

Simulated dataset #2

A sample size of 500 was generated with no missing values.

Chemical concentrations were generated using the estimated correlations from 14 log-transformed PCBs, dioxins and furans using NHANES data with observed means and standard deviations. The covariates were based on the poverty index ratio (z1), age (z2), and sex (Z3); their correlations with the chemicals were used in the data generation. The response variable, Y, was generated to have the same SD as estimated from log(ALT).

The mean model for Y was assumed to be different across sex.

For z3=0, the mean model was:

$$\mu = 3 + 0.05*x4 + 0.1*x6 + 0.1*x11 + 0.5*x12 + 0.1*x14 + 0.01*z1 + 0.003*z2 ;$$

for z3=1, the mean model was:

$$\mu = 3 + 0.01*x1 + 0.05*x4 + 0.1*x11 + 0.1*x14 + 0.01*z1 + 0.003*z2 - 0.32*(z3=1).$$

NOTE: the x terms in these models are generated data on the log scale.

Thus, for z3=0 only x4, x6, x11, x12 and x14 are positively associated with Y. When z3=1, only x1, x4, x11 and x14 are associated with Y. Interactions were not assumed among the chemical concentrations.