On testing simultaneously non-inferiority in two multiple primary endpoints and superiority in at least one of them.

Röhmel J1, Gerlinger C2, Benda N2, Läuter J3

1Institute for Biometry and Clinical Epidemiology, Charité University Medicine, Berlin, Germany
2Schering AG, D-13342 Berlin, Germany
3Otto-von-Guericke University Magdeburg
Joachim.roehmel@t-online.de

In a clinical trial with an active treatment and a placebo the situation may occur that two (or even more) primary endpoints may be necessary to describe the active treatment’s benefit. We were interested in a more specific situation in so far as superiority in one of the primary endpoints would suffice given that non-inferiority is observed in the remaining. Several proposals exist in the literature for dealing with this or similar problems, but prove insufficient or inadequate at a closer look (e.g. [1], [2], [8], [9]). We propose a hierarchical three step procedure, where non-inferiority in both variables is the aim in the first step, overall tests for superiority ([4], [6], [3], [5]) or a bootstrap procedure based on ideas presented [5] build the second step, and (for the case of two primary endpoints) two separate superiority tests are performed in the third step. All statistical tests are conducted at the same one-sided significance level alpha. From the above mentioned overall superiority tests we preferred the SS test from [5] or adjustments according to [4] for the reason that these have been proven to strictly control the type I error. A simulation study reveals that the performance regarding power of the overall test depends to a considerable degree on the correlation and on the magnitude of the expected effects of the two primary endpoints. Therefore, the recommendation which test to choose depends on knowledge of the possible correlation between the two primary endpoints. In general, procedures based on [5] in step 2 shows the best overall properties, whereas the procedure based on [4] shows an advantage if both, a positive correlation between the two variables and a considerable difference between their standardized effect sizes can be expected.

References