Please join the virtual Biometric Colloquium

LORENA HAFERMANN

Charité – Universitätsmedizin Berlin, Germany

STATISTICAL MODEL BUILDING: BACKGROUND "KNOWLEDGE" BASED ON INAPPROPRIATE PRESELECTION CAUSES MISSPECIFICATION

November 17th, 2021 at 9:00 am

JOIN WEBEX MEETING
https://meduniwien.webex.com/meduniwien/j.php?MTID=mab6618e8724e128f7223c37d4e9288ed

HOSTS: Georg Heinze

ABSTRACT

Background: Statistical model building requires selection of variables for a model depending on the model's aim. In descriptive and explanatory models, a common recommendation often met in the literature is to include all variables in the model which are assumed or known to be associated with the outcome independent of their identification with data driven selection procedures. An open question is, how reliable this assumed "background knowledge" truly is. In fact, "known" predictors might be findings from preceding studies which may also have employed inappropriate model building strategies.
Methods: We conducted a simulation study assessing the influence of treating variables as "known predictors" in model building when in fact this knowledge resulting from preceding studies might be insufficient. Within randomly generated preceding study data sets, model building with variable selection was conducted. A variable was subsequently considered as a "known" predictor if a predefined number of preceding studies identified it as relevant.
Results: Even if several preceding studies identified a variable as a "true" predictor, this classification is often false positive. Moreover, variables not identified might still be truly predictive. This especially holds true if the preceding studies employed inappropriate selection methods such as univariable selection.
Conclusions: The source of "background knowledge" should be evaluated with care. Knowledge generated on preceding studies can cause misspecification.