Testing for Non-Inferiority

A form of equivalence helps to compare a new lab, instrument or test method with an established one.

Q What is non-inferiority, and how is this related to equivalence in comparing two sets of test results?

A A previous Data Points article covered the concept of equivalence, defined as the condition that the true average test results measured on the same material by two testing sources differ by no more than pre-determined limits.[1] The ASTM Practice for Conducting Equivalence Testing in Laboratory Applications (E2935) provides calculation details for equivalence studies used to qualify a new laboratory, instrument or test method by comparison with an established one. This article discusses another form of equivalence called non-inferiority that is useful in situations where the equivalence is required in only one direction.

A quick review of equivalence testing will be useful before discussing non-inferiority. Equivalence is supported if the average test results from the two data sources differ by no more than a predetermined equivalence limit, denoted as *E,* at a given level of confidence. The statistical procedure for determining equivalency from test result data is based on calculating a confidence interval for the true mean difference as $D \pm t\, s_D$, where D is the difference between the two test result averages, s_D is the standard error of that difference, and t is a tabulated multiplier based on the number of data and a preselected confidence level. The calculation for s_D is based on the standard deviations of the two sets of data and the study design. Then equivalence is supported if this entire confidence interval is completely contained within the equivalence interval $0 \pm E$. Stated more formally, equivalence is supported if the lower confidence limit, $LCL = D - t\, s_D$, is greater than the lower equivalence limit, $-E$, and the upper confidence limit, $UCL = D + t\, s_D$, is less than the upper equivalence limit, E. This is known as the two one-sided tests or TOST procedure.

Non-inferiority is the special case of equivalence needed in only one direction. For example, the candidate source may be slightly inferior to the established source with respect to a particular performance characteristic, such as assay sensitivity, but the candidate may have off-setting advantages such as lower cost or faster delivery of results. If the margin of inferiority is shown to be in the worst case as less than some difference *E*, then the candidate method is said to be non-inferior to the established method. Other terms used in practice for non-inferior are “equivalent or better” or “at least equivalent as.”

“The concept of non-inferiority deserves greater use in the field of test method evaluation. Standards development work is now in process by Committee E11 on Quality and Statistics to include non-inferiority procedures in ASTM E2935.”

The statistical procedure for noninferiority uses the same confidence interval test as equivalency, but the decision depends only on the outcome in one direction. In the situation where the performance characteristic is defined as "higher is better," such as method sensitivity or probability of detection, the statistical test supports noninferiority when $LCL > -E$. An example comes from environmental testing for microbial contamination where the traditional method involves counting microbial colonies after plating and incubating the sample, which may take place over time periods measured in days. Newer rapid methods have benefits in timeliness and cost even though they may have slightly lower sensitivity than the traditional method. Due to test results that may run over several orders of magnitude, microbiologists often use a logarithmic scale for counts with a margin allowance of 0.5 for $\log_{10}$(counts), and this can be used to define the magnitude of E. In terms of relative counts this margin amounts to a roughly 30 percent reduction allowance for the noninferiority margin.

Conversely, when the performance characteristic is defined as "lower is better," such as incidence of misclassifications, the statistical test would support non-inferiority when $UCL < E$. Note that the equivalence procedure comprises two one-sided statistical tests while the non-inferiority procedure performs only a single one-sided statistical test.

Up to this point, the performance parameter being discussed has been based on differences in means, but another important parameter in test method evaluation is the test precision, measured inversely as the population standard deviation or variance, and lower is better for these parameters. Because variances are a scale parameter, the non-inferiority test would use the ratio R of the two variances instead of their difference; thus $R = s_C^2 / s_E^2$, where s_C^2 and s_E^2 are the calculated variances of the test results from the candidate and established test methods, respectively. An upper confidence limit for R, denoted UCL_R, for a given confidence level and amount of data can be found from the tabulated F distribution. The equivalence limit E would also be in the form of a ratio. For example, if $E = 2$, this non-inferiority margin would allow the candidate method to have up to twice the variance of the established method or up to about 1.4 times the standard deviation. The statistical test would then support noninferiority if $UCL_R < E$.

An important topic in the design of non-inferiority procedures involves the amount of data required to control the power of the procedure; that is, the probability of accepting non-inferiority when the candidate method is truly non-inferior to the established method. Fortunately, the power computation for the non-inferiority procedure is simpler than that for equivalence because only a single one-sided statistical test is involved. Unfortunately, the amount of data needed for good power in inferiority procedures for precision must be much larger (at least a half order of magnitude higher) than those for means.

A good reference for non-inferiority procedures is *Design and Analysis of Non-Inferiority Trials*, although their context is clinical trials for pharmaceuticals.[2] However, many numerical examples are included and these are easily translatable to test method evaluation.

In summary, the concept of non-inferiority deserves greater use in the field of test method evaluation. Standards development work is now in process by Committee E11 on Quality and Statistics to include non-inferiority procedures in ASTM E2935.

References

1. Murphy, T.D., "Testing for Equivalence," ASTM *Standardization News*, Sept./Oct. 2014, Vol. 42, No. 5, pp. 16-17.

2. Rothmann, M.D., Wiens, B.L. and Chan, I.S.F., *Design and Analysis of Non-Inferiority Trials*, Chapman and Hall/CRC, 2012.

Thomas D. Murphy, retired statistical consultant, Fredericksburg, Virginia, is chairman of Subcommittee E11.20 on Test Method Evaluation and Quality Control, a part of Committee E11 on Quality and Statistics. He served as chairman of E11 during 2002-2003.

John Carson, Ph.D., of CB&I Federal Services LLC and P&J Carson Consulting LLC, Ohio, is the Data Points column coordinator. Vice chairman of E11.30 on Statistical Quality Control, part of E11 on Quality and Statistics, he is also a member of Committees E50 on Environmental Assessment, Risk Management and Corrective Action and of D02 on Petroleum Products, Liquid Fuels and Lubricants.