Chapter 9
Noninferiority Testing

Just like equivalence studies noninferiority studies are very popular in modern clinical research with many treatments at hand and new compounds being mostly only slightly different from the old ones. Unlike equivalence studies (Chap. 6), noninferiority studies have a single boundary, instead of two boundaries, with an interval of equivalence in between. Noninferiority studies have been criticized for their wide margin of inferiority making it virtually impossible to reject noninferiority.

As an example, two parallel-groups of patients with rheumatoid arthritis are treated with either a standard or a new nonsteroidal anti-inflammatory drug (NSAID). The reduction of gamma globuline levels (g/l) after treatment is used as the primary estimate of treatment success. The underneath three steps constitute an adequate procedure for noninferiority analysis.

Step 1: Determination of the Margin of Noninferiority, the Required Sample, and the Expected p-Value and Power of the Study Result

1. The left boundaries of the 95% confidence intervals of previously published studies of the standard NSAID versus various alternative NSAIDS were never lower than -8 g/l. And, so, the margin was set at -8 g/l.
2. Based on a pilot-study with the novel compound the expected mean difference was 0 g/l with an expected standard deviation of 32 g/l. This would mean a required sample size of

$$\text{n} = \text{power index} \times (\text{SD} / (\text{margin} - \text{mean}))^2$$

$$\text{n} = 7.8 \times (32 / (-8 - 0))^2 = 125 \text{ patients per group.}$$

A power index of 7.8 takes care that noninferiority is demonstrated with a power of about 80% in this study (see also Chap. 8).

T.J. Cleophas and A.H. Zwinderman, *Statistical Analysis of Clinical Data on a Pocket Calculator: Statistics on a Pocket Calculator*, DOI 10.1007/978-94-007-1211-9_9,

3. The mean difference between the new and standard NSAID was calculated to be 3.0 g/l with a standard error (SE) of 4.6 g/l. This means that the t-value of the study equaled t = (margin − mean)/SE = (−8 − 3)/4.6 = −2.39 SE-units or SEM-units. This t-value corresponds with a p-value of <0.05 (page 21 bottom row, why the bottom row can be applied is explained in the next Chapter). Non-inferiority is, thus, demonstrated at $p<0.05$.

Step 2: Testing the Significance of Difference Between the New and the Standard Treatment

The mean difference between the new and standard treatment equaled 3.0 g/l with an SE of 4.6 g/l. The 95% confidence of this result is 3.0 ± 2*4.6, and is between −6.2 and 12.2 g/l (* = sign of multiplication). This interval does cross the zero value on the z-axis, which means no significant difference from zero ($p>0.05$).

Step 3: Testing the Significance of Difference Between the New Treatment and a Placebo

A similarly sized published trial of the standard treatment versus placebo produced a t-value of 2.83, and thus a p-value of 0.0047. The t-value of the current trial equals 3.0/4.6 = 0.65 SE-units. The add-up sum 2.83 + 0.65 = 3.48 is an adequate estimate of the comparison of the new treatment versus placebo. A t-value of 3.48 corresponds with a p-value of <0.002 (see page 21, bottom row, the use of bottom row will be explained in the next Chapter). This would mean that the new treatment is significantly better than placebo at $p<0.002$.

Conclusion

We can now conclude that

(1) noninferiority is demonstrated at $p<0.05$, that
(2) a significant difference between the new and standard treatment is rejected at $p>0.05$, and that
(3) the new treatment is significantly better than placebo at $p<0.002$. Non-inferiority has, thus, been unequivocally demonstrated in this study.