$$H_0 : \mu T - \mu R \leq -\delta \text{ or } \mu T - \mu R \geq \delta$$
$$H_a : -\delta < \mu T - \mu R < \delta$$

In these formulas, δ is a positive number denoting the largest acceptable difference between the proposed biosimilar and reference product that is considered to not have clinical impact (i.e., the "equivalence margin"). Analytical similarity is supported if the null hypothesis of non-equivalence, H_0, is rejected. In other words, the statistical equivalence in means is established if the results of the statistical analysis indicate, with high confidence, that

$$-\delta < \mu T - \mu R < \delta$$

A test of the equivalence hypothesis can be conducted by requiring the simultaneous rejection of the following two one-sided null hypotheses:

$$H_{01} : \mu T - \mu R \leq \delta \text{ vs. } H_{a1} : \mu T - \mu R > -\delta$$
$$H_{02} : \mu T - \mu R \geq \delta \text{ vs. } H_{a2} : \mu T - \mu R < \delta$$

The probability of making a Type I error (i.e., declaring incorrectly that a biosimilar product's particular attribute is equivalent to a reference product's particular attribute) for a test of the equivalence hypothesis is controlled at the prespecified level α, provided each of the two one-sided hypotheses, H_{01} and H_{02}, is tested at the same level α.

A convenient way to simultaneously test the two null hypotheses defining equivalence is through a confidence-interval-based test. If the (1–2α)100% two-sided confidence interval of the mean difference lies within (–δ, δ), then both null hypotheses are rejected, and the Type I error probability is controlled at level α for a conclusion of equivalence. For example, a 5% Type I error probability is obtained by requiring a 90% confidence interval to lie within (–δ, δ).

9.3.2.1.2 Margin determination

Determining an appropriate margin is a critical but challenging step for equivalence testing in any setting. Ideally, it would be possible to establish and prespecify a biologically or clinically meaningful equivalence margin based on scientific knowledge or past experience. Often, however, such a margin is not readily available for every quality attribute deemed important enough for Tier 1 testing in a biosimilar development program. With this limitation, the FDA currently recommends use of an equivalence margin that is a function of the reference product's variability for the attribute being tested. Specifically, the equivalence margin should be in the form of $f \times \sigma_R$, where f is a fixed constant, and σ_R is the standard deviation of the quality attribute of the reference product. This suggested form of the equivalence margin is based on three criteria: