and Pettersson, 1978). In other words, 55% more patients achieved a 75% improvement of their psoriasis lesions compared to placebo. To set up an equivalence trial design that retains at least 60% of this effect size, we could propose a margin of ±18% covering 37%–77% of patients having a 75% PASI improvement.

When looking at this range, physicians may say that a product that only produces a 37% PASI 75 response is not equivalent to a product that produces a 59% PASI 75 response; however, this does not reflect the clinical reality. For this, we have to look at how equivalence is assessed in such a trial.

The ultimate goal is to get an understanding of how different the biosimilar and Enbrel could potentially be in the whole psoriasis population based on the data generated in the clinical study—that is, based on a sample of all psoriasis patients eligible for etanercept treatment. This potential difference is estimated by a confidence interval, which will contain the true difference between the two products assessed at a certain probability, usually 90% (e.g., required by FDA) or 95% (e.g., required by EMA). In order for the whole confidence interval for the difference in PASI responses to fall within the justified equivalence margins, the actual difference between the two products must be far less than the 18% margin. In other words, the "point estimate" of the difference between the two products may only be a few percentage points different to meet this criterion. The actual clinical difference between the two products is best represented by the point estimate difference. It is clearly not clinically relevant if the point estimate difference is only a few percentage points between the two products. Clinicians would not consider it problematic or clinically important if 55% of patients achieved a PASI 75 response in one group of patients, while 58% of patients in a second group achieved a PASI 75 response. This is well within the variability of using PASI as a clinical tool. Therefore, to actually meet clinical equivalence statistically retaining 50% or 60% of the effect size requires the point estimates of the two treatments to be fairly similar (Figure 7.3).

One unique aspect that impacts a sponsor's ability to harmonize the clinical trial design assumptions across regions is the confidence interval used for justification of the statistical approach as outlined above. There is a longstanding disagreement in the design of equivalence studies between FDA and EMA regarding the confidence intervals needed to demonstrate equivalence. The FDA along with other regulatory authorities is comfortable with a 90% confidence interval (CI), whereas the EMA routinely insists that sponsors use a 95% CI for equivalence testing in pharmacodynamic and efficacy studies. There is no scientifically valid reason why PD or efficacy should be subject to a higher level of scrutiny (significance level of 2.5%) as compared to showing bioequivalence in PK studies where the use of 90% confidence intervals (significance level of 5%) is a globally accepted norm. Using a 95%, instead of a 90%, confidence interval as described above has a dramatic impact on the sample size in that it adds approximately 20% to the total number of test subjects. If the sponsor is willing to take a risk in harmonizing between regulatory authorities, it may justify a 90% confidence interval outside EMA with comfortable power to achieve success. However, using the same number of subjects while increasing the confidence interval for EMA to 95% reduces the statistical power by approximately 10%. If complying with EMA requirements and retaining reasonable power (≥80%),