

"Qualitätssicherung statistischer Signifikanzaussagen durch Ergebnisvalidierung" (Diplomarbeit, Moritz Heinz)

Background: Statistical analysis of large biomedical data bases, such as the Sylvania Lawry Centre, even when formally correct, is prone to produce a substantial number of erroneous conclusions. This is because frequently a large number of analyses are performed on the same data set before one focuses on a set of selected hypotheses and/or models, in particular if computer power is cheap and the data are expensive or time consuming to collect.

Problem: The above described multiple comparison approaches are usually not capable of identifying valid conclusions, a fact that is frequently not emphasized in corresponding scientific publications. On the other hand, if erroneous conclusions are drawn and published from such large data bases, other groups of scientists will typically not have enough power to prove them wrong based on single clinical trials.

Aim: The aim of this diploma thesis is to establish a sound statistical background to the validation procedure of the Sylvania Lawry Centre. Main issues include random splitting of the dataset without stratification, model selection, loss of power by not using the whole database for the analysis.

Rationale: When performing a meta-analysis based on a large number of pooled trials, typically more hypotheses are being tested than are reported. Especially when graphical examination of the data is followed by statistical hypothesis tests, the significance level is not retained.