- The goal of ensuring that values of the attribute being tested for the proposed biosimilar tend to fall within the reference product distribution
- The desire to have a unified representation of the margin for all Tier 1 quality attributes despite different levels of variability
- The goal of having sufficient power for practical sample sizes

After examining a range of possible values for the constant *f*, the FDA determined that a reasonable value should be 1.5. With $\delta = 1.5\ \sigma_R$, the test generally should support equivalence if the 90% confidence interval of the difference in means lies within the interval $(-1.5\ \sigma_R, 1.5\ \sigma_R)$ (i.e., the lower limit of the 90% confidence interval for the difference in means is greater than $-1.5\ \sigma_R$ and the upper limit is less than $1.5\ \sigma_R$). Use of this multiplier in computing the equivalence margin results in a test with reasonable properties under what the FDA considers are realistic conditions. For example, if 10 biosimilar and 10 reference product lots are available, and the variability of the attribute for the reference product (σ_R) is known and not estimated from the sponsor's data, this test has adequate power (i.e., at least 85%) to reject the null hypotheses in favor of equivalence when the true underlying mean difference between the proposed biosimilar and the reference products is small, namely, equal to $\sigma_R/8$, assuming a test of size $\alpha = 0.05$. If the true difference between products is less than $\sigma_R/8$, power will be increased.

A limitation of the proposed approach to setting the equivalence margin is that σ_R is usually not known and must be estimated from the current reference product lots available to the sponsor. If one uses a *t*-test and does not consider the uncertainty in the estimate of the margin, the Type I error probability may be inflated. Alternative tests can be constructed to account for this additional uncertainty, but additional research is needed to better understand the operating characteristics of these tests (such as the small sample size performance of a Wald test based on large-sample approximations; Bickel and Doksum 2007).

9.3.2.2 Tier 2 (quality range approach)

For Tier 2, the similarity acceptance criteria based on reference product results for a specific quality attribute should be defined as $(\hat{\mu}_R - X\hat{\sigma}_R, \hat{\mu}_R + X\hat{\sigma}_R)$, where $\hat{\mu}_R$ is the sample mean and $\hat{\sigma}_R$ is the sample standard deviation based on the reference product lots. The multiplier (X) should be scientifically justified for that attribute and discussed with the FDA. Based on our experience, methods such as the tolerance interval approach and the min-max approach are not recommended (Dong et al. 2015).

Analytical similarity generally should be demonstrated for a quality attribute if a sufficient percentage of test lot values (e.g., 90%) fall within the quality range defined above for that attribute. The lots used for Tier 2 testing should, if possible, be the same as those used for Tier 1 testing.