

Categorizing variables measured with error

Subgroup of TG4: Measurement error and misclassification



Background

- Categorization of food consumption often done in nutritional studies
- Reason: to study shape / no assumption on shape
- Categorization in general has disadvantages
 - Loss of information / power
 - Can be manipulated
 - Misspecified model (in presence confounders)
 - Inferior to splines

Plans of subgroup

- Paper explaining effect of categorizing variable measured with errors
- What to do (feasible for researchers):
 - Regression calibration → categorize calibrated values (using pre-specified cut-offs)
 - Simple correction factor for when categories are quantiles:
Divide by correlation coefficient true/observed

Use calibrated values

calibrated value = expected value of true intake:

$$C_i = E(X_i | X_i^*)$$

C_i = "calibrated value"

X_i = true value

X_i^* = measured value with imperfect instrument

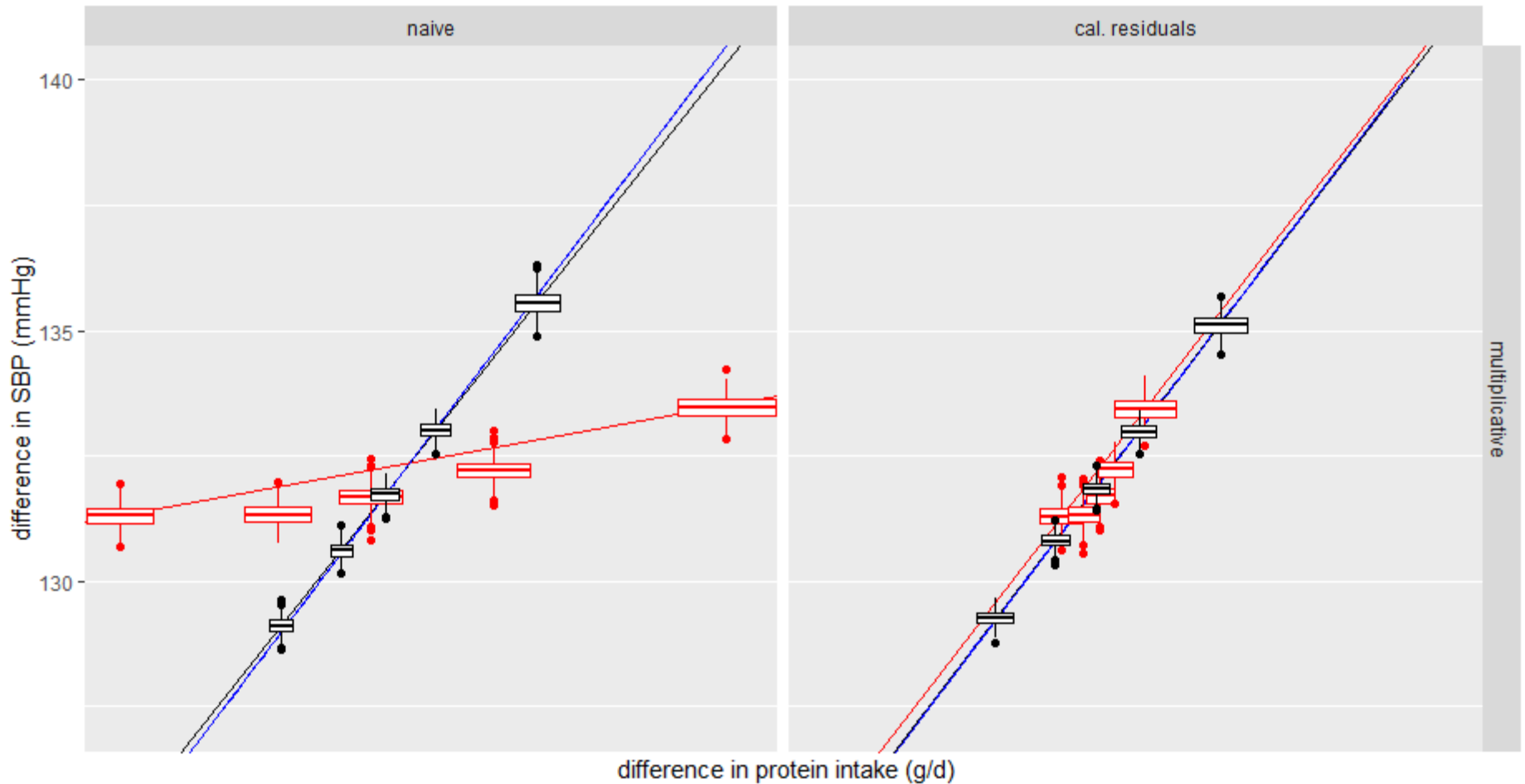
Proposal based on:

$$E_{C_i \in (X_{low}, X_{high}]}(C_i) = E_{C_i \in (X_{low}, X_{high}]}(E(X_i)) \stackrel{?}{=} E_{X_i \in (X_{low}, X_{high}]}(X_i)$$

Results simulation

x=lognormal , strong association , pos. confounding

blue: simulated relation; black: no error; red = with error

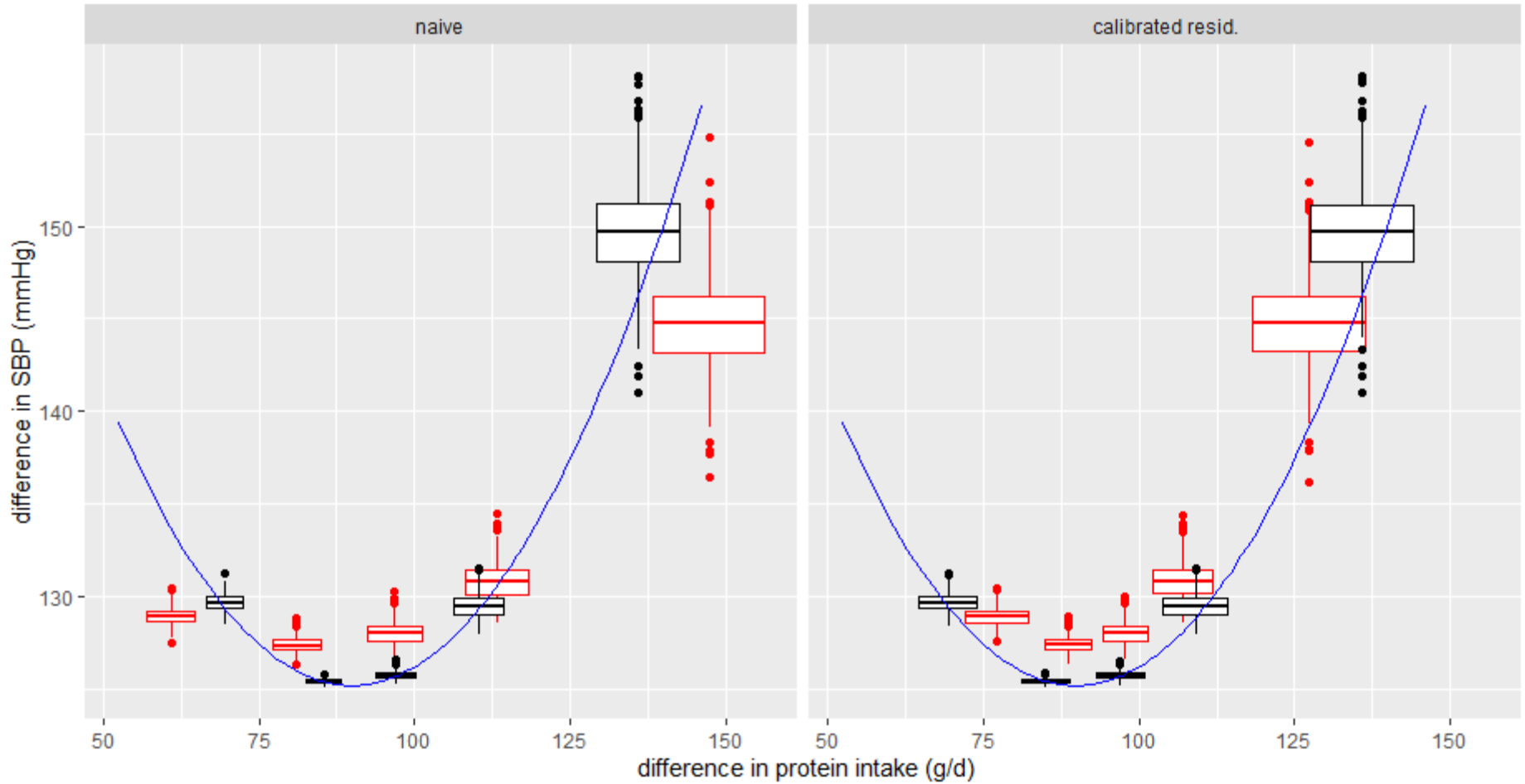


- Ground truth relation
- Relation from error free variable
- Variable error-free
- Variable with error

Results simulation

x=lognormal , strong association , no confounding

blue: simulated relation; black: no error; red = with error



— Ground truth relation

■ Variable error-free

■ Variable with error

Possible collaboration TG4

- Splines: TG2 (Selection of variables and functional forms in multivariable analysis)
- Categorizing with confounders → misspecified model : TG2?
- Include validation studies, sample size calculation → TG5 study design
- Measurement error → missing true values → TG1 missing values
- Prediction models → Evaluating diagnostic tests and prediction models