University of Zurich[UZH]

Biostatistics

Multicollinearity

Jerome Sepin

April 14, 2023

Why this presentation?

From Schwabe et al. (2022):

"The primary pharmacokinetic parameters were log-transformed and analyzed using an analysis of covariance (ANCOVA) model that included treatment as a fixed effect, baseline weight category ($\geq 60- \leq 80kg / > 80kg- \leq 100kg$), baseline BMI (continuous variable), and study site as covariates."

What is the problem with this?

$$\text{BMI} = \frac{\text{weight (kg)}}{\text{height (m)}^2}$$

Relation of variables

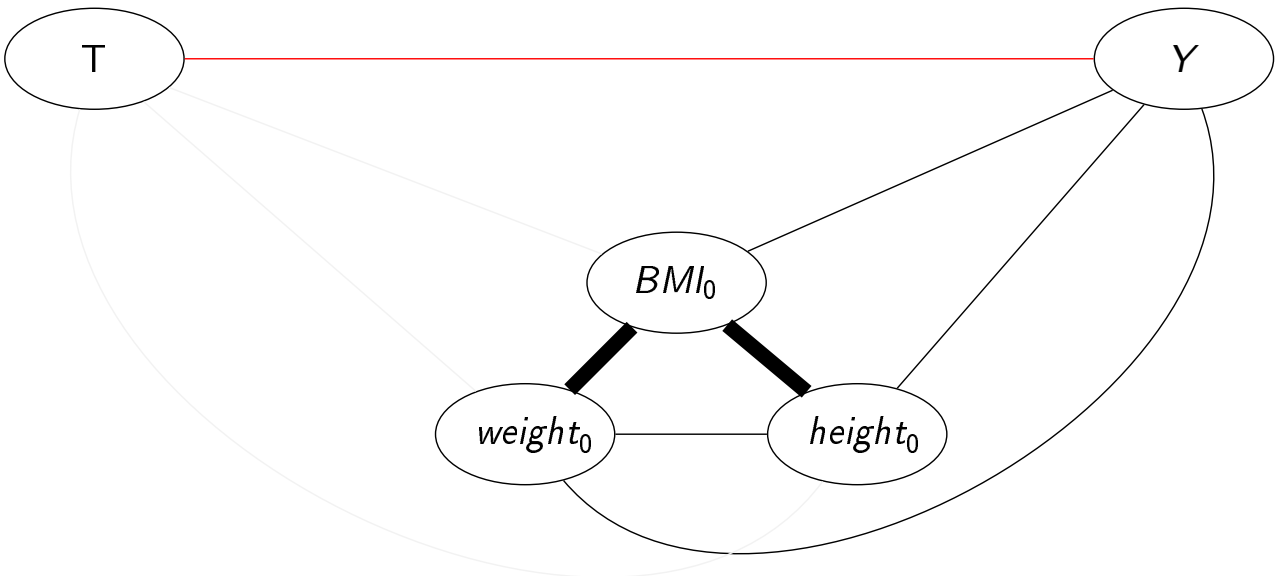


Figure: Successful randomization removes confounding

$$\boxed{\text{BMI} = \frac{\text{weight (kg)}}{\text{height (m)}^2}}$$

But it may not be successful...

ICH E9: *"In multicentre trials (see Glossary) the randomisation procedures should be organised centrally. It is advisable to have a separate random scheme for each centre, i.e. to stratify by centre or to allocate several whole blocks to each centre. More generally, stratification by important prognostic factors measured at baseline (e.g. severity of disease, age, sex, etc.) may sometimes be valuable in order to promote balanced allocation within strata; this has greater potential benefit in small trials. The use of more than two or three stratification factors is rarely necessary, is less successful at achieving balance and is logistically troublesome. The use of a dynamic allocation procedure (see below) may help to achieve balance across a number of stratification factors simultaneously provided the rest of the trial procedures can be adjusted to accommodate an approach of this type.*

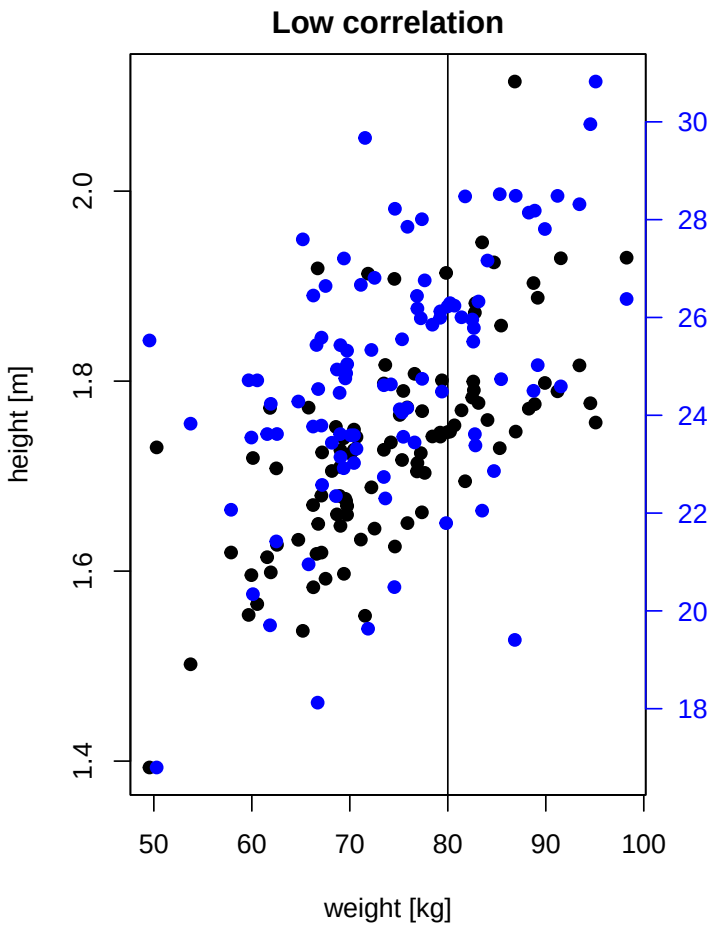
Factors on which randomisation has been stratified should be accounted for later in the analysis. "

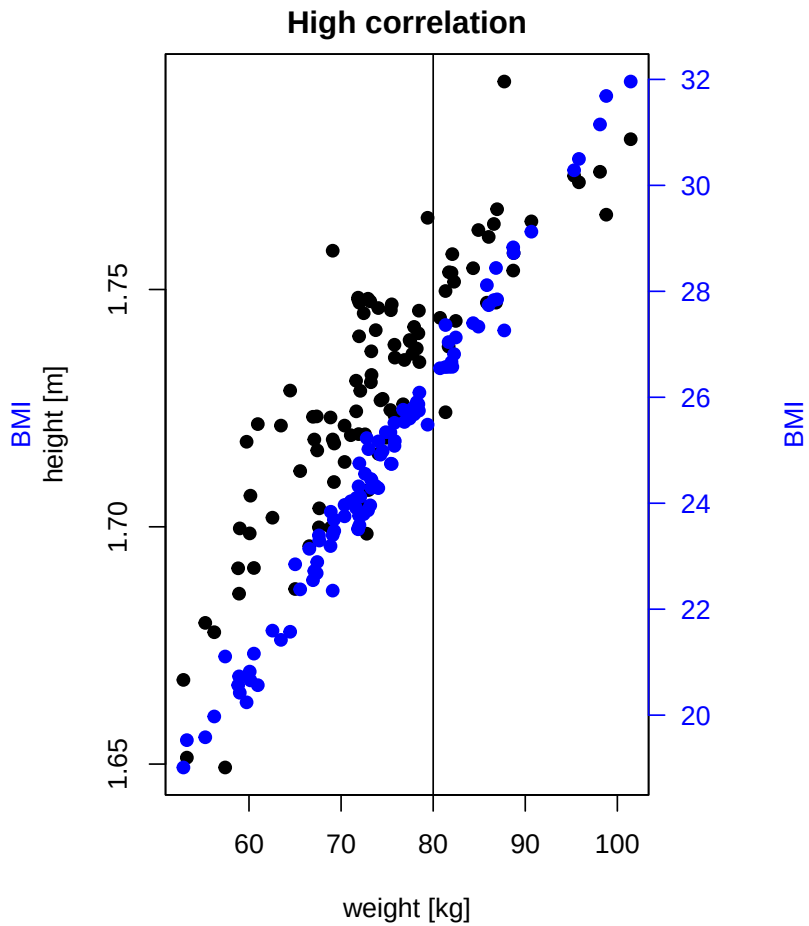
Relation of variables

1. paper: $AUC \sim trt + weight_{cat.} + BMI_{cont.} + site$
   $AUC \sim trt + weight_{cat.} + \frac{weight_{cont.}}{weight^2_{cont.}} + site$
2. a1: $AUC \sim trt + weight_{cat.} + height_{cont.} + site$
3. a2: $AUC \sim trt + weight_{cont.} + height_{cont.} + site$

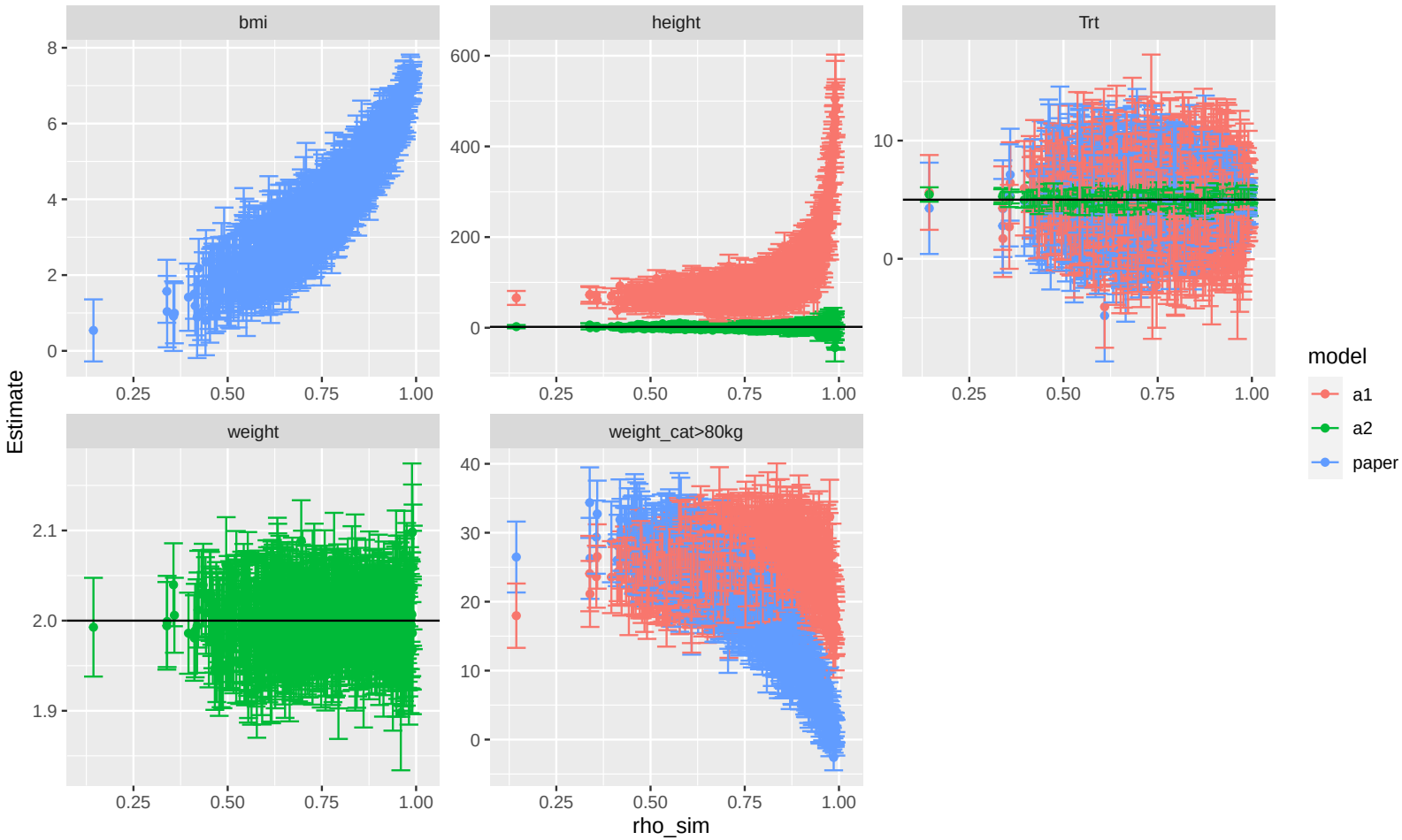
Let's see what happens when the correlation between *height* and *weight* changes!

Relation of variables





Results


bmi
height
Trt
weight
weight_cat>80kg
Estimate
rho_sim
model
a1
a2
paper

Conclusion

- Collinearity increases uncertainty (only the variables that are included)
- The closer we model "nature", the lower our residual standard error -> more power! (avoid dichotomization)
- A bit more sophisticated randomization may not hurt ...

References

Schwabe, C., Illes, A., Ullmann, M., Ghori, V., Vincent, E., Petit-Frere, C., Monnet, J., Racault, A. S., and Wynne, C. (2022). Pharmacokinetics and pharmacodynamics of a proposed tocilizumab biosimilar MSB11456 versus both the US-licensed and EU-approved products: a randomized, double-blind trial. *Expert Review of Clinical Immunology*, 18(5):533–543.