this should be a starting point; however, it is well known that not all pharmacopeia methods are suitable for the purpose of demonstrating analytical similarity—the developer may want to explore cross pharmacopeia methods or other reported methods. Obviously, choosing a pharmacopeia method reduces the workload as only a verification step is involved, not the validation. This consideration is highly important for analytical methods to have provided a high degree of variability such as bioassays. The FDA expects the developer to choose and/or develop a method that has lower variability. What is low will depend on the nature of test, but the developer may want to conduct a literature search to establish to quote what is a generally accepted variability for the stated methods to justify its own coefficient of variation of the test method.

3. *A selection of statistical tiers*—As shown above, the FDA classifies all statistical testing into three categories depending on the relative impact of the testing on clinical efficacy and safety. At this stage, a test will be classified as Tier 1, 2, or 3.
4. *A number of lots tested side by side*—A common question asked by the biosimilar developer is, "how many lots should be tested?" While the FDA has not clearly described its expectations, except for Tier 1, wherein a difference equal to $\sigma/8$ should be detectable, this allows calculation of the power of testing as shown later in the chapter. Ideally, if 10 different lots of reference and test products are available, this should provide sufficient power, generally over 80%, which should be acceptable to the FDA.
5. *Selection of power and alpha*: The Type I error (based on alpha) is the FDA risk, and the Type 2 (based on beta) is the developer's risk; a proper selection of both is needed to establish an equivalence acceptance criterion. As we will see, the FDA can be flexible on Type I error if the developer can show that the attribute can be tightened in its limits through manufacturing controls.
6. The number of reference lots to establish acceptance criteria is based on the standard deviation of the reference lots. How many lots to use will depend on the degree of variability. In most instances, between 5 and 10 lots will be needed to establish a reasonably narrow standard deviation. Know that there are two ways of doing it—first by adopting less variable testing methods and second by increasing the number of test lots.
7. Randomization of lots is something that the FDA recommends, meaning that the lots of both the reference and the test are not picked out with bias—only good lots of the test and only bad lots of the reference. This expectation of the FDA can be very