# Course 14: Non-Inferiority, Data Collection, & Trial Closeout

## Non-Inferiority: Issues in Closeness
- Introduction: 299
- Overview: 302
- The Problem: 304
- Examples: 305
- The Goal and General Methodology: 306
- Methodology with Means of Normal Distributions: 309
- Methodology with Binomial Distributions: 312
- Other Issues: 314
- References: 315

## Data Collection and Quality Control
- Problems in Data Collection: 320
- Minimizing Poor Quality Data: 321

## Trial Closeout
- Termination Procedures: 322
- Data Storage: 323
- Dissemination of Results: 323
NON-INFERIORITY: ISSUES IN CLOSENESS

Introduction
To confirm the efficacy of a new treatment, phase III studies use a comparison of the treatment to a control. Efficacy can be confirmed by showing either:

1. The new treatment has superior efficacy to the control, or
2. The new treatment has efficacy equivalent to that of the active control, or
3. The new treatment is not much worse than the active control.

The first is commonly called a superiority trial; the second, an equivalence trial; and the third, a non-inferiority trial.

Superiority
The superiority study design aims to show that a new drug is better than an active control with respect to the efficacy variable of interest.

The statistical hypothesis of the study is that there is a difference in efficacy that favors the new treatment. A hypothesis test designed to reject no difference in effect between the treatment and control is usually used to confirm this hypothesis. A treatment and control also may be used.

Non-inferiority
This study design aims to show that a new drug is not much worse than an active control with respect to the efficacy variable of interest.

The statistical hypothesis of the study is that the difference in efficacy is at worst a small amount (delta) that favors the control. A confidence interval that estimates the difference in effect between the treatment and control is usually used to confirm this hypothesis.

Interpretation of non-inferiority trials can be much more difficult than that of superiority trials. Some specific and important issues to consider are:
• The quality of the trial and compliance of the subjects to the trial
• The selection and justification of delta (the acceptance difference)

Considerations
Can a trial be designed for both superiority and non-inferiority?
Yes, if the definition of non-inferiority (particularly delta) is pre-specified and accepted, then it is generally appropriate for a trial to lead to either conclusion.
What is the difference between a non-inferiority trial and an equivalence trial?
The terminology is often used synonymously. A subtle difference is that equivalence implies that the interest is in showing that the two therapies are similar, whereas non-inferiority allows for the treatment to be better than the control.

Strict definitional equivalence is used more for bio-equivalence trials that compare different formulations of the same drug.


Interpretations of trial results depend on where the confidence interval falls relative to delta and zero.

<table>
<thead>
<tr>
<th>Superior?</th>
<th>N</th>
<th>N</th>
<th>N</th>
<th>N</th>
<th>N</th>
<th>N</th>
<th>Y</th>
<th>Y</th>
<th>Y</th>
<th>Y</th>
</tr>
</thead>
<tbody>
<tr>
<td>Non-inferior?</td>
<td>N</td>
<td>N</td>
<td>Y</td>
<td>Y</td>
<td>Y</td>
<td>Y</td>
<td>Y</td>
<td>Y</td>
<td>Y</td>
<td>Y</td>
</tr>
<tr>
<td>Equivalent?</td>
<td>N</td>
<td>N</td>
<td>Y</td>
<td>Y</td>
<td>Y</td>
<td>Y</td>
<td>Y</td>
<td>N</td>
<td>N</td>
<td>N</td>
</tr>
<tr>
<td>Clinical interpretation</td>
<td>Worse</td>
<td>Possibly worse</td>
<td>Comparable</td>
<td>Comparable</td>
<td>Comparable</td>
<td>Comparable</td>
<td>Possibly better</td>
<td>Better</td>
<td></td>
<td></td>
</tr>
</tbody>
</table>
Statistical testing procedures for exhibiting a difference between treatments are used by researchers in many fields, especially clinical trials; however, there are still some misinterpretations when the results are not statistically significant. These misinterpretations may result in inappropriate, or even incorrect, conclusions. Although the lack of statistical significance provides some information regarding the “closeness” of the treatments, formal conclusions cannot be drawn using the classic approach of testing the null hypothesis of equality. The potential misinterpretation is even greater when the purpose of the investigation is to exhibit “closeness.”

Here methodology for use when the goal is to exhibit non-inferiority of treatments is discussed. The techniques allow for direct conclusions regarding non-inferiority. A review is presented, with discussion of appropriate interpretation.
Overview

In recent years, there seems to have been an increase in attention to the problem of exhibiting that treatments are “close.” Although this problem may be encountered in many research situations, the focus here is on the comparison of treatments in clinical trials, especially as it pertains to drug development. The problem of closeness has been examined primarily from the perspective of bioequivalence, where the treatments are two formulations of the same drug or one drug given under different conditions, such as fed or fasted. In such studies, the goal is usually to establish that the treatments are not very different with respect to pharmacokinetic parameters. Extensive research has been conducted on the statistical methodology for such studies.

Much of the seminal research on bioequivalence was performed by Westlake, with Westlake (1972, 1975, 1976) among a lengthy list of early contributions. Westlake (1979, 1981, 1988) provided further results and discussion. More recently, Steinjians and several coauthors provided additional details regarding various aspects; see, for example, Steinjians and Hauschke (1990); Schulz and Steinjians (1991); Diletti, Hauschke, and Steinjians (1991); Steinjians, Hauschke, and Jonkman (1992); and many others. Other references that should be mentioned are Metzler (1974), Kirkwood (1981), Anderson and Hauck (1983), and Schuirmann (1987). Not surprisingly, each paper generates further research; for example, sample size determination for the procedures in the latter two references was considered by Frick (1987); Müller-Cohrs (1990); Phillips (1990); Hauschke, Steinjians, Diletti, and Burke (1992); Liu and Chow (1992); and Bristol (1993a).

When the goal is to establish the closeness of two treatments with respect to a clinical variable, some of the philosophy and methodology from bioequivalence studies can be applied. Such studies are referred to as non-inferiority (or therapeutic equivalence) trials and are usually conducted with the goal of showing that the new treatment is not much worse than the active control.

The new treatment may be preferred because of:
- Safety factors (e.g., fewer side effects)
- Convenience (e.g., improved dosing regimen)
- Tolerance (e.g., loss of efficacy with the control)
- Lack of response (e.g., nonresponders to control may respond to the new treatment)
- Resistance (e.g., pathogens may no longer respond to a certain antibiotic)
However, the new treatment may be less efficacious. If the decrease in efficacy is too great, the preference with respect to factors unrelated to efficacy may not be sufficient to advocate use (e.g., approval) of the new treatment. Thus, it suffices to show that the new treatment is not much worse than the control with respect to the efficacy variable of interest.

In some cases, the new treatment may be superior to the control with respect to efficacy, but the difference is not clinically significant, or clinically significant but small. This small difference would require a very large sample size if a comparative study was designed to test for a statistically significant difference. If the required sample size is not economically feasible and it suffices to exhibit non-inferiority, then the techniques presented here could be used.

Three major differences between non-inferiority trials and bioequivalence trials are the goal, the design, and the distributional assumptions. Bioequivalence trials are usually conducted as crossover studies, whereas non-inferiority trials are usually conducted as parallel-group studies. More importantly, the goal of bioequivalence trials is the two-sided problem of showing that the two treatments are not very different, whereas the goal of non-inferiority trials is the one-sided problem of showing that the new treatment is not much worse than the control. Finally, the pharmacokinetic variables are assumed to have a normal distribution, although a logarithmic transformation may be used to improve the validity of this assumption. Non-inferiority trials have been discussed and conducted for various parametric models, but the binomial responses tend to dominate the literature.
The Problem

Statistical significance is often misinterpreted, especially when the results of a study are “negative,” in that the difference between treatments was not statistically significant. This has been discussed by many authors, such as Wade and Waterhouse (1977), Matthyse (1978), Spriet and Bieler (1979), and Detsky and Sackett (1985). In these papers and the early work on bioequivalence, it is noted that the lack of a statistically significant difference does not allow one to conclude non-inferiority of distributions.

To avoid the misinterpretations associated with p-values and the concept of significance testing, confidence intervals have received considerable attention. In addition to providing information regarding statistical significance, confidence intervals also provide an estimate of the difference between the treatments. Simon (1986), Bristol (1989), Borenstein (1994), and many others have advocated and discussed the use of confidence intervals in clinical trials. The last reference provides an extensive review of the subject.

With the above issues in mind, studies conducted to show that two treatments are equivalent should have this goal explicitly stated, and appropriate methodology should be employed.
**Examples**

**Example 1**
Strickland et al. (1977) presented the results of a crossover study, comparing IV iron-dextran complex to oral ferrous sulfate in 20 renal dialysis patients, of whom 15 completed the study. The mean within-patient differences (oral vs. IV) for hemoglobin and blood loss are given in the following table.

<table>
<thead>
<tr>
<th></th>
<th>Mean Diff.</th>
<th>95% Conf. Int.</th>
</tr>
</thead>
<tbody>
<tr>
<td>Hemoglobin (g)</td>
<td>0.20</td>
<td>(-0.31, 0.71)</td>
</tr>
<tr>
<td>Blood Loss (ml)</td>
<td>86</td>
<td>(-82, 250)</td>
</tr>
</tbody>
</table>

Conclusion: “There was no significant difference between the treatments in mean hemoglobin during treatment,” and “changes in hemoglobin of only 0.5 are associated with significant improvement.” Thus, “oral iron is as efficient as intravenous iron.”

**Example 2**
Sides (1993) presented the results of a comparison of dirithromycin 500 mg QD and erythromycin 250 mg TID for seven days in the treatment of acute exacerbations of chronic bronchitis. The paper states that the comparison was “not statistically significant” (using a chi-square test). The success rates for the post-treatment clinical response were 86.1% (87/101) for dirithromycin and 88.9% (72/81) for erythromycin [p=0.579, with 95% CI = (-12.5%, 7.0%)]. The success rates for the clinical response at the extended follow-up visit were 87.5% (70/80) for dirithromycin and 89.6% (60/67) for erythromycin [p=0.698, with 95% CI = (-12.4%, 8.3%)].

**Example 3**
Rex et al. (1994) presented the results of a comparison of fluconazole and amphotericin B in the treatment of candidemia in patients without neutropenia. They stated that the goal was to establish the non-inferiority of fluconazole and amphotericin B. The success rates were 70% (72/103) for fluconazole and 79% (81/103) for amphotericin B [p=0.22, with 95% CI = (-5%, 23%)]. The statistical analysis plan was to test $H_0$: $|\pi_A - \pi_F| \leq 0.2$ against $H_1$: $|\pi_A - \pi_F| > 0.2$, using the two-sided Cochran-Mantel-Haenszel test. The decision rule for declaring non-inferiority was not explicitly stated, but the lack of statistical significance for this testing formulation does not allow the conclusion to be made.
The Goal and General Methodology

As noted earlier, the goal of non-inferiority trials is to show that the new treatment is not much worse than the control. This goal can be achieved using a **hypothesis-testing** formulation or a **confidence interval** formulation (both of these approaches will be discussed in detail below). The hypothesis-testing formulation focuses on the difference in the **parameter space** and examines the consistency of the data with the hypothesized difference, whereas the confidence interval formulation focuses on the difference in the **observed values** and relates the observed difference to a region in the parameter space. It is well known that there is a **relationship** between hypothesis testing and confidence intervals when the conventional null hypothesis of no difference is tested; a similar relationship also holds for non-inferiority testing.

Using either the hypothesis-testing approach or the confidence interval approach, a parametric model is usually assumed, and a distance function is defined on the parameter space to define non-inferiority and also is used to measure the observed distance between the treatments. The choice of the **distance function** $d(\theta_T, \theta_C)$, where $\theta_T$ and $\theta_C$ are the parameters for the new treatment and the control, respectively, depends on the problem being solved. For location parameters, the distance function should be the **difference**; for scale parameters, it should be the **ratio**.

For **survival studies**, Fleming (1990) considered the hazard ratio; Com-Nogue, Rodary, and Patte (1993) also considered the difference in the survival function at a fixed time; and Munk (1996) used a specific function for the proportional hazards model, similar to that considered by Desu and Bristol (1986) for various selection problems with this model. Related problems were considered by Rodary, Com-Nogue, and Tournade (1989); Wellek (1993); and Bristol (1993b).

For **binomial distributions**, the difference in the success probabilities is usually used with asymptotic normality used to **compute the probabilities of interest**. Using the results of Gart (1971), Dunnett and Gent (1977) presented a procedure that used the odds ratio and compared the results to the chi-square test for 2x2 tables. Other models that have been considered include a procedure by Lee and Lusher (1991) based on **McNemar’s test for paired dichotomous data**, as well as exact tests proposed by Mehta, Patel, and Tsiatis (1984) for ordered categorical data.
**Hypothesis-testing approach**

Given the parameters $\theta_T$ and $\theta_C$ for the new treatment and control, respectively, and the appropriate distance function $d(\theta_T, \theta_C)$, the hypothesis-testing approach is:

\[
\text{Test } H_0: d(\theta_T, \theta_C) \in D_C \text{ against } H_1: d(\theta_T, \theta_C) \in D_E,
\]

where the region $D_C$ corresponds to the **superiority of the control**, and $D_E$ is the complement of $D_C$, such that $D_E$ corresponds to (one-sided) non-inferiority (i.e., the new treatment is superior, or neither treatment is superior). This approach allows only for a conclusion of non-inferiority or a decision that this conclusion cannot be made.

One could partition $D_E$ into $D_E = D_T \cup D_N$, where $D_T$ corresponds to the **superiority** of the new treatment, and $D_N$ is the region where **neither treatment is superior**. For example, when the treatments are being compared with respect to location parameters, $d(\theta_T, \theta_C) = \theta_T - \theta_C$, $D_C = (-\infty, -\delta]$, $D_N = (-\delta, \delta)$, and $D_T = [\delta, \infty)$, so $D_E = (-\delta, \infty)$. With this partition, a three-decision procedure could be defined to allow for a formal decision to be made if $H_0$ is not rejected in favor of $H_1$ (i.e., if non-inferiority is not concluded).

Morikawa and Yoshida (1995) proposed a strategy using a closed testing procedure to allow testing for both non-inferiority and superiority. A similar procedure was recently proposed by Dunnett and Gent (1996). Desu and Bristol (1985a, 1985b) presented selection procedures to select the better of two treatments or to declare the treatments equivalent.
**Confidence interval approach**

Although the confidence interval approach can be used for any model, the presentation here is restricted to location parameters. Let \((L,U)\) denote a \((1 - \alpha)100\%\) confidence interval for \(\theta_T - \theta_C\), where large values of the parameter are preferable and indicate superiority. The decision rule is:

Declare **non-inferiority if** \(L \geq -d\).

This formulation is considered here for use when large means indicate superiority because a negative difference is consistent with superiority of the control.

Although the problem of non-inferiority is a one-sided problem, the confidence interval approach defined above uses two-sided methodology. One-sided methodology could be used by replacing the \((1 - \alpha)100\%\) confidence interval by the \((1 - \alpha)100\%\) **lower confidence limit**. An issue here is the choice of \(d\); the choice of \(d\) can be made by 1) non-inferiority (clinically meaningful difference), 2) statistical considerations, or 3) regulatory considerations.
Methodology with Means of Normal Distributions

Let $\mu_T$ and $\mu_C$ denote the means of normal distributions for the new treatment and control, respectively, with a common variance $\sigma^2$.

**Normal means: hypothesis-testing approach**

In this case, the hypothesis-testing approach is:

Test $H_0: \mu_T - \mu_C \leq -\delta$ against $H_1: \mu_T - \mu_C > -\delta$,

where $\delta$ is a specified positive constant.

The testing procedure is:

Reject $H_0: \mu_T - \mu_C \leq -\delta$ in favor of $H_1: \mu_T - \mu_C > -\delta$ if

$$\overline{X}_T - \overline{X}_C \geq -d,$$

where $d$ satisfies:

$$P\{\overline{X}_T - \overline{X}_C \geq -d | \mu_T - \mu_C = -\delta\} = \alpha,$$

So $d = \delta - z_\alpha \sigma (2/n)^{1/2}$.

Thus the rejection rule is:

Reject $H_0: \mu_T - \mu_C \leq -\delta$ in favor of $H_1: \mu_T - \mu_C > -\delta$ if

$$\overline{X}_T - \overline{X}_C \geq -\delta + z_\alpha \sigma (2/n)^{1/2}.$$

Note that this is very demanding if $-\delta + z_\alpha \sigma (2/n)^{1/2} \geq 0$ [i.e., $n \leq 2(z_\alpha \sigma / \delta)^2$], as this would require the observed difference to be non-negative to show that the difference in the parameter space exceeds $-\delta$.

The probability of declaring non-inferiority is the power of the test, given by:

$$P\{\text{declare non-inferiority}\} = P\{\text{reject } H_0\}$$

$$= P\{\overline{X}_T - \overline{X}_C \geq -\delta + z_\alpha \sigma (2/n)^{1/2}\}$$

$$= \Phi\left[0.5n^{1/2}(\delta + (\mu_T - \mu_C)/\sigma - z_\alpha)\right].$$
The sample size can be chosen such that, for specified $\delta^*$:

$$P\{\text{declare non-inferiority} \mid \mu_T - \mu_C = \delta^*\} = 1 - \beta$$

[i.e., $n = 2(z_\alpha + z_\beta)^2\sigma^2/(\delta + \delta^*)^2$].

In particular, if it is assumed that there is a loss in efficacy associated with using the new treatment instead of the control, so $\mu_T < \mu_C$, then $\delta^* < 0$ is appropriate. A popular choice is $\delta^* = 0$, which results in $n = 2(z_\alpha + z_\beta)^2\sigma^2/\delta^2$. Although this may be a reasonable choice, it will result in a sample size that is too small if the parameters satisfy $\mu_T < \mu_C$. The choice of $\delta^*$ should be driven by the true value of the difference (or the best preliminary estimate available), as discussed in Bristol (1995a). If it is expected that the new treatment is superior to the control, then the choice $n = 2(z_\alpha + z_\beta)^2\sigma^2/\delta^2$, corresponding to $\delta^* = 0$, will result in a conservative sample size.

**Normal means: confidence interval approach**

The confidence interval approach for the difference in the means of normal distributions is:

Declare non-inferiority if $L \geq -d$,

where $L = \bar{X}_T - \bar{X}_C - z_{\alpha/2}\sigma(2/n)^{1/2}$.

Thus, the probability of declaring non-inferiority is:

$$P\{L \geq -d\} = \Phi\left(\frac{0.5n}{\sigma} \sqrt{d + \mu_T - \mu_C}/\sigma - z_{\alpha/2}\right)$$

A common choice of $d$ is $d = \delta$, which requires all values of the difference consistent with the data, as defined by the confidence interval, to satisfy the non-inferiority criterion. In some cases, such as when the sample size is chosen by non-statistical constraints, $d$ may be chosen according to a probability constraint. Finally, $d$ may be specified by regulatory authorities.

The sample size may be chosen such that:

$$P\{\text{declare non-inferiority} \mid \mu_T - \mu_C = 0\} = 1 - \beta$$

[i.e., $n = 2(z_{\alpha/2} + z_\beta)^2\sigma^2/\delta^2$].
Note that for $d=\delta$, the difference between the sample sizes required for the hypothesis-testing approach and the confidence interval approach is that $z_\alpha$ is used in the former and $z_{\alpha/2}$ is used in the latter. This difference suggests that the $(1 - \alpha)100\%$ lower confidence limit should be used, to be consistent with the one-sided test procedure. Then the probability of declaring non-inferiority would be the same for the two approaches for all values of $\mu_T - \mu_C$. Furthermore, for $d=\delta$, the observed difference must satisfy

$$\bar{X}_T - \bar{X}_C \geq -\delta + z_{\alpha/2} \sigma (2/n)^{1/2}$$

to establish $\mu_T - \mu_C \geq -\delta$; this bound provides a description of the impact of the sample size and variance on the decision rule.
Methodology with Binomial Distributions

The problem of non-inferiority for binomial distributions is usually solved using the asymptotic normality. A major complication with this approach is the dependence of variance on the unknown probabilities of interest. For the two-sided problem of showing a difference, this problem was discussed by Robbins (1977) and Eberhardt and Fligner (1977) for the two-sample problem, and by Bristol (1993c) for multiple comparisons with a control. This problem can be overcome by using a procedure based on Fisher’s exact test or by basing the inference on the logit transformation.

Let $\pi_T$ and $\pi_C$ denote the probabilities of success for the new treatment and control, respectively, and let $p_T$ and $p_C$ denote the corresponding observed proportions of successes. Furthermore, let $\sigma^2 = \{(\pi_T(1 - \pi_T) + \pi_C(1 - \pi_C))/n \}$ denote the variance of the difference $p_T - p_C$.

**Binomial distributions: hypothesis-testing approach**

In this case, the hypothesis-testing approach is:

Test $H_0$: $\pi_T - \pi_C \leq -\delta$ against $H_1$: $\pi_T - \pi_C > -\delta$.

The rejection rule is:

Reject $H_0$: $\pi_T - \pi_C \leq -\delta$ in favor of $H_1$: $\pi_T - \pi_C > -\delta$ if

$$p_T - p_C \geq -d,$$

where $d$ satisfies $P\{p_T - p_C \geq -d \mid \pi_T + \delta = \pi_C = \pi \} = \alpha$.

Then $d = \delta - z_\alpha \sigma_1$, where $\sigma_1^2 = \{(\pi - \delta)(1 - (\pi - \delta)) + \pi(1 - \pi)/n \}$.

However, $\sigma_1$ is unknown. Thus, in practice, the rejection rule is:

Reject $H_0$: $\pi_T - \pi_C \leq -\delta$ in favor of $H_1$: $\pi_T - \pi_C > -\delta$ if

$$n^{\frac{1}{2}} \left(p_T - p_C + \delta\right)/\left\{\left(p - \delta\right)(1 - (p - \delta)) + p(1 - p)\right\}^{\frac{1}{2}} \geq z_\alpha,$$

where $p = 0.5(p_T + p_C)$ or $p$ is the maximum likelihood estimate of $\pi$ under $\pi_T + \delta = \pi_C = \pi$. This problem has been considered by many authors, such as Farrington and Manning (1990) and Dunnett and Gent (1977).
The common sample size can be chosen such that:

\[ P\{\text{declare non-inferiority} \mid \pi_T = \pi_C = \pi_i = 1 - \beta, \} = 1 - \beta, \]

so \( \phi(d/\sigma_0) = 1 - \beta \), where \( \sigma_0^2 = 2\pi(1 - \pi)/n \). Thus, the common sample size is:

\[
n = \left[ z_{\alpha/2}\left(\left(\pi - \delta\right)(1-(\pi - \delta)) + \pi(1-\pi)\right)^{1/2} + z_\beta\left(2\pi(1-\pi)\right)^{1/2} \right] / \delta^2.
\]

This sample size has been discussed by Makuch and Simon (1978), Makuch and Johnson (1986), and others.

**Binomial distributions: confidence interval approach**

The confidence interval approach for the difference of binomial success probabilities is:

Declare non-inferiority if \( L \geq -d \),

where \( L = p_T - p_C - z_{\omega/2}\left[p_T(1 - p_T) + p_C(1 - p_C)\right]^{1/2} / n^{1/2} \)

Because of the different variances, the hypothesis testing approach and the confidence interval approach result in different decision rules, even for \( d = \delta \). In particular, for the confidence interval approach:

\[ P\{\text{declare non-inferiority}\} = \phi([d + \pi_T - \pi_C] / \sigma - z_{\omega/2}) \]

The common sample size can be chosen such that:

\[ P\{\text{declare non-inferiority} \mid \pi_T = \pi_C = \pi_i = 1 - \beta, \} = 1 - \beta, \]

so \( \phi(d/\sigma) = 1 - \beta \), which results in \( n = 2(z_{\omega/2} + z_\beta)^2\pi(1-\pi)/d^2 \).

Because the probability of declaring non-inferiority increases as the success probabilities differ from 0.5, the FDA Division of Anti-Infective Drug Products (1992) proposed a choice of \( d \) that depends on the data:

\[
\begin{align*}
\text{if } p_C &\geq 0.90, \quad \delta = 0.10 \\
\text{if } 0.80 &\leq p_C < 0.90, \quad \delta = 0.15 \\
\text{if } p_C &< 0.80, \quad \delta = 0.20
\end{align*}
\]

Because the procedure is used for comparison to active controls, values are constant. Properties of this procedure are examined in Bristol (1996).
Other Issues

Any issue encountered in a comparative clinical trial conducted to observe a difference can be encountered in non-inferiority trials. Durrleman and Simon (1990), Jennison and Turnbull (1993a, 1993b), Bristol and Nezamis (1995), and others have considered the problem of interim analyses. Issues associated with multiple comparisons have been considered by Giani and Finner (1991), Giani and Strasburger (1994), Bristol (1994, 1995b), and others.

Issues involved in switching between superiority and non-inferiority endpoints are considered by:
European Agency for the Evaluation of Medicinal Products
*Evaluation of Medicines for Human Use:*
References


Spriet A, Bieler D. (1979). When can “non-significantly different” treatments be considered as “equivalent”? *Br J Clin Pharmacol* **7:**623-624.


DATA COLLECTION AND QUALITY CONTROL

During all phases of the study, sufficient effort should be made to ensure that all key data are of high quality.

Problems in Data Collection

- Incomplete or missing data

- Erroneous data. Sources of error include:
  - Using definitions different from those given in the study protocol
  - Mislabeled specimens
  - Badly calibrated equipment
  - Transcription errors when entering data on a form
  - Keypunch errors when entering data into computer

- Variability in the observed characteristics. Sources of variability include:
  - Inherent variability in the characteristic being measured
  - Instrument variability
  - Observer/technician variability (both between and within observer/technician)
  - Variability in reading x-rays, scans, histologic specimens, etc.

Example: Studies of Intra- and Inter-observer Disagreement

<table>
<thead>
<tr>
<th>Study</th>
<th>Reading of:</th>
<th>Percent Disagreement</th>
</tr>
</thead>
<tbody>
<tr>
<td>Intra-observer:</td>
<td></td>
<td></td>
</tr>
<tr>
<td>Davies et al. (1958)</td>
<td>Electrocardiograms</td>
<td>12.5%</td>
</tr>
<tr>
<td>Feinstein et al. (1970)</td>
<td>Lung cancer pathology</td>
<td>20%</td>
</tr>
<tr>
<td>Inter-observer:</td>
<td></td>
<td></td>
</tr>
<tr>
<td>Smyllie et al. (1965)</td>
<td>Respiratory signs</td>
<td>25%</td>
</tr>
</tbody>
</table>
Minimizing Poor Quality Data

- Establish clear definitions of entry and diagnostic criteria and methodology.
  - Have definitions on data forms if possible.
  - Forms should have references to manual pages for details.
  - Prepare a manual of procedures (in addition to the protocol), outlining procedures for the given coordinating center.

- Train investigators and staff to promote standardization of procedures.
  - Questions to patients should always be asked the same way. For example, “Have you had chest pain in the last 3 months?” is different from, “You haven’t been having chest pain, have you?”
  - Training session for technicians should include duplicate testing, discussion of differences, and the use of reference sets.
  - Institute certification procedures for certain types of data collection (e.g., pulmonary function tests, blood pressure readings).
  - Periodic re-training and certification may be necessary in long-term studies.
  - For patient interview data involving subjective or sensitive issues, special training is required.
TRIAL CLOSEOUT

The closeout of a clinical trial requires careful planning if it is to be accomplished in an orderly fashion.

Termination Procedures

1. Follow each patient for a fixed period of time.

![Study start to Study end diagram]

**Pros**
- Closeout is smooth, with patients tapering off as the study ends.
- Costs less than a trial with more extensive follow-up.
- In some studies, limited follow-up may be more appropriate.

**Cons**
- Unblinding usually occurs at a patient’s last visit. Unblinding early patients may inadvertently unblind others.
- Recommendations to early patients for post-trial treatment may be based on early trial results. Word could spread to others still on the trial.
2. Follow subjects to a common termination date.

<table>
<thead>
<tr>
<th>Study start</th>
<th>Study end</th>
</tr>
</thead>
</table>

**Pros**
- Avoids many problems listed above.
- Increased follow-up adds power to the treatment comprise.
- Easier to justify full support of staff until the end of the trial.

**Cons**
- Conducting a large number of “close-out” visits at the very end of the trial may be logistically difficult.
- The last visit is important, especially if the final response measure is obtained at that time.

**Data Storage**
The investigator should keep one set of documents (including the data). The institution should maintain a list of all subjects who participated in the trial. Subject data can be filed with individual medical records.

**Dissemination of Results**
Sequence of informing others of trial results:
Study leadership → other investigators → subjects (and referring physicians, if appropriate) → publications in the literature and press release to the public