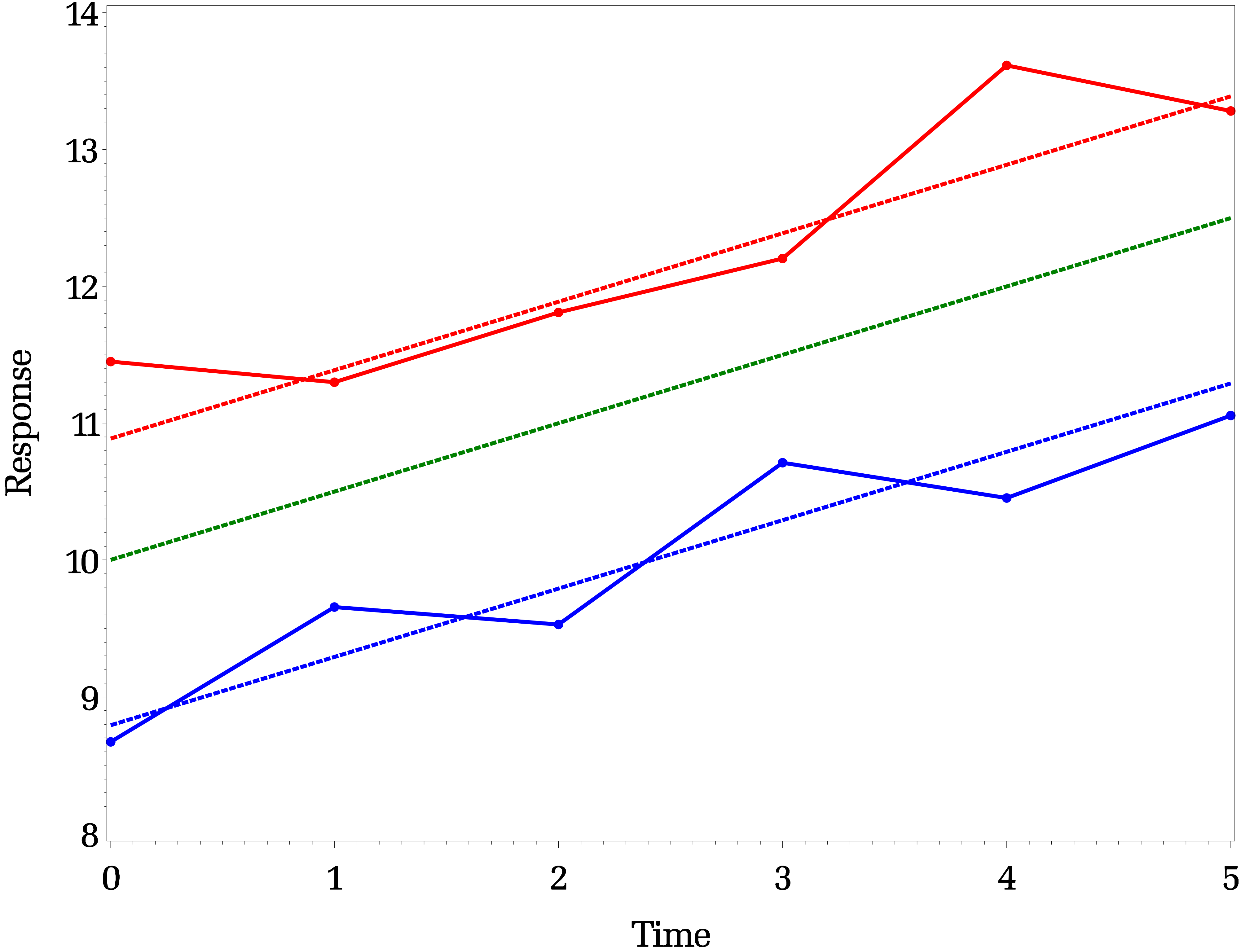

14
13
12
11
10
9
8
Response
0
1
2
3
4
5
Time



14
13
12
11
10
9
8
Response
0
1
2
3
4
5
Time



14
13
12
11
10
9
8
Response
0
1
2
3
4
5
Time



Response
Time
14
13
12
11
10
9
8
0
1
2
3
4
5

Longitudinal studies: three effects

- Longitudinal studies generally lead to *three effects* of exposure on response:
  - *within-subject* (*individual level*) *effect* of the exposure for a particular subject on this subject's mean response
  - *between-subject effect* of the mean exposure for a particular subject on mean response
  - *marginal* (*population-average*) *effect* of the exposure (whether within or between subjects) on mean response

Statistical analysis of longitudinal studies

- Distinctive feature: observations on the same subject are typically positively *correlated*, and this correlation needs to be accounted for in the statistical analysis
- Two major approaches: marginal analysis and mixed models

Statistical analysis: Marginal modeling

– regression with "working" specification of within-subject correlation structure using Generalized Estimating Equations

– equivalent to a cross-sectional analysis with (some) dependent observations

– allows consistent estimation of *population-average* effect, even if the "working"correlation structure is misspecified

– within- and between-subject effects cannot be estimated

Statistical analysis: Mixed effects models

– include both *fixed* and *random effects*

– *fixed effects* are functions of covariates which are the same for all subjects

– *random effects* are subject-specific realizations of latent random variables; account for between-subject heterogeneity and induce within-subject correlation structure

– allows estimation of all three effects but requires specification of latent random effects

<u>Linear mixed model (LMMs)</u>

- **Traditional assumption** in mixed models: random effects are independent of covariates
- In LMMs, the traditional assumption leads to all three effects being the same
- Yet, Neuhaus & Kalbfleisch (1998) empirically demonstrated that three effects could be different in LMMs
- Three exposure effects are *always* different if random effects in LMM are *correlated* with exposure (e.g., Neuhaus & McGulloch, 2006)

Linear mixed model

- Let x_{ij}, y_{ij} denote the exposure and outcome for person i, $i = 1, ..., n$, time $j = 1, ..., m_i$
- Simple linear mixed effects model

$$y_{ij} = \beta_o + \beta_x x_{ij} + u_{yi} + \epsilon_{yij}$$

- Exposure may also vary with time and be specified as

$$x_{ij} = \mu_0 + u_{xi} + \epsilon_{xij}$$

Linear mixed model

- Model:

$$y_{ij} = \beta_o + \beta_x x_{ij} + u_{yi} + \epsilon_{yij}$$

$$x_{ij} = \mu_0 + u_{xi} + \epsilon_{xij}$$

- Traditional assumption that u_{yi} is independent of x_{ij} may be too strong: both random effects u_{yi} and u_{xi} represent heterogeneity between subjects in response and exposure, respectively, and therefore may be correlated

Linear mixed model

- Correlation between u_{yi} and u_{xi} leads to linear regression

$$u_{yi} = \frac{\sigma_{u\mu}}{\sigma_{\mu}^2}(\mu_{xi} - \mu_0) + \eta_{yi}, \;\; \eta_{yi} \perp (\mu_{xi} = \mu_0 + u_{xi}, \epsilon_{xij})$$

- Then the model can be reparameterized as

$$y_{ij} = \left(\beta_o + \frac{\sigma_{u\mu}}{\sigma_{\mu}^2}\mu_0\right) + \left(\beta_x + \frac{\sigma_{u\mu}}{\sigma_{\mu}^2}\right)\mu_{xi} + \beta_x \epsilon_{xij} + \eta_{yi} + \epsilon_{yij}$$

Linear mixed model

- Generally, there are *three different effects* of x_{ij} on y_{ij}:
  - within-subject $\beta_W = \frac{cov(x_{ij}, y_{ij}|\mu_{xi})}{var(x_{ij}|\mu_{xi})} = \beta_x$
  - between-subject $\beta_B = \frac{cov(x_{ij}, y_{ij}|\epsilon_{xij})}{var(x_{ij}|\epsilon_{xij})} = \beta_x + \frac{\sigma_{u\mu}}{\sigma_\mu^2}$
  - marginal $\beta_M = \frac{cov(x_{ij}, y_{ij})}{var(x_{ij})} = \beta_x + \frac{\sigma_{u\mu}}{\sigma_x^2}$;
- Also $\beta_M = \frac{\sigma_\delta^2}{\sigma_\mu^2 + \sigma_\delta^2}\beta_W + \frac{\sigma_\mu^2}{\sigma_\mu^2 + \sigma_\delta^2}\beta_B$

Linear mixed model

- Assuming $\sigma_{u\mu} = 0$, MLE of the common effect β_x is given by

$$\widehat{\beta}_x = \left[\sum_{i=1}^{n}(X_i - \overline{X}\,)^t V^{-1}(X_i - \overline{X}\,)\right]^{-1}\left[\sum_{i=1}^{n}(X_i - \overline{X}\,)^t V^{-1}(Y_i - \overline{Y}\,)\right]$$

where $X_i = (x_{i1}...x_{im})^t$, $Y_i = (y_{i1}...y_{im})^t$, $V = \frac{1}{\sigma_\epsilon^2} var(Y_i|X_i)$

- $\widehat{\beta}_x$ converges in probability to

$$\widehat{\beta}_x \xrightarrow{\text{P}} \beta_x + \frac{\sigma_{u\mu}\mathbf{1}_m^t V^{-1}\mathbf{1}_m}{\sigma_x^2 tr(V^{-1}R_x)},\ \mathbf{1}_m = (1...1)^t, R_x = \frac{1}{\sigma_x^2} var(X_i)$$

and is generally biased for all three effects if $\sigma_{u\mu} \neq 0$

Linear mixed model

- To sum up:
  – three effects are *always* different if random effects in LMM are *correlated* with exposure
  – ignoring such correlation leads to biases in all three estimated effects
- Neuhaus et al. solution: partitioning exposure x_{ij} into two covariates – subject's mean μ_{xi} and temporal deviations ϵ_{xij} from the mean; then the slope for ϵ_{xij} represents the within-subject effect, and slope for μ_{xi} the between-subject effect.

Recent longitudinal studies with objective measures

- In three recent studies, only Ekstedt et al. used temporal deviations of PA measurements from subject-specific means
- In presence of correlation between random effect and exposure, only their analysis would consistently estimate within-subject effect, while analyses in other two studies would be biased
- Complication – measurement error in accelerometry assessment may lead to attenuated effects

Joint multivariate mixed model

- **Our approach**: joint modeling of responses and covariates
- After some *appropriate transformations*, let
  $X_{ij} = (x_{ij1}...x_{ijp})^t$ be a $(p \times 1)$ vector of time-varying covariates
  $Y_{ij} = (y_{ij1}...y_{ijq})^t$ a $(q \times 1)$ vector of responses
  $Z_i = (z_1...z_k)^t$ a $(k \times 1)$ vector of non time-varying covariates

Joint multivariate mixed model

- Joint multivariate mixed model with correlated random effects

$$
\begin{aligned}
X_{ij} &= \boldsymbol{\alpha}_{0j} + \boldsymbol{\alpha}_z Z_i + \boldsymbol{u}_{Xi} + \boldsymbol{\epsilon}_{Xij} \\
Y_{ij} &= \boldsymbol{\beta}_{0j} + \boldsymbol{\beta}_X X_{ij} + \boldsymbol{\beta}_z Z_i + \boldsymbol{u}_{Yi} + \boldsymbol{\epsilon}_{Yij}
\end{aligned}
$$

$$
\begin{aligned}
&\boldsymbol{\epsilon}_{Xij} \sim N(0,\, \Sigma_{\epsilon_X}),\, \boldsymbol{\epsilon}_{Yij} \sim N(0,\, \Sigma_{\epsilon_Y}), \\
&\boldsymbol{u}_i = (\boldsymbol{u}_{Xi}^t, \boldsymbol{u}_{Yi}^t)^t \sim N(0,\, \Sigma_u)
\end{aligned}
$$

Joint multivariate mixed model

- Joint specification of appropriately transformed time-varying outcome and exposure using linear mixed models with correlated random effects allows for:
  - correlations between random effects and covariates in the outcome model
  - modeling a *non-linear* relationship on the original scale
  - consistent estimation of all three different effects

Joint multivariate mixed model

- Joint modeling allows for specification of and adjustment for measurement errors in both exposure and outcome
- Denoting observed vectors by M_{Xij} and M_{Yij}, respectively, assume

$$M_{Xij} = X_{ij} + \nu_{Xij}$$
$$M_{Yij} = Y_{ij} + \nu_{Yij}$$

where $\nu_{ij} = \left(\nu^t_{Xij}, \nu^t_{Yij}\right)^t \sim N(0,\, \Sigma_\nu)$

- Model may be fitted using SAS proc CALIS

BodyMedia study

- BodyMedia Inc. FIT is a multi-sensor 3-axis accelerometer designed to be worn 24 hours/day
- For 935 men and 3647 women, analysis included *log-transformed* daily minutes of moderate to vigorous PA (MVPA) as well as sleep and lying awake minutes averaged over a week for 12 consecutive weeks, along with information on age, height, and weight
- Goal: estimate within-subject effect of MVPA on sleep characteristics, including sleep efficiency

Application to BodyMedia

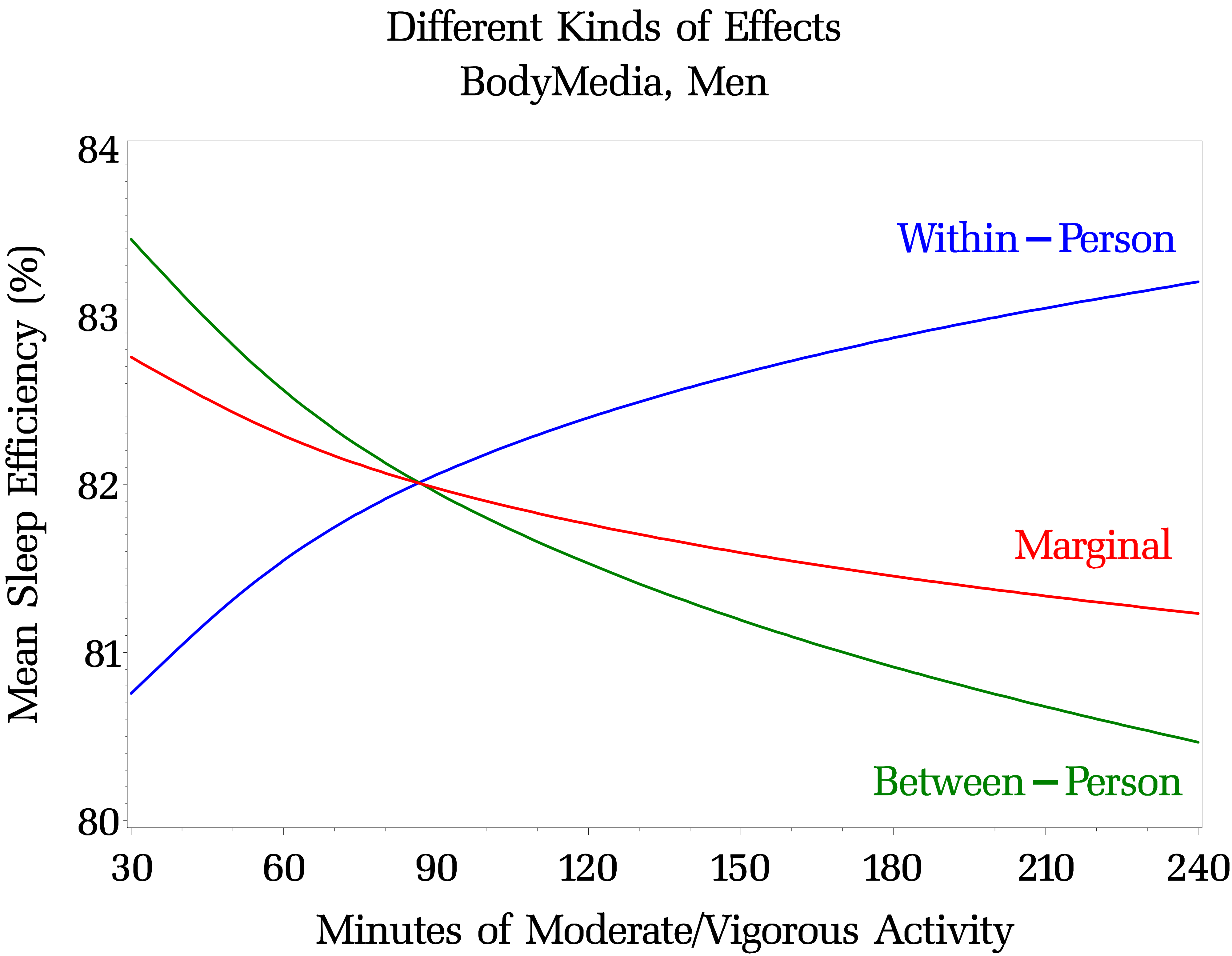
- In our application, X_{ij} is log MVPA, $Y_{ij} = (y_{ij1}, y_{ij2})^t$ where y_{ij1} is log sleep minutes and y_{ij2} is log lying awake minutes, $\mathrm{Z}_i = (z_{i1}, z_{i2})^t$ where $z_{i1}$is age and z_{i2} is log BMI

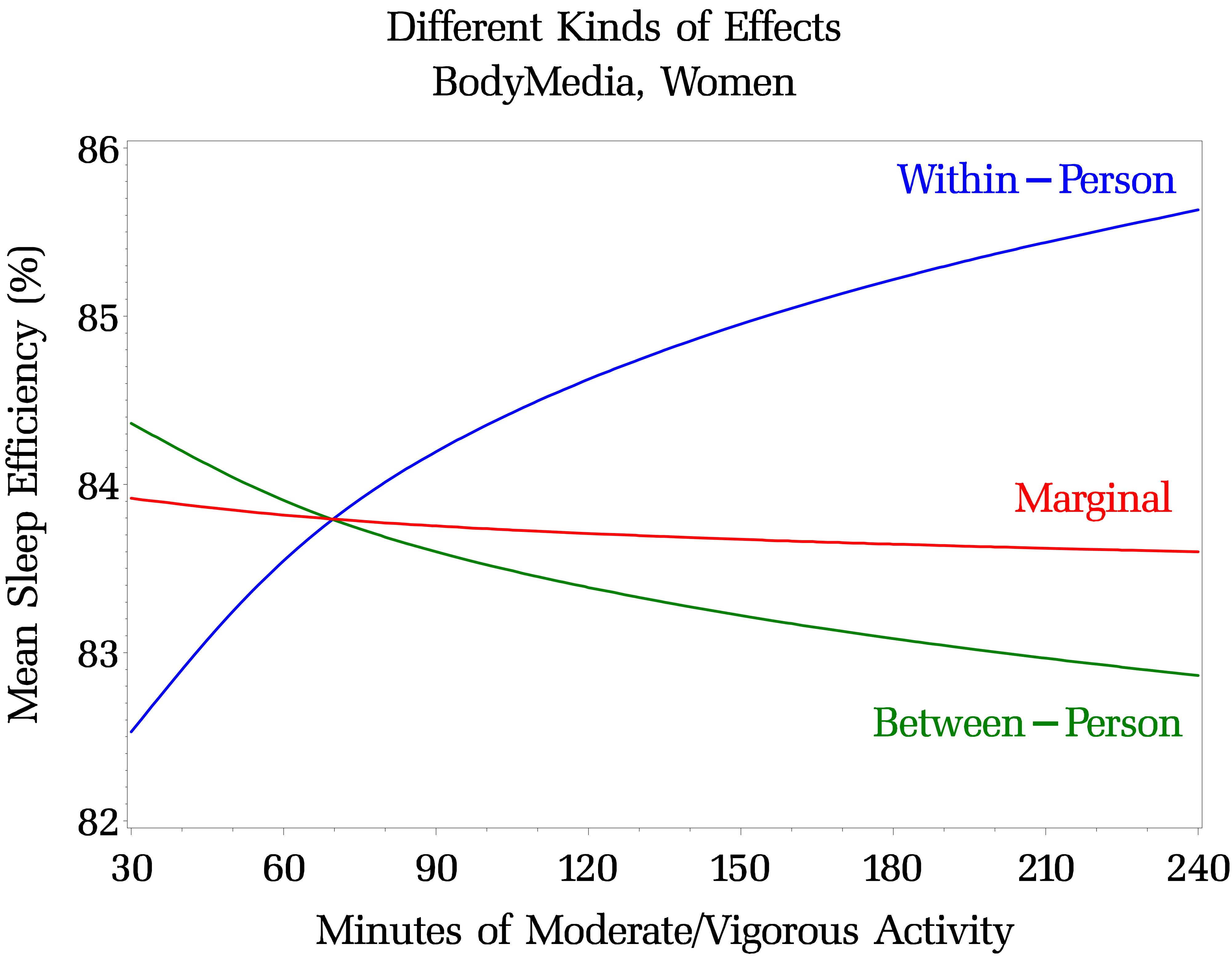
- $\boldsymbol{\epsilon}_{Xij}$ and $\boldsymbol{\epsilon}_{Yij}$ are specified as first order autoregressions, e.g.

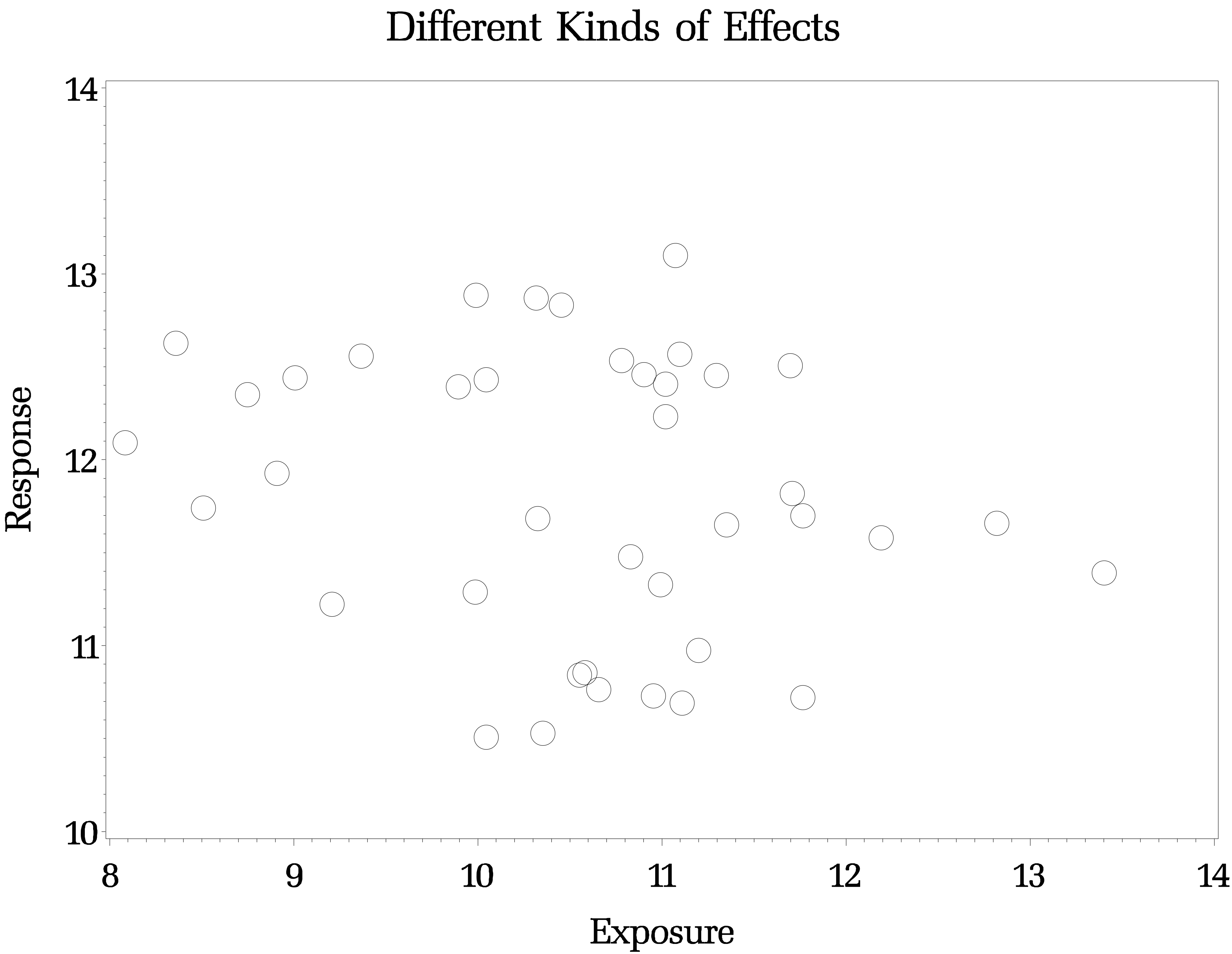
$$\boldsymbol{\epsilon}_{Yij} = \boldsymbol{\Gamma}\boldsymbol{\epsilon}_{Yij-1} + \boldsymbol{\phi}_{Yij},\, \boldsymbol{\phi}_{Yij} \sim N(0,\, \Sigma_{\phi})$$

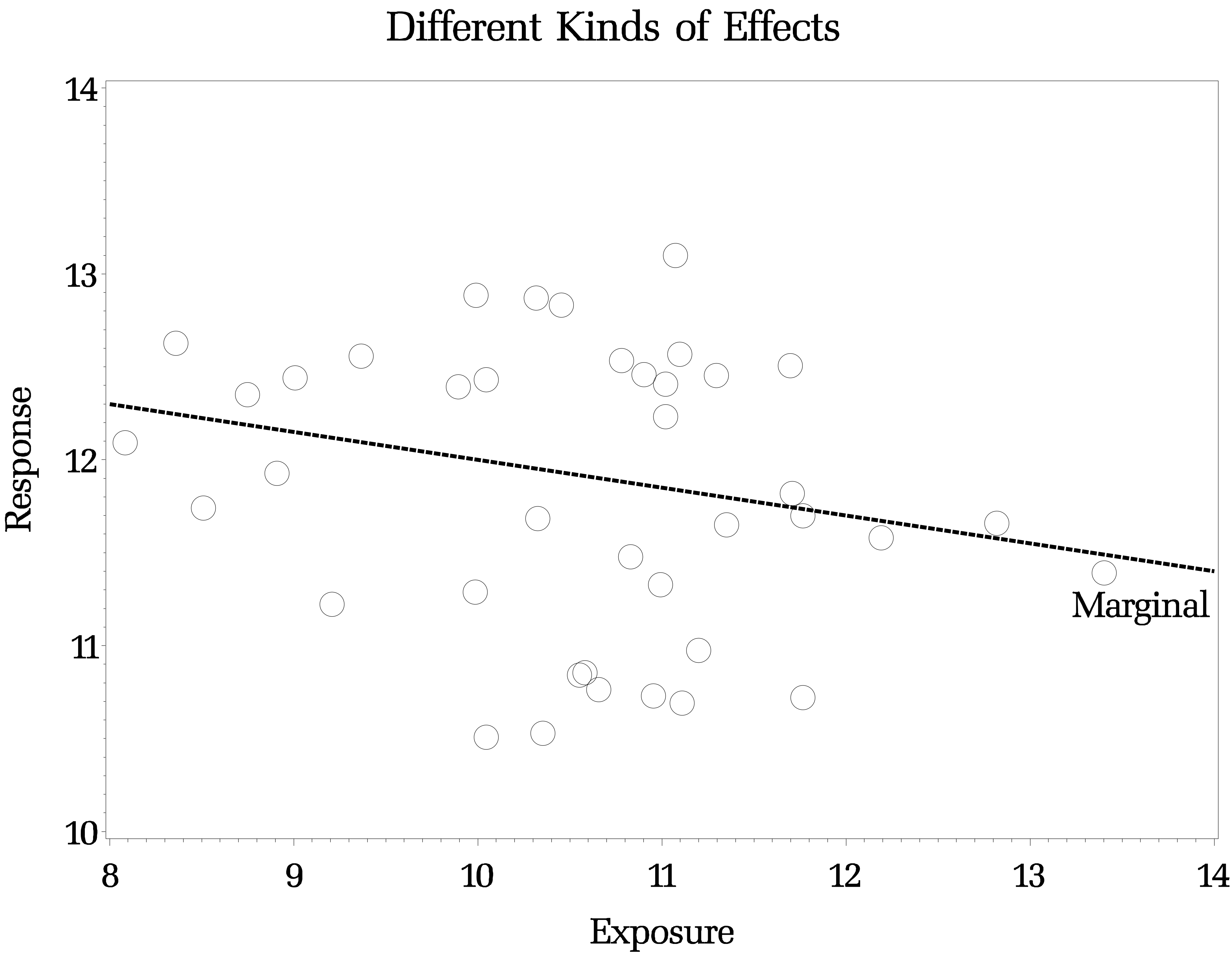
- Measurement error VAR is specified as $\Sigma_{\nu} = \begin{pmatrix} \Sigma_{\nu_X} & 0 \\ 0 & \Sigma_{\nu_Y} \end{pmatrix}$

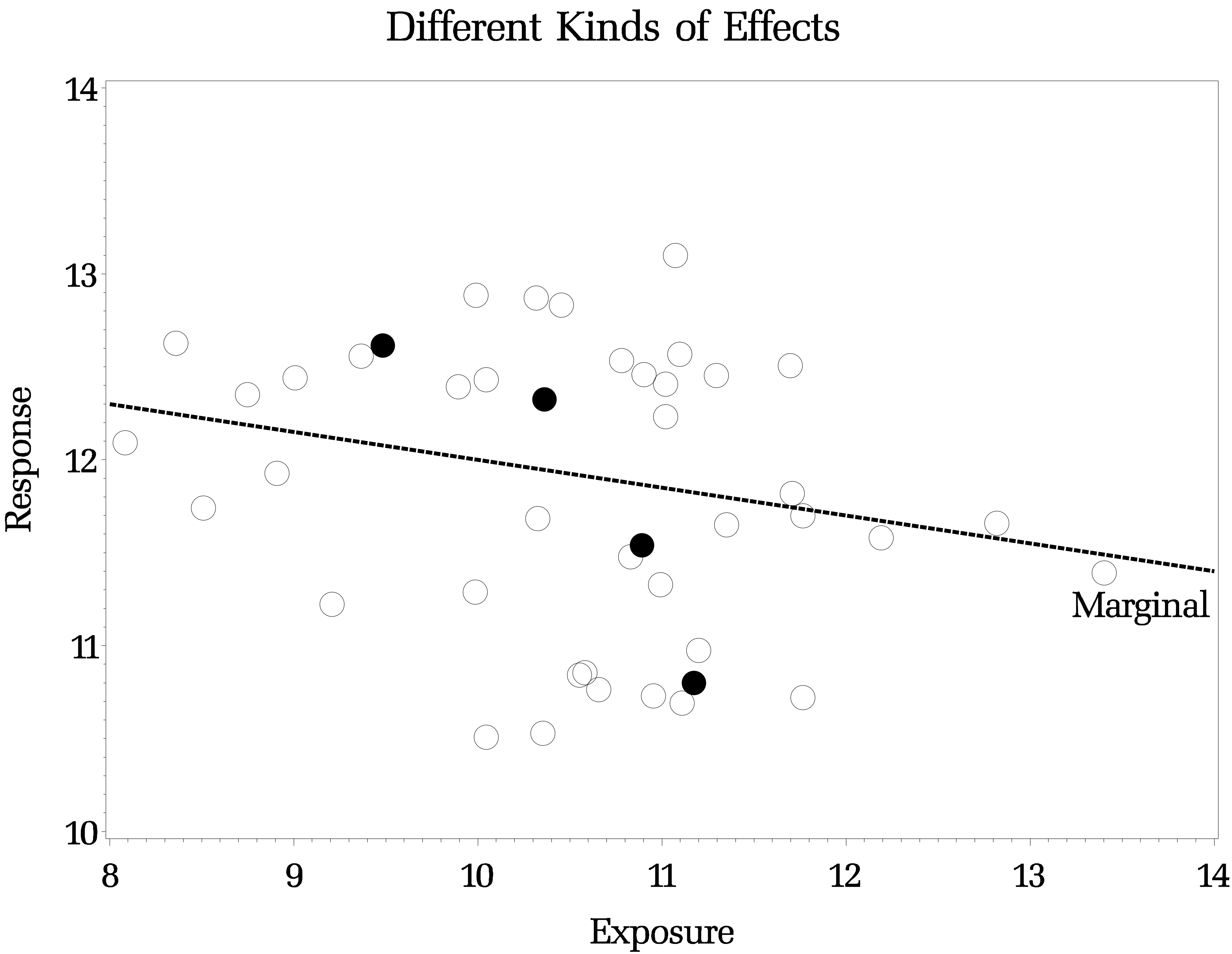
- Goal: estimate different effects of X_{ij} on sleep efficiency $e^{y_{ij1}}/(e^{y_{ij1}} + e^{y_{ij2}})$


Different Kinds of Effects
BodyMedia, Men
Mean Sleep Efficiency (%)
84
83
82
81
80
30
60
90
120
150
180
210
240
Minutes of Moderate/Vigorous Activity
Within−Person
Marginal
Between−Person

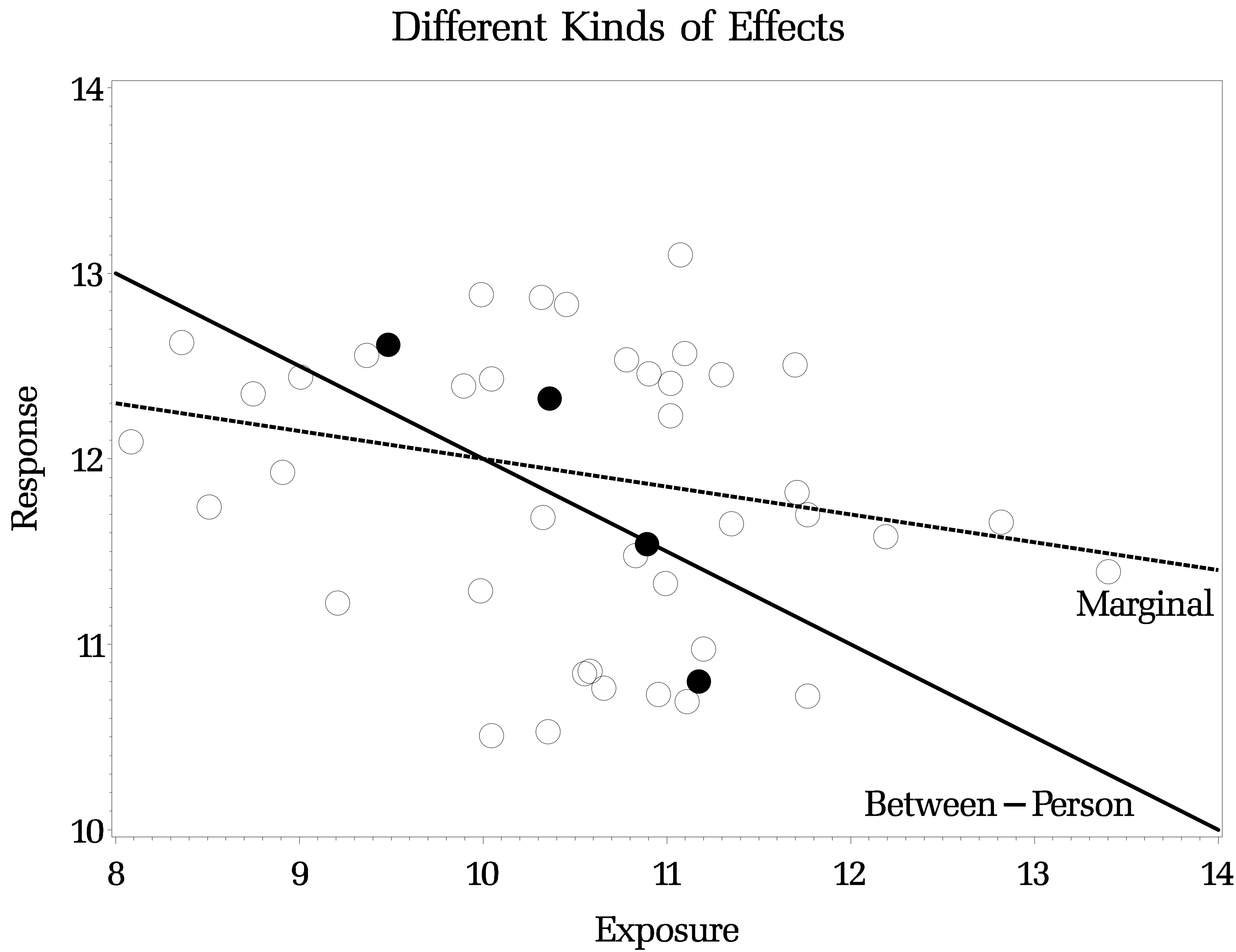

Different Kinds of Effects
BodyMedia, Women
Within−Person
Marginal
Between−Person
Mean Sleep Efficiency (%)
82
83
84
85
86
30
60
90
120
150
180
210
240
Minutes of Moderate/Vigorous Activity


Different Kinds of Effects
Response
Exposure
14
13
12
11
10
8
9
10
11
12
13
14

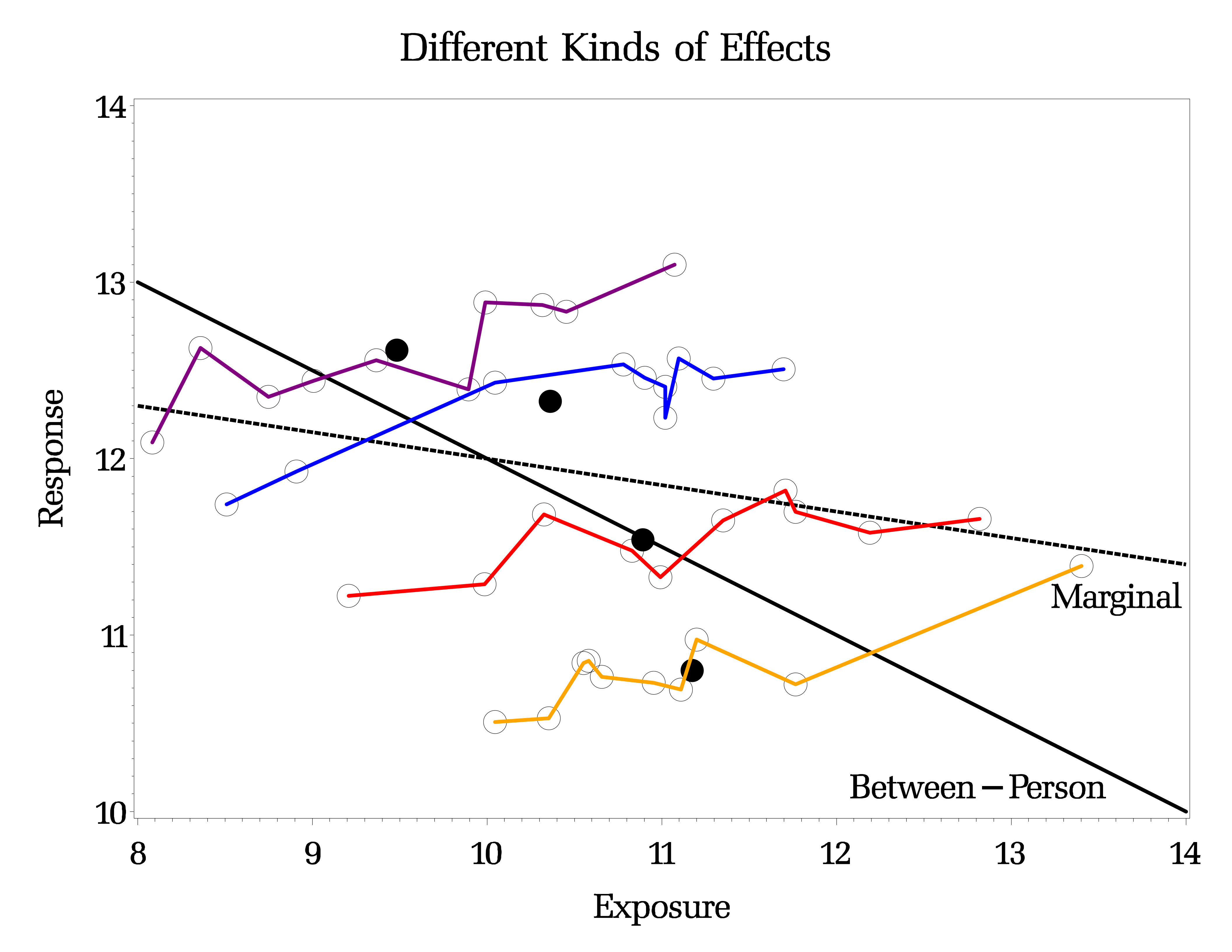

Different Kinds of Effects
Response
Exposure
Marginal
14
13
12
11
10
8
9
10
11
12
13
14

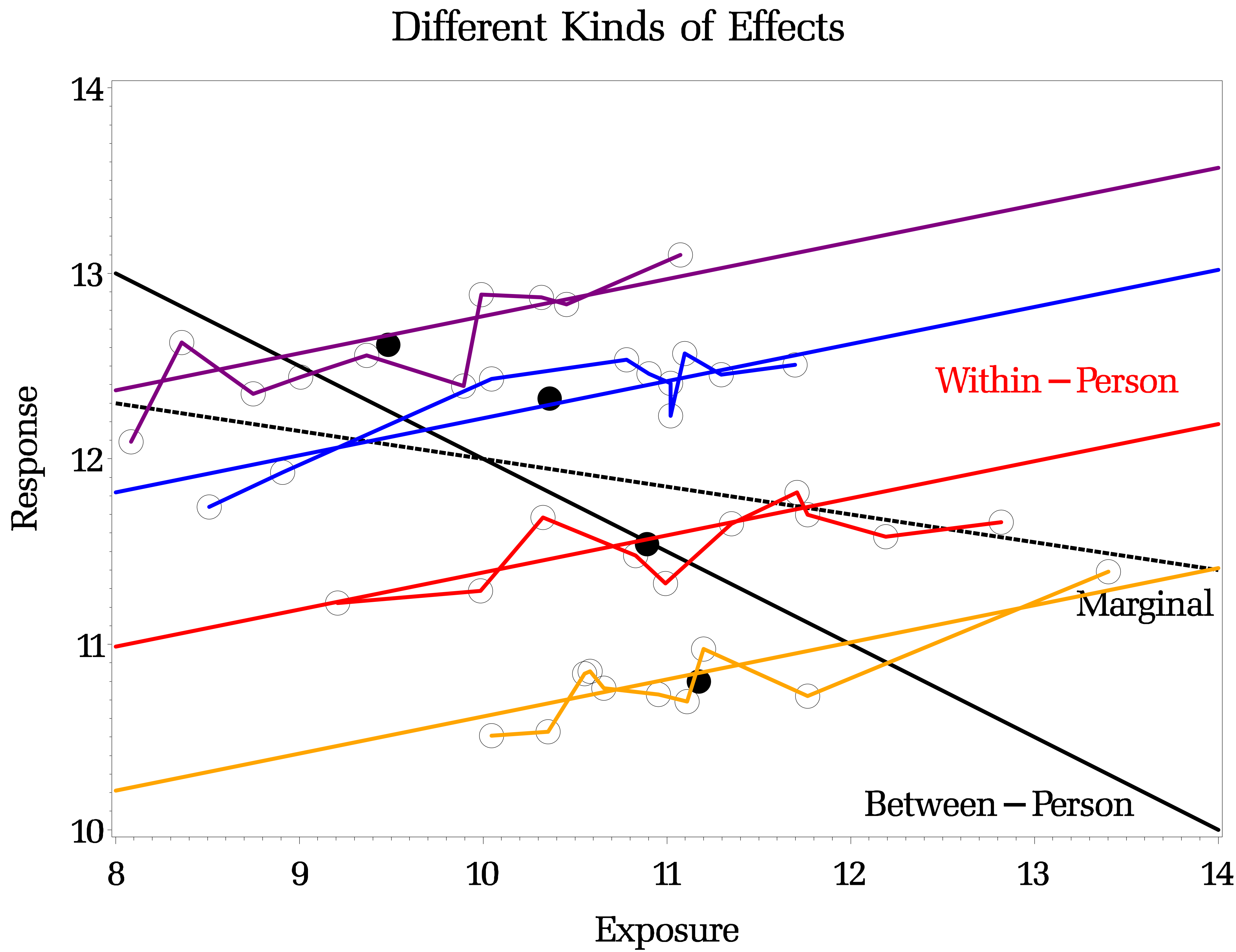


Different Kinds of Effects
Response
Exposure
Marginal
14
13
12
11
10
8
9
10
11
12
13
14

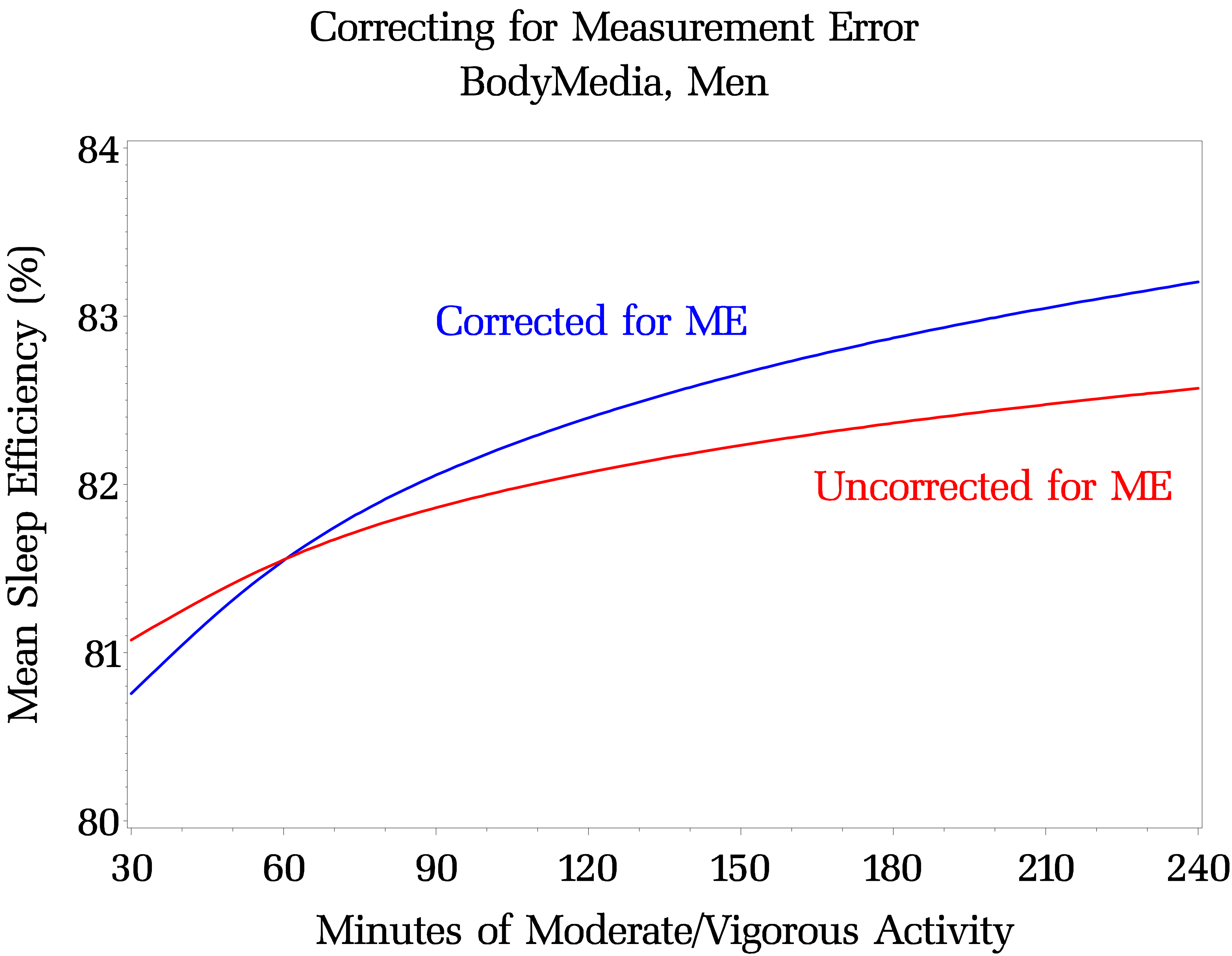


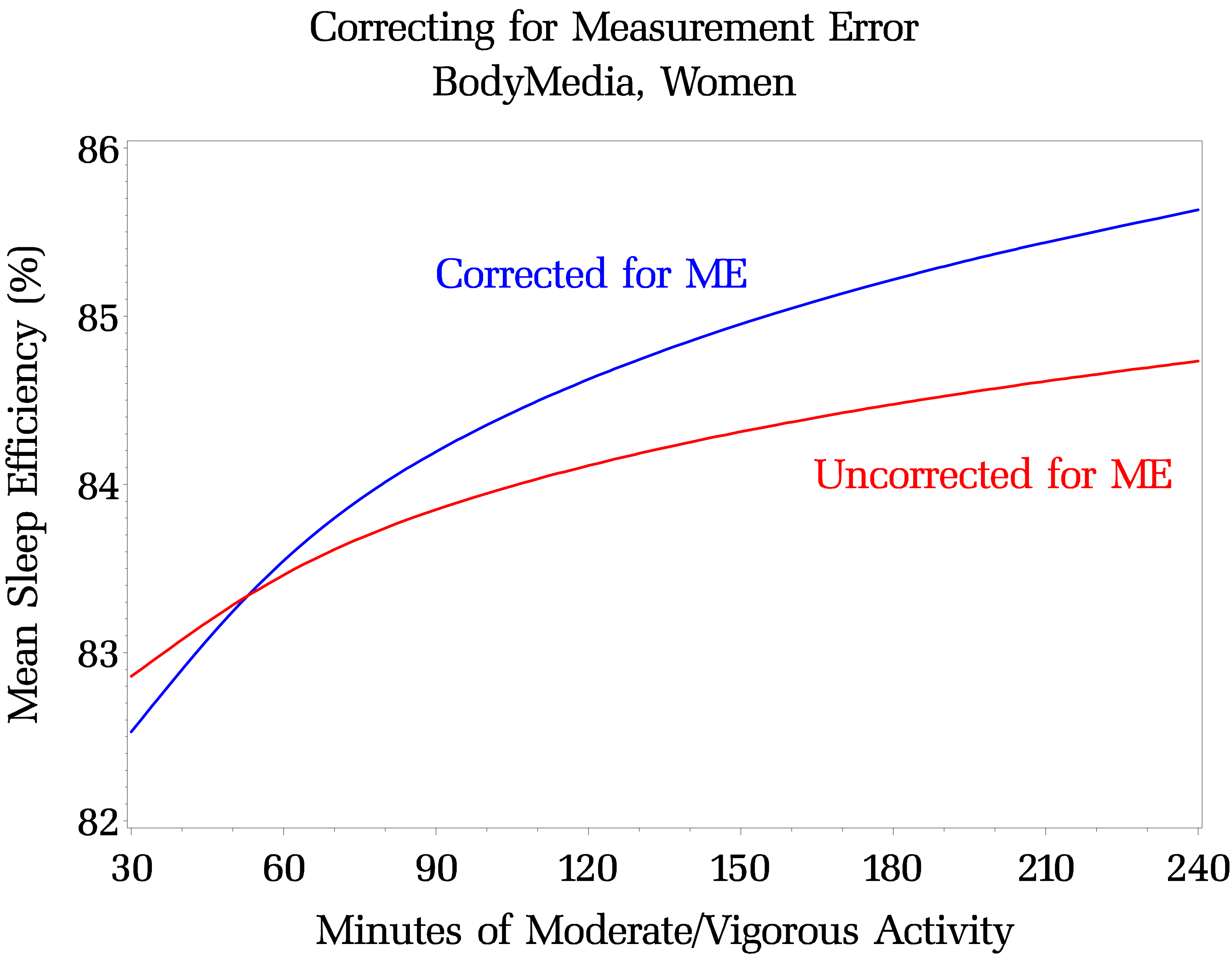
Different Kinds of Effects
Response
Exposure
14
13
12
11
10
8
9
10
11
12
13
14
Marginal
Between – Person


Different Kinds of Effects
Response
Exposure
14
13
12
11
10
8
9
10
11
12
13
14
Marginal
Between – Person



Different Kinds of Effects
Within – Person
Marginal
Between – Person
Response
Exposure
14
13
12
11
10
8
9
10
11
12
13
14


Correcting for Measurement Error
BodyMedia, Men
Mean Sleep Efficiency (%)
84
83
82
81
80
Corrected for ME
Uncorrected for ME
30
60
90
120
150
180
210
240
Minutes of Moderate/Vigorous Activity


Correcting for Measurement Error
BodyMedia, Women
Corrected for ME
Uncorrected for ME
Mean Sleep Efficiency (%)
86
85
84
83
82
30
60
90
120
150
180
210
240
Minutes of Moderate/Vigorous Activity

Simulation study

- To evaluate performance of the suggested model in finite samples, we performed a simulation study based on BodyMedia results for men
- For each set of simulations, we simulated 1000 data sets with 1000 subjects each, varying the correlation between random effects and the number of repeat observations

Simulation Results
Within-Person Effect of MVPA Minutes
on Sleep Efficiency
Each subject has 6 observations

Sim	Corr[*]	True Effect	Correlated Random Effects	Uncorrelated Random Effects
1	-0.1	0.085	0.085 (0.001)	0.051 (0.001)
2	-0.2	0.085	0.085 (0.001)	0.018 (0.001)
3	-0.4	0.085	0.086 (0.001)	-0.063 (0.001)

[*] Corr = correlation of random effects for log MVPA and logit sleep efficiency

Discussion

- Although traditionally ignored in mixed models, possible correlations between random effects and exposure should be accounted for to consistently estimate three different effects
- Joint modeling of time-varying exposure and response as random variables, instead of commonly specifying responses conditional on the exposures, allows explicit specification of those correlations
- Another benefit of joint modeling is explicit specification of and adjustment for measurement error in both exposure and outcome