Abstract

A frequently addressed question in clinical trials is the comparison of different populations with regard to their response, as for example when comparing different treatment groups or clinically relevant subgroups, such as different gender, age classes, geographic regions, or grades of disease severity. Commonly used statistical tests, as for example the well-known t-test, typically address the question whether there is a significant difference between groups. However, there are numerous situations where the goal is actually the reverse, i.e. to demonstrate similarity rather than a difference.

Such situations typically arise during drug development and clinical trials. Considering for example pediatric research, studies involving children are difficult to conduct, due to ethical and safety reasons. One solution to this is the extrapolation of pediatric data from adult data if adults and children are proven to behave sufficiently similar regarding the particular disease. Another example, where assessing similarity between populations becomes a matter, is the approval process for generics, i.e. drugs, which are aiming to have “the same” treatment effect as the original product at a much lower cost. There, bioequivalence has to be demonstrated, which means proving similarity of the two products in terms of efficacy and safety. While classical statistical tests like the t-test can be used to determine if there is a significant difference between two groups, it is not able to provide evidence of the absence of an effect in case of a non-significant result. A flexible solution for this problem is equivalence testing, providing the opportunity to test whether two groups vary not more than a pre-specified threshold, and hence, are sufficiently similar. This talk introduces the concept of equivalence testing, illustrated by some of the numerous application areas. Statistical theory will be derived, including both asymptotic theory and a testing approach based on a parametric bootstrap procedure, particularly providing an alternative for trials where sample sizes are small and asymptotic methods fail.