

Superiority, Equivalence and Non-Inferiority

by I. Elaine Allen and Christopher A. Seaman

Most statistical tests are performed to show whether two measurements, processes, products or treatments are significantly different from each other and whether we can reject the null hypothesis. The null hypothesis will be rejected if the test statistic is sufficiently large compared to the critical value or if the p-value of the test statistic is sufficiently small compared to a prespecified level of α .

For example, when a new product is introduced, the goal is often to show its superiority compared to an existing product. In many cases, particularly in healthcare, it is equally likely that the goal is a degree of equivalence between the old and the new product, rather than superiority. Instead of testing for equivalence, the use of non-inferiority experimental designs is a useful technique to accomplish this goal. This column will outline the need for non-inferiority testing and specifying the non-inferiority margin using clinical trials and process improvement techniques as examples.

Equivalence and Sample Size

When testing the difference between two independent measurements, as the difference between the two measurements approaches zero, basic statistical methods show the number of observations required to test the significance of this difference as it approaches infinity. For example, to test the hypothesis that a new product Y is significantly different from an existing product X, we can use the following formula:

$H_0: X = Y$ or $X - Y = \Delta$ as $\Delta \rightarrow 0$ for continuous X and Y vs. $H_1: X \neq Y$, with the test statistic calculated as

$$Z = \left(\frac{X - Y}{\sigma/\sqrt{n}} \right), \text{ giving a required}$$

$$\text{sample size of } n = \frac{(Z^2 \sigma^2)}{(X - Y)^2},$$

$$\text{and } n \rightarrow \infty \text{ as } (X - Y) \rightarrow 0$$

Solving for n, the sample size, shows that as the difference approaches zero, the sample size increases without bound. For this reason we change the test to one in which we identify an interval for the difference between X and Y, in which for any dif-

Useful technique can provide more cost effective sampling and analysis when applied to statistical quality control.

ference less than that specified, Y is declared not inferior to X. For testing non-inferiority, rather than using the zero point estimate for strict equivalence, we specify a non-inferiority margin.

Developing a Non-inferiority Hypothesis

In clinical trials in the biopharmaceutical industry, testing for the non-inferiority of a new product compared with an existing approved competitor has become an acceptable practice. The goal of these trials is to show that the new treatment or product is statistically (and clinically) not inferior to the standard approved product. Using X and Y as above, we can formulate a non-inferiority hypothesis test as follows:

$$H_0: X - Y \geq \Delta, \text{ in which } \delta \text{ strictly } > 0$$

$$H_1: X - Y < \Delta$$

This hypothesis implies that X is strictly superior to Y under the null hypothesis versus Y is not inferior to

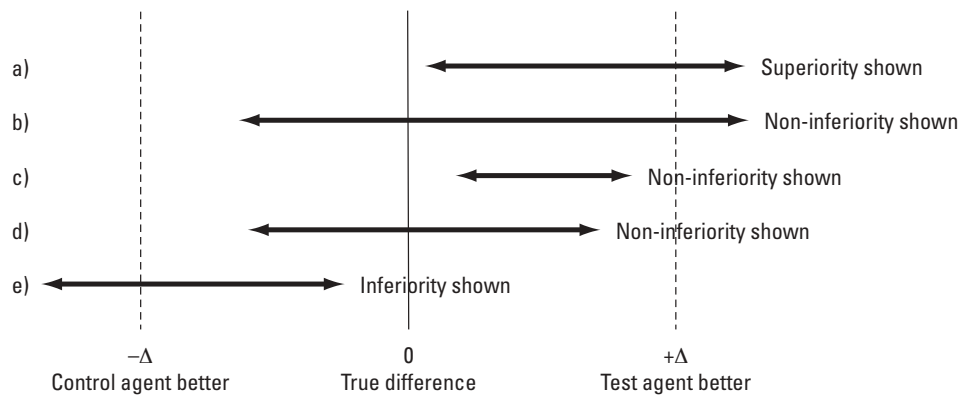
X under the alternate hypothesis. Here Y represents a new product, treatment or process, and X is the standard approved product, treatment or process in use. The boundary Δ is the prespecified non-inferiority margin, and Δ is defined to be strictly greater than zero.

Clearly, the major issue facing the experimenter in designing a non-inferiority trial is the choice of the non-inferiority margin Δ . While the simplest formulation of the non-inferiority margin in a hypothesis test is the difference between two continuous variables, these margins and tests can be formulated as ratios, ordinal estimates and percentages (for example, comparing the percentage defective from an old manufacturing process to the percentage defective in a proposed streamlined, new manufacturing process).

To assess whether the null hypothesis can be rejected (or, in this case, that the non-inferiority margin is met) it is usual practice to perform a one-sided hypothesis test at the α level of significance. One-sided tests are used here because our alternative hypothesis is examining whether the new product is the same (non-inferior to) or better than the existing product.

We are not interested in the hypothesis that the new product is significantly worse than the existing product. By not rejecting the non-inferiority hypothesis test, we will have accomplished this goal. This test will be identical to calculating a two-sided $100(1 - 2\alpha)$ percentage confidence interval for the difference between the two products ($X - Y$). If the upper bound on the confidence interval of the difference between X and Y is less than Δ (the prespecified non-inferiority margin), you can conclude that the standard product is more efficacious than the new product by no more than Δ . Then the null hypothesis is rejected in favor of the

FIGURE 1 Non-inferiority Experiment Outcomes



non-inferiority of the new product.

Determining The Inferiority Margin

Several assumptions must be met prior to the specification of Δ .

First, our outcome variable from the standard comparator cannot be a placebo. The standard comparator is assumed to have some effect (for example, efficacious treatment or reduction in defectives) based on the measurement of the outcome variable.

Second, the study must be designed so that if Y is not inferior to X it will be shown (for example, sufficient sample size, control of variability within and between observations).

Third, the outcomes of X and Y in the non-inferiority trial are greater than would be seen in a placebo or would be seen by chance alone.

In summary, the non-inferiority margin must be chosen to test Y as being superior to placebo and non-inferior to X. The value of the non-inferiority margin should be chosen as the maximum acceptable difference between X and Y.

For example, when introducing a new manufacturing process, Y, with efficacy of the new and old processes being the number or percentage of defective products produced, the maximum acceptable difference between processes X and Y would be the increase in the percentage of defectives when moving to the new process that would still render the new process as “equivalent” or not inferior to the old process.

Often the old process is time inten-

sive and more costly. Finding the new process not inferior to the old process gives sufficient reason to replace it. In clinical trials, with patient and treatment variability, a new treatment that performs within 10 to 20% of an old treatment is often the margin used to be called non-inferior. Often the safety profile of the new treatment is superior to the standard treatment, making the new treatment preferable even though it is not superior in efficacy.

Outcomes of a Non-inferiority Experiment

At the conclusion of a non-inferiority experiment, the confidence interval can be plotted and examined on a chart showing the non-inferiority

a. Superiority shown: The confidence bound of $(X - Y)$ does not include the 0 difference line, and the confidence bound exceeds the upper margin of non-inferiority. The result shows the superiority of Y. This result is identical to that of a hypothesis comparing X and Y in which superiority of Y is shown.

b. Non-inferiority shown: Non-inferiority is shown because the lower bound of $(X - Y)$ crosses the 0 difference line; although the upper bound does exceed the upper non-inferiority boundary, but the lower bound does not cross the lower non-inferiority boundary.

c. Non-inferiority shown: Here non-inferiority is shown because the confidence bounds for $(X - Y)$ are totally within the non-inferiority boundaries.

d. Non-inferiority shown: Non-inferiority is shown because, although the confidence bounds for $(X - Y)$ cross the 0 line, the lower bound does not cross the lower non-inferiority boundary.

e. Inferiority shown: The lower bound of the confidence interval of $(X - Y)$ crosses the lower boundary of the inferiority margin.

margin representing the non-inferiority difference between X and Y. Several outcomes are possible, from superiority of the new process, product or treatment, as shown in Figure 1. Note both the upper and lower non-inferiority margins are given; although for this hypothesis test we are comparing $(X - Y)$ with the lower non-inferiority boundary.

Switching Between Non-inferiority and Superiority

An experiment might be designed to show non-inferiority. Through testing, the outcome actually might show a statistically significant improvement for Y compared with X. The issue arises as to whether the experimenter is entitled to claim superiority for Y over

X, rather than merely the non-inferiority for which the trial was planned. Statistically this is not a problem, as is readily seen from the confidence interval (a) depicted in Figure 1 (p. 53). If the upper confidence bound is greater than the non-inferiority margin and the lower confidence bound exceeds zero, a conclusion of superiority is warranted.

It might also be possible to support a non-inferiority conclusion in the case of an experiment designed to show superiority. This conclusion can be made only if both the hypothesis of superiority and the hypothesis of inferiority are stated before the experiment.

Summary

The use of non-inferiority experimental designs is well established in evaluating new clinical entities and devices in the biotech and pharmaceutical industries. Adapting this technique for identifying processes and products as non-inferior vs.

equivalent within a prespecified non-inferiority margin can provide more cost effective sampling and analysis when applied to statistical quality control.

BIBLIOGRAPHY

D'Agostino, R.B., J.M. Massaro and L.M. Sullivan, "Non-inferiority Trials: Design Concepts and Issues—The Encounters of Academic Consultants in Statistic," *Statistics in Medicine*, Vol. 22, No. 2, 2003, pp. 169-186.

European Medicines Agency, "Points to Consider on Switching Between Superiority and Non-inferiority," www.emea.europa.eu, 2000.

Morikawa T., and M. Yoshida, "A Useful Testing Strategy in Phase III Trials: Combined Test of Superiority and Test of Equivalence," *J Biopharm Stat*, Vol. 5, 1995, pp. 297-306.

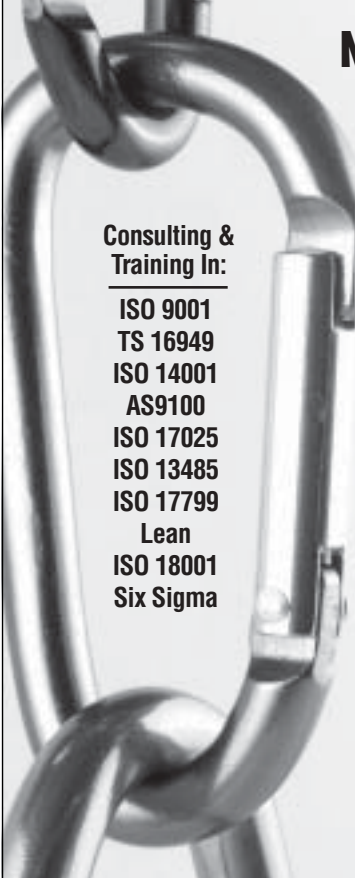
Pater, Cornel, "Equivalence and Non-inferiority Trials—Are They Viable Alternatives for the Registration of New Drugs?" *Business Briefing: Pharmagenetics*, 2004, pp. 50-55.

I. ELAINE ALLEN is professor of statistics and entrepreneurship at Babson College in Wellesley, MA. She earned a doctorate in statistics from Cornell University in Ithaca, NY, and is a member of ASQ.

CHRISTOPHER A. SEAMAN is a doctoral student in mathematics at the Graduate Center of City University of New York.

Please comment

If you would like to comment on this article, please post your remarks on the *Quality Progress* Discussion Board at www.asq.org, or e-mail them to editor@asq.org.



Consulting & Training In:

- ISO 9001
- TS 16949
- ISO 14001
- AS9100
- ISO 17025
- ISO 13485
- ISO 17799
- Lean
- ISO 18001
- Six Sigma

Never decide the mountain is too high... We can help you reach the top.

We provide a *personal approach* and *real-world, practical experience*.

AUTO SUPPLIERS:
TS 16949 Implementation;
Core Tools & Auditor Training

6 SIGMA CLIENTS:
Green Belt, Black Belt Classes

AUDITING SERVICES:
ISO, TS, SOX, CISA

ISO/FDIS 27001 Security Standard Implementation

Environmental Systems Implementation

ISO 9001 Implementation and Training

Making CA Work:
Eliminating Recurring Problems

Public Training Calendar
(see our website)



YOUR PARTNER IN QUALITY MANAGEMENT

An ISO Registered Company





Schedule
Contract GS-23F-00001



Phone: 800.355.3876 or 770.395.0124 • mail@euroquest.net • www.EuroQuest.net