



London, 27 July 2000
CPMP/EWP/482/99

**COMMITTEE FOR PROPRIETARY MEDICINAL PRODUCTS
(CPMP)**

**POINTS TO CONSIDER ON SWITCHING BETWEEN SUPERIORITY
AND NON-INFERIORITY**

DISCUSSION IN THE EFFICACY WORKING PARTY (EWP)	February 1999/ September 1999
TRANSMISSION TO THE CPMP	September 1999
RELEASE FOR CONSULTATION	September 1999
DEADLINE FOR COMMENTS	December 1999
RE-SUBMISSION TO THE EFFICACY WORKING PARTY	June 2000
APPROVAL BY THE CPMP	July 2000

Points to Consider have been developed to provide advice on selected areas relevant to the development of medicinal products in specific therapeutic fields.

This document will be revised in accordance with the scientific advances made in this area.

SWITCHING BETWEEN SUPERIORITY AND NON-INFERIORITY

I. INTRODUCTION

A number of recent applications have led to CPMP discussions concerning the interpretation of superiority, non-inferiority and equivalence trials. These issues are covered in ICH E9 (Statistical Principles for Clinical Trials). There is further relevant material in the Step 2 draft of ICH E10 (Choice of Control Group) and in the CPMP Note for Guidance on the Investigation of Bioavailability and Bioequivalence. However, the guidelines do not address some specific difficulties that have arisen in practice. In broad terms, these difficulties relate to switching from one design objective to another at the time of analysis.

The types of trials in question are those designed to compare a new product with an active comparator. The objective may be to demonstrate:

- the superiority of the new product
- the non-inferiority of the new product or
- the equivalence of the two products.

When the results of the trial become available, they may suggest an alternative interpretation. Thus the results of a superiority trial may only appear to be sufficient to support non-inferiority, while the results of a non-inferiority trial may appear to support superiority. Alternatively, the results of an equivalence trial may appear to support a tighter range of equivalence.

A satisfactory approach to this subject requires an understanding of confidence intervals and the manner in which they capture the results of the trial and indicate the conclusions that can be drawn from them. Such an understanding also leads to an appreciation of why power calculations are of relatively little interest when a trial is complete.

For simplicity, this paper addresses the issues of superiority, non-inferiority and equivalence from the perspective of an efficacy trial with a single primary variable. Some comments on other situations are made in Section VI. It is assumed throughout this document that switching the objective of a trial does not lead to any change in the selection or definition of the primary variable.

II. TRIAL OBJECTIVES

II.1 Superiority trial

A superiority trial is designed to detect a difference between treatments. The first step of the analysis is usually a test of statistical significance to evaluate whether the results of the trial are consistent with the assumption of there being no difference in the clinical effect of the two treatments. In a trial of good quality, the degree of statistical significance (p-value) indicates the probability that the observed difference - or a larger one - could have arisen by chance assuming that no difference really existed. The smaller this probability is, the more implausible is the assumption that there really is no difference between the treatments.

Once it is accepted that the assumption of "no difference" is untenable, it then becomes important to estimate the size of the difference in order to assess whether the effect is clinically relevant. This has two aspects. First there is the best estimate of the size of the difference between treatments (point estimate). For normally distributed data this is usually taken as the observed difference between the mean values on each. Next, there is the range of values of the true difference that are plausible in the light of the results of the trial (confidence interval). It is clear that this range should not include zero since the possibility of a zero

difference has already been rejected as unreasonable. The method of constructing confidence intervals generally ensures that this is so, provided it corresponds to the choice of significance test. Thus the following two statements are usually equivalent:

- The two-sided 95% confidence interval for the difference between the means excludes zero.
- The two means are statistically significantly different at the 5% level ($p < 0.05$) two-sided.

The above text addresses the situation where the difference between two mean values is the statistic of interest and a zero difference represents no effect. In practice a number of other summary statistics are used for the evaluation of differences between treatments, for example the odds ratio for proportions or the ratio of geometric means in bio-equivalence studies. (The latter arises from the logarithmic transformation used for bioavailability data.) In such cases the same principles apply but ‘no difference’ may be represented by a value other than zero - a value of 1 in both the examples quoted here. In these cases it is the position of the confidence interval for the test statistic relative to this ‘no difference’ value that is of interest.

When significance tests are carried out in practice, precise numerical values of probabilities are usually quoted, for example $p = 0.032$, because this is more informative than $p < 0.05$. This allows judgement to be based more precisely on the extent of the disagreement between the null hypothesis and the observed data rather than on the approximations implied by using cut-off points of 0.05, 0.01 and 0.001. However, confidence intervals have to be associated with a specific probability value (coverage probability) and this is nearly always taken as 95% (0.95). When a difference is statistically significant at a more extreme level, e.g. $p = 0.002$, the two-sided 95% confidence interval will exclude zero by a wider margin. Figure 1 illustrates these points.

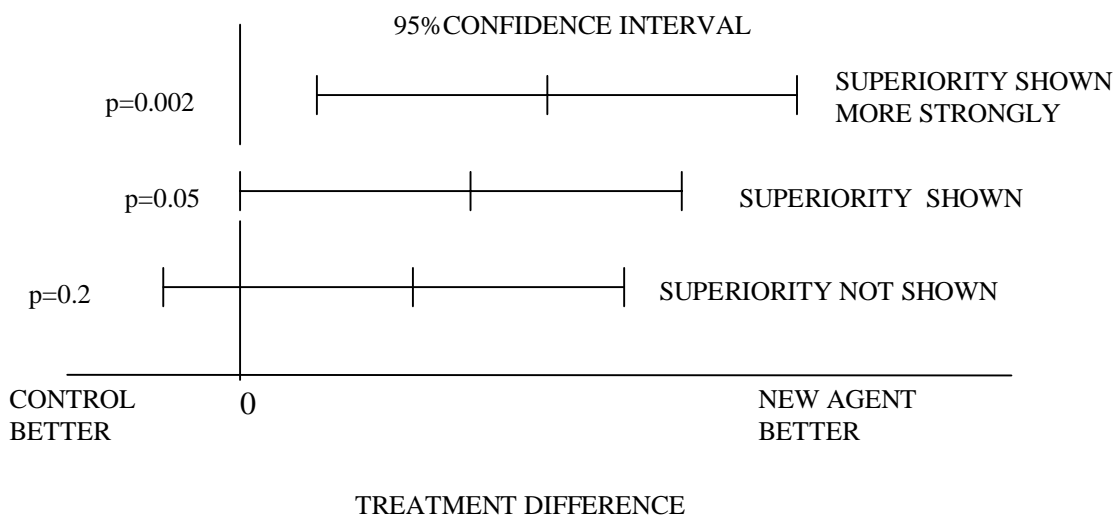


Figure 1: Relationship between significance tests and confidence intervals

Whether the observed difference is indeed clinically relevant is a matter of judgement. In contrast to an equivalence or non-inferiority trial where clinical relevance is addressed through the pre-study choice of Δ (see II.2 and II.3), in a superiority trial clinical relevance requires separate consideration: a statistically significant difference may not be clinically relevant. The difference taken as the basis of the power calculation in a superiority trial cannot be assumed to provide a suitable value.

Note that in Figure 1, and throughout the rest of the document, it is assumed that values to the right of zero correspond to a better response on the new treatment so that values to the left are worse i.e. better on the control treatment.

II.2 Equivalence trial

An equivalence trial is designed to confirm the absence of a meaningful difference between treatments. In this case it is more informative to conduct the analysis by means of the calculation and examination of the confidence interval although there are closely related methods using significance test procedures. (See also II.3.) A margin of clinical equivalence (Δ) is chosen by defining the largest difference that is clinically acceptable, so that a difference bigger than this would matter in practice. There are well-recognised difficulties associated with this task which will not be discussed in any detail here. If the two treatments are to be declared equivalent, then the two-sided 95% confidence interval - which defines the range of plausible differences between the two treatments - should lie entirely within the interval $-\Delta$ to $+\Delta$. See Figure 2. There are situations in which the equivalence margins may be chosen asymmetrically with respect to zero.

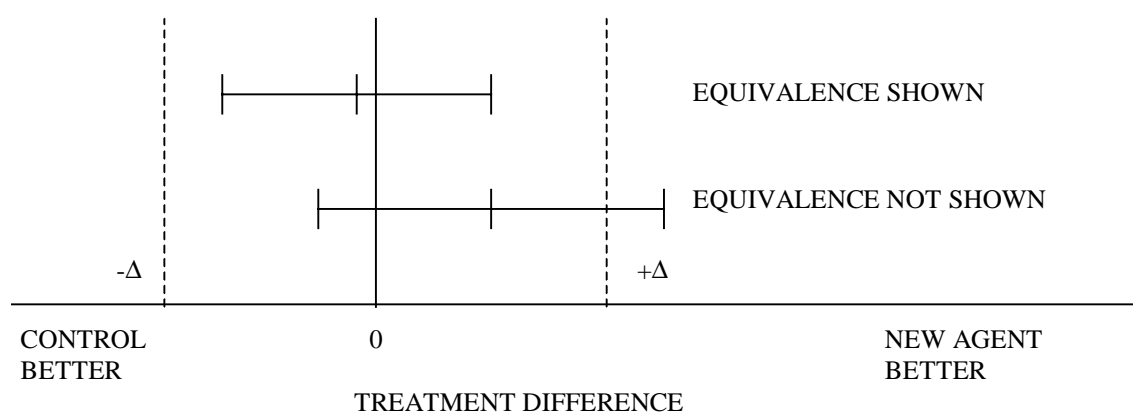


Figure 2: Confidence interval approach to analysis of equivalence trial.

In the case of bioequivalence studies a coverage probability of 90% for the confidence interval has become the accepted standard when evaluating whether the average values of the pharmacokinetic parameters of two formulations are sufficiently close.

Clinical equivalence trials, with two-sided 95% confidence intervals, may be carried out when conventional bio-equivalence trials are impossible, for example in the case of a generic inhaled or topically applied product.

II.3 Non-inferiority trial

In Phase III drug development, non-inferiority trials are more common than equivalence trials. In these we wish to show that a new treatment is no less effective than an existing treatment - it may be more effective or it may have a similar effect. Again a confidence interval approach is the most straightforward way of performing the analysis but now we are

only interested in a possible difference in one direction. Hence the two-sided 95% confidence interval should lie entirely to the right of the value $-\Delta$. See Figure 3. Non-inferiority trials are sometimes mistakenly referred to, and designed as, equivalence trials. This distinction is important and can be a source of confusion.

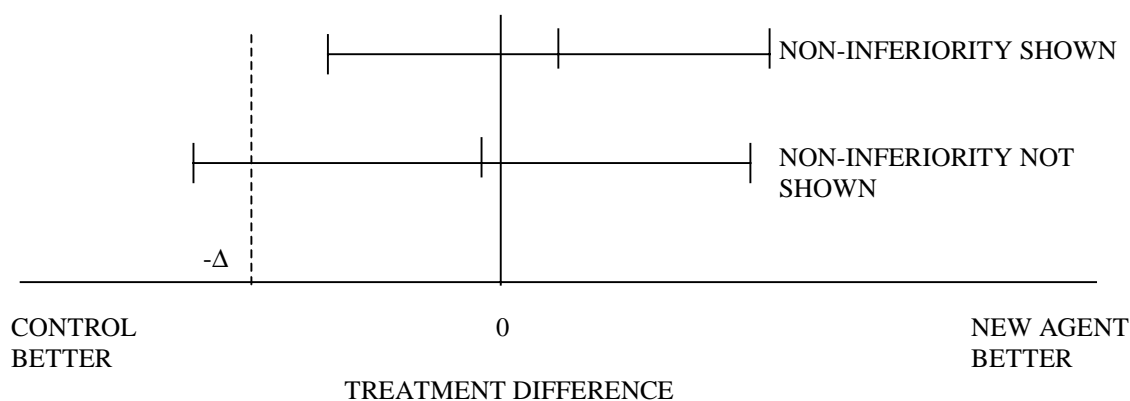


Figure 3: Confidence interval approach to analysis of non-inferiority trial

Note also that by using the closely related significance testing procedures referred to in II.2, it is possible to calculate a p-value associated with the null hypothesis of inferiority. This is a valuable further aid to assessing the strength of the evidence in favour of non-inferiority.

II.4 One-sided and two-sided confidence intervals

It will be assumed throughout this document that two-sided 95% confidence intervals are to be used for all clinical trials whatever their objective. Among other benefits, this preserves consistency between significance testing and subsequent estimation. It is also consistent with the guidance provided in the ICH E9 Note for Guidance. If one-sided intervals are used, then they should be used with a coverage probability of 97.5%.

In the special case of bioequivalence studies, two-sided 90% confidence intervals have been established as the norm as recommended, for example, in the CPMP Note for Guidance on the Investigation of Bioavailability and Bioequivalence.

III. RELEVANCE OF PRE-DEFINITION

A conclusion of equivalence or non-inferiority clearly depends upon the value of Δ chosen as the maximum acceptable difference. It is always possible to choose a value of Δ which leads to a conclusion of equivalence or non-inferiority if it is chosen after the data have been inspected. Since the choice of Δ is generally a difficult one, there is ample room for bias here, however well intentioned the researcher may be. Plausible arguments may often be advanced for a retrospective choice. In the design of equivalence and non-inferiority trials, this reason (amongst others) makes it necessary for the choice of Δ , and the reasoning behind the choice, to be set down in advance by the researcher in the study protocol. The corresponding coverage probability for the confidence interval (usually 95%) should also be chosen at this time. (See Section IV.2 for how these requirements apply when objectives are changed.) The question of how to choose an appropriate Δ will be addressed in a subsequent CPMP Points to Consider.

Pre-definition of a trial as a superiority trial, an equivalence trial or a non-inferiority trial is necessary for numerous reasons including the following:

- to ensure that comparator treatments, doses, patient populations and endpoints are appropriate (see ICH E10)
- to allow sample size estimates to be based on the correct power calculations
- to ensure that equivalence and non-inferiority criteria are pre-defined
- to permit appropriate analysis plans to be described in the protocol
- to ensure that the trial has sufficient sensitivity to achieve its objectives (see ICH E10)

If the objective of a trial is switched from superiority to non-inferiority, or vice versa, these aspects may lead to greater difficulty than the interpretation of significance tests and confidence intervals.

IV. SWITCHING THE OBJECTIVE OF THE COMPARISON

The only switching which is likely to have any practical relevance is switching between superiority and non-inferiority. The place of equivalence trials is so specific that they stand alone.

IV.1 Interpreting a non-inferiority trial as a superiority trial

If the 95% confidence interval for the treatment effect not only lies entirely above $-\Delta$ but also above zero then there is evidence of superiority in terms of statistical significance at the 5% level ($p < 0.05$). See Figure 4. In this case it is acceptable to calculate the p-value associated with a test of superiority and to evaluate whether this is sufficiently small to reject convincingly the hypothesis of no difference. There is no multiplicity argument that affects this interpretation because, in statistical terms, it corresponds to a simple closed test procedure. Usually this demonstration of a benefit is sufficient on its own, provided the safety profiles of the new agent and the comparator are similar. When there is an increase in adverse events, however, it is important to estimate the size of the effect to evaluate whether it is sufficient in clinical terms to outweigh the adverse effects.

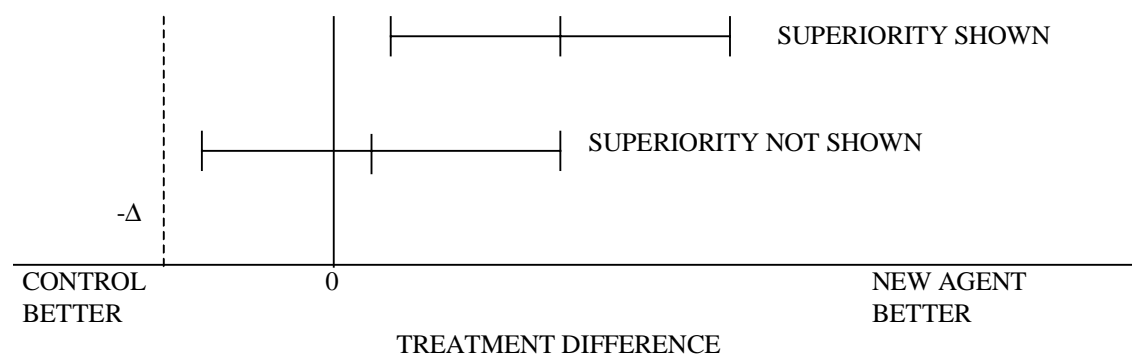


Figure 4: Non-inferiority to superiority

There are a number of other factors that might be affected by this changed objective.

IV.1.1 Appropriateness of comparator

If the comparator was suitable for a demonstration of non-inferiority, then there should be well-controlled data to show that it is an effective treatment. Hence, for proof of efficacy, a clear demonstration of superiority to the comparator in terms of statistical significance should be acceptable.

IV.1.2 Power calculations

Non-inferiority trials are generally large because of their need to exclude the possibility of a small degree of inferiority of a new agent relative to an active control. However if the new agent is actually superior to control by a small amount, then the power to show its non-inferiority is increased. Demonstrating the small amount of superiority to control might in principle require the planning of an even larger trial. When the trial is completed, however, the results provided by the confidence interval supply a concrete assessment of the precision actually achieved, superseding any calculations of power carried out before the trial was undertaken.

IV.1.3 Size of additional clinical benefit

Since the comparator in a non-inferiority trial must be an effective agent, any superiority to that agent should carry the implication of acceptable superiority to no treatment (placebo). For this reason the size of the additional clinical benefit demonstrated is not likely to be relevant to a claim of efficacy except in relation to any increase in adverse effects and hence relative risk/benefit. However, when the proposed licence includes a claim of superiority to the comparator, the size of the additional benefit should be discussed in clinical terms.

IV.1.4 Choice of analysis set

In a superiority trial the full analysis set, based on the ITT (intention-to-treat) principle, is the analysis set of choice, with appropriate support provided by the PP (per protocol) analysis set. In a non-inferiority trial, the full analysis set and the PP analysis set have equal importance and their use should lead to similar conclusions for a robust interpretation. A switch of objective would require this difference of emphasis to be recognised. More details of the relative importance of these two analysis sets in superiority and non-inferiority trials can be found in the ICH E9 Note for guidance.

IV.1.5 Trial quality

A trial to show equivalence or non-inferiority must show a high degree of consistency with protocolled plans if it is to be reliable. Deviations from the inclusion criteria, from the intended treatment regimen, from the schedule, manner and precision of taking measurements, and so on, all tend to reduce the sensitivity of a trial and to make a conclusion of 'no difference' more likely, even when the deviations are of an unsystematic or random nature. The size of the bias associated with these and other departures from the protocol is generally unknown and may render such a trial uninterpretable. Failure to show a difference between two treatments can also arise when both treatments are inefficacious, perhaps as a result of being inappropriately administered. This problem does not affect superiority trials to the same extent because the demonstration of a difference is itself validation of the sensitivity of the trial. The estimate of the size of the effect may however be similarly affected. For these reasons, switching from non-inferiority to superiority is likely to carry with it a greater degree of confidence in the conclusion.

IV.1.6 Conclusion

Switching the objective of a trial from non-inferiority to superiority is feasible provided:

- The trial has been properly designed and carried out in accordance with the strict requirements of a non-inferiority trial.
- Actual p-values for superiority are presented to allow independent assessment of the strength of the evidence.
- Analysis according to the intention-to-treat principle is given greatest emphasis.

IV.2 Interpreting a superiority trial as a non-inferiority trial

If a superiority trial fails to detect a significant difference between treatments, there may be interest in the lesser objective of establishing non-inferiority. If the results of the superiority trial are summarised by means of a 95% confidence interval for the treatment difference, the lower end of that confidence interval provides a quantitative estimate of the minimum estimated effect of the new treatment relative to the comparator. When the study protocol contains an acceptable, prospectively defined margin $-\Delta$ for non-inferiority, downgrading the objective presents less methodological problems. See Figure 5. Pre-defining a non-inferiority margin would appear only to make sense in trials with non-inferiority as an aim. However, in any superiority trial where non-inferiority may be an acceptable outcome for licensing purposes, it is prudent to specify a non-inferiority margin in the protocol in order to avoid the serious difficulties that can arise from later selection. Under such circumstances it is also prudent to pay attention during design to the possible later need to justify that the study had sufficient sensitivity to detect the drug effects of interest (see IV.2.4). It is important to note that there are areas of medicine where non-inferiority to an active control is unlikely to be acceptable as the sole or main evidence of efficacy, and placebo controlled trials are required.

In trials where there is no prospectively defined non-inferiority margin, such a quantity has to be justified after the event and in many situations this will not be possible. It is likely that the justification will have to be written after the results have been seen and there may be little external basis for an objective choice of margin.

Although there does not appear to be a statistical multiplicity issue *per se* related to this switch of objective, that does not diminish the difficulties associated with the post hoc definition of Δ .

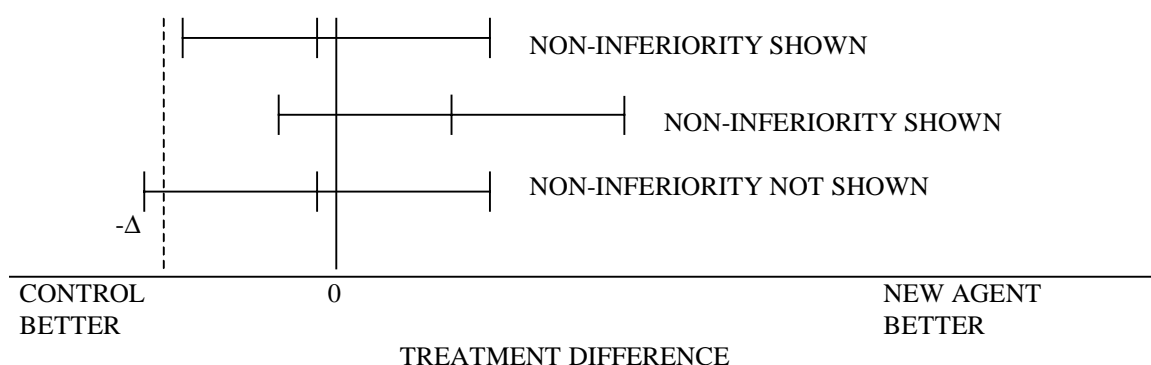


Figure 5: Superiority to non-inferiority.

A number of other issues require comment:

IV.2.1 Appropriateness of comparator, doses, patient population and endpoints

A comparator chosen for a demonstration of superiority may not be acceptable for a conclusion of non-inferiority. In order for it to be acceptable, it will be necessary to establish that there are data from good quality controlled superiority trials showing consistent evidence that the comparator is an effective treatment with reproducible effects, and establishing the size of its effect relative to no treatment. There should also be a reasonable basis for expecting that the same degree of efficacy would be shown in the current trial. For example, the patient population and the endpoints should be similar. These issues are covered fully in ICH E10.

IV.2.2 Power calculations

As indicated in IV.1.2, the results provided by the confidence interval supply a concrete assessment of the precision actually achieved by a clinical trial, superseding any calculations of power carried out before the trial was undertaken. The position of the lower end of the confidence interval relative to the agreed criterion of non-inferiority provides the key information for making decisions about non-inferiority.

IV.2.3 Choice of analysis set

In a superiority trial the full analysis set, based on the ITT (intention-to-treat) principle, is the analysis set of choice, with appropriate support provided by the PP (per protocol) analysis set. In a non-inferiority trial the full analysis set and the PP analysis set have equal importance and their use should lead to similar conclusions for a robust interpretation. A switch of objective would require this difference of emphasis to be recognised. More details of the relative importance of these two analysis sets in superiority and non-inferiority trials can be found in the ICH E9 Note for Guidance.

IV.2.4 Trial quality

A trial to show equivalence or non-inferiority must show a high degree of consistency with protocolled plans if it is to be reliable. Deviations from the inclusion criteria, from the intended treatment regimen, from the schedule, manner and precision of taking measurements, and so on, all tend to reduce the sensitivity of a trial and to make a conclusion of 'no difference' more likely, even when the deviations are of an unsystematic or random nature. The size of the bias associated with these and other departures from the protocol is generally unknown and may render such a trial uninterpretable. Failure to show a difference between two treatments can also arise when both treatments are inefficacious, perhaps as a result of being inappropriately administered. This problem does not affect superiority trials to the same extent because the demonstration of a difference is itself validation of the sensitivity of the trial. For these reasons, switching from superiority to non-inferiority is likely to carry with it a lesser degree of confidence in the conclusions. It will be necessary to pay particular attention to demonstrating the sensitivity of the trial by

- finding evidence, direct or indirect, that the control treatment is showing its usual efficacy
- comparing the trial with earlier trials which demonstrated the efficacy of the control agent
- in particular demonstrating levels of non-compliance and of loss of patients and data that are at least comparable to those in the earlier trials
- showing similar results from the full analysis set and PP analysis set.

IV.2.5 Conclusion

Switching the objective of a trial from superiority to non-inferiority may be feasible provided:

- The non-inferiority margin with respect to the control treatment was pre-defined or can be justified. (The latter is likely to prove difficult and to be limited to rare cases where there is a widely accepted value for Δ .)
- Analysis according to the intention-to-treat principle and PP analysis, showing confidence intervals and p-values for the null hypothesis of inferiority, give similar findings.
- The trial was properly designed and carried out in accordance with the strict requirements of a non-inferiority trial (see ICH E9 and E10).

- The sensitivity of the trial is high enough to ensure that it is capable of detecting relevant differences if they exist.
- There is direct or indirect evidence that the control treatment is showing its usual level of efficacy.

V. CHANGING THE EQUIVALENCE MARGINS

A further related issue that has arisen in connection with equivalence and non-inferiority trials concerns making changes to the equivalence margins when the trial is complete. Let us suppose that a bioequivalence trial finds a 90% confidence interval for the relative bioavailability of a new formulation that ranges from 0.90 to 1.15. Can we only conclude that the relative bioavailability lies between the conventional limits of 0.80 and 1.25 because these were the pre-defined equivalence margins? Or can we conclude that it lies between 0.90 and 1.15?

The narrower interval based on the actual data is the appropriate one to accept. Hence, if the regulatory requirement changed to $\pm 15\%$, this study would have produced satisfactory results. There is no question here of a data-derived selection process.

However, if the trial had resulted in a confidence interval ranging from 0.75 to 1.20, then a post hoc change of equivalence margins to $\pm 25\%$ would not be acceptable because of the obvious conclusion that the equivalence margin was chosen to fit the data.

These considerations apply equally to the 95% confidence intervals used for clinical equivalence and for non-inferiority. The confidence interval based on the results of the trial is always the best summary of the outcome. It is the choice of equivalence margin that is subject to bias. This should be chosen on the basis of external information and not chosen to fit the data.

VI. DISCUSSION

This Points to Consider has been written from the perspective of an efficacy trial comparing active agents with a single primary variable. In practice some studies have more than one primary variable, and most studies have secondary variables. With respect to switching objectives, each of these variables requires separate consideration in the context of the specific drug development, drawing separate conclusions regarding superiority or non-inferiority for each variable in turn. Any overall judgement about whether the trial as a whole has established the superiority or non-inferiority of the new treatment will depend upon the regulatory requirements for that clinical area and the pattern of results across all relevant variables.

The concepts covered in these Points to Consider can also be applied to specific safety variables when these have been pre-identified as major endpoints of a trial to compare active agents.

In practice the issue of switching objectives is not relevant to placebo controlled trials, even where non-inferiority to placebo is a valuable outcome i.e. for safety variables.

The problem of switching objectives *post hoc* can be avoided by designing a trial prospectively in the knowledge that both non-inferiority and superiority are outcomes of potential value. In this case all the issues identified in this document should be addressed during design. In particular the statistical analysis should be conducted using an appropriate stepwise procedure, progressing from non-inferiority to superiority.

VII. OVERALL CONCLUSIONS

The interpretation of superiority trials as non-inferiority trials and vice versa is best approached by expressing the results as a confidence interval for the difference between the test treatment and control. There is no fundamental problem associated with the use of this confidence interval as a basis for either mode of interpretation. For a well-designed and conducted trial, there are few difficulties connected with the change from non-inferiority to superiority that cannot be addressed by appropriate analysis. However, there are more severe difficulties associated with the switch from superiority to non-inferiority because of the possible need to find a basis for, and agree on, a margin of equivalence after seeing the outcome, and because of the inherent difficulties of non-inferiority trials. There are consequences for the design of a superiority trial in which non-inferiority might be an acceptable outcome.

When examining the results with respect to alternative choices of the equivalence margins during interpretation, the main problem derives from attempts to switch to wider acceptable margins. Data that satisfy narrower equivalence margins may be safely interpreted in this manner.