by the manufacturing process used by the biosimilar product developer. Whereas the compendial specifications for drug substance may be useful in extrapolating the specification for the drug product with regard to primary and higher-order attributes, the goal of demonstrating that the variability between the biosimilar product and the originator product is comparable can only be reached in a side-by-side testing. A good example is the level of protein content, which may be allowed a range of establishing ±5% for the release purpose, but for the analytical similarity demonstration, the difference in the mean of the test and originator product will be based on $1.5*\sigma_R$. In the case of EP2006, the FDA reported the acceptance intervals to be 2.08–2.26%. The same applies to Tier 2 range testing, where the goal is to demonstrate a similar variability. As a result, it is entirely possible for the lots of the biosimilar product to meet all release specifications, yet fail in the analytical similarity demonstration. For example, the biosimilar product may use release criteria where total impurities are not more than 2% and no single impurity is more than 1%; however, if minor impurities are found in the biosimilar product, but not in the originator product, this may be a reason for failing similarity.

9.8.25 Mixed graphic and numerical data

Some testing provides a graphical output, but the values such as peak height and area under the peak curve can be quantitated. Should these data output be treated as numerical or graphical?

Whenever there is a graphical output, it should demonstrate no extraordinary peak, no extraordinary heights of the peaks, and no extraordinary baseline—this is an overall evaluation of the graphical output and applies to all tiers (1–3). However, peaks have known significance relating to potency, purity, and safety, and as a result, the quantifiable graphical attributes have clinical meaningfulness; these can be compared for Tier 1 and Tier 2 analysis. For Tier 3 testing, there is no further need to perform any quantitative evaluation. In this situation, the data analysis will be presented both as highly similar graphical presentation as well as a tier-sensitive equivalent interval or range. However, if the numerical comparisons do not have any clinical relevance, a graphical representation alone would be sufficient, regardless of the attribute.

9.8.26 Non-inferiority testing

There can be several situations where the attribute tested less desirable in Tier 1 and Tier 2 testing; the comparison test may fail if the TOST approach is used instead of one-sided testing or non-inferiority testing. How should these data be analyzed and interpreted?

The developer should define the attributes that are characterized in this category, such as level of aggregates, impurities adversely affecting activity or safety, degradation rate, HCP, residual DNA, etc., for which