Design and analysis of biomedical studies

Biomedicine is a field that has great influence on the majority of mankind. The constant development has considerably changed our way of life during the last centuries. This has been achieved through the dedication of biomedical researchers along with the tremendous resources that over time have been allocated this field. It is utterly important to utilize these resources responsibly and efficiently by constantly striving to ensure high-quality biomedical studies. This involves the use of a sound statistical methodology regarding both the design and analysis of biomedical studies. The focus of this project is on statistical aspects that arise within the field of biomedicine.

Two types of errors are frequently accentuated within the framework of statistics, namely type I and type II errors. Type I errors occur when a null hypothesis erroneously is rejected. An acceptable type I error rate is specified prior to conducting the statistical analysis. However, all statistical models make assumptions and if violated the actual type I error rate may deviate from the pre-specified type I error rate. Type II errors occur when we fail to reject a false null hypothesis. On contrary to the type I error rate, the type II error rate is not explicitly specified during the statistical analysis and this entails that assessment of the type II error rate in practice is at risk of being neglected altogether. Concerns regarding type I errors, type II errors and adherence (or lack thereof) to model assumptions for biomedical studies are a recurring theme in this thesis.

Data collected in some biomedical studies are positively skewed; hence methods relying on the normal distribution are not directly applicable. We investigated how data from one of these studies are suitably analyzed. We extracted 23 different summary statistics from data gathered from eleven studies. The degree of adherence to the model assumptions evaluated for each of these summary statistics form basis for our conclusions. Hierarchically structured data are frequently encountered in biomedical studies. For one type of studies entailing such data we have conducted a literature study strongly indicating that this structure commonly is neglected in the statistical analysis. Based on this closed-form expressions for the approximate type I error rate are formulated. The type I error rates are assessed for a number of factor combinations as they appear in practice and in all cases the type I error rates are demonstrated to be severely inflated.

Prior to conducting a study it is important to perform power and sample size determinations to ensure that reliable conclusions can be drawn from the statistical analysis. We have formulated closed-form expressions for the statistical power of studies with a hierarchical structure to guide biomedical researchers designing future studies of this type. Upon model fitting it is important to examine if the model assumptions are met to avoid that spurious conclusions are drawn. While the range of diagnostic methods is extensive for models assuming a normal response it is generally more limited for non-normal models. An R package providing diagnostic tools suitable for examining the validity of binomial regression models have been developed. The binom Tools package is publicly available at the CRAN repository.