the acceptance criteria should be established to show non-inferiority. In the case of Tier 1, testing should be done with this model = $1.5^*\sigma_R \leq 90\%$ CI $(\mu_T - \mu_R)$, and in Tier 2 testing, values outside the lower limits of $X^*\sigma_R$ are excluded. The statistical power will increase in one-sided testing.

To make sure that the similarity testing does not fail where it is supposed to be not within the interval or range, but desirably higher or lower, the acceptance criteria can be developed accordingly. For some attributes, the test should be a non-inferiority testing and not a testing of equivalence. Examples of these tests would include the level of aggregates, potentially toxic or undesirable impurities, process-related impurities that impact negatively (e.g., tungsten level), and all other attributes where a lower level is preferred. These should be tested as $(\mu_T - \mu_R) \leq -c^*\sigma_R$, where μ_T is the test product sample mean, μ_R is the reference product sample mean, and σ_R is the standard deviation of the reference product.

9.8.27 Use of public domain values

There may be situations where detailed information on reference product may be available in regulatory documents that are made public, in research publications, or in patent documents that will qualify as the population values of variation (σ_R). A good example is the EAC reported by the FDA for a filgrastim product in its ODAC Committee Meetings. Based on samples ranging from 12 to 39, this range represents in all likelihood the narrowest range; for protein content it was ±2.26% and ±2.08% for U.S. Neupogen; for bioassay, the EAC was ±9.26% in Tier 1 testing for the 90% CI for the difference in the mean.

A justification for using this approach can be that the reported data on the reference product are based on a larger number of lots with different expiry dates and thus provide a better sampling of the population standard deviation, increasing the reliability of the ranges established. This is more applicable to bioassays where the variation is likely going to be higher, and thus a larger number of samples are required to narrow down the standard deviation. However, the developer must also demonstrate that the test methods used are not providing any testing bias. One way to demonstrate this is to compare the 90% CI of the reference products tested side by side with the acceptance criteria established above; if that test passes, then there is strong justification for using this approach. This is a bridging study. However, for some tests where sample handling may contribute to variation such as stability testing, etc., publicly available data may not be used.

9.8.28 Testing across

The FDA recommends that all lots of the test products be tested for all attributes. Are there any situations where not all lots are tested for specific tests?