It is well known that bias may be introduced when there is an opportunity to select the most desirable result from a number of results obtained; consequently, the probability of a false positive result may be increased, and any estimated differences between the products are likely to be biased toward equivalence. Therefore, to minimize bias and the chance of erroneous conclusions, the statistical analysis plan should be prespecified to the fullest extent possible. In some cases, it may be necessary to first collect preliminary data (e.g., to get an initial estimate of the variability of the reference product's attribute or to select an assay at the outset before finalizing the statistical analysis plan).

9.3.1.4 Finalization of the analytical similarity assessment plan

The final analytical similarity assessment plan should include the risk ranking of attributes, the specification of tiers of evaluation to be used for each attribute/assay, and the final statistical analysis plan. The plan should specify the anticipated availability of both proposed biosimilar and reference product lots for evaluation of each attribute/assay and should include a rationale as to why the proposed number of lots will be sufficient for evaluation purposes. The analytical similarity assessment plan should be discussed with the FDA as early in the biosimilar development program as possible so that agreement can be reached on which attributes/assays should be evaluated in each tier. The final analytical similarity assessment plan should be submitted to the FDA prior to initiating the final analytical assessments; typically this would be done in connection with a meeting with the FDA.

9.3.2 Statistical methods for evaluation

The FDA's current thinking on the statistical evaluation of analytical similarity is described in this section. Sponsors that intend to propose alternative statistical approaches to the FDA should do so during the analysis planning stage.

9.3.2.1 Tier 1 (equivalence test)

9.3.2.1.1 Hypotheses and statistical tests

Analytical similarity of the quality attributes determined to have the highest potential clinical impact (based on the risk ranking and other factors, as described in Section 9.3.1) should be evaluated through formal statistical tests of equivalence. Equivalence of attributes measured on a continuous scale can be assessed by testing the difference in means between the proposed biosimilar and reference product.

In the following formulas, μT and μR denote the population means, and σ_T^2 and $\sigma_R^2\sigma$ denote the population variances of the proposed biosimilar and reference product, respectively. To test for equivalence in means, the null and alternative hypotheses are given by