



**COMMITTEE FOR MEDICINAL PRODUCTS FOR HUMAN USE
(CHMP)**

GUIDELINE ON THE CHOICE OF THE NON-INFERIORITY MARGIN

DRAFT AGREED BY THE EFFICACY WORKING PARTY	December 1999 – January 2004
ADOPTION BY COMMITTEE FOR RELEASE FOR CONSULTATION	February 2004
END OF CONSULTATION (DEADLINE FOR COMMENTS)	May 2004
AGREED BY WORKING PARTY	June 2004
ADOPTION BY COMMITTEE	July 2005
DATE FOR COMING INTO EFFECT	January 2006

GUIDELINE ON THE CHOICE OF THE NON-INFERIORITY MARGIN

TABLE OF CONTENTS

INTRODUCTION.....	3
1. BACKGROUND.....	4
2. GENERAL CONSIDERATIONS	4
3. DEMONSTRATING EFFICACY.....	5
4. ESTABLISHING ACCEPTABLE EFFICACY RELATIVE TO THE ACTIVE COMPARATOR.....	8
5. AREAS WHERE IT IS DIFFICULT TO JUSTIFY A NON-INFERIORITY MARGIN OF ANY SIZE.....	9
6. CONCLUSIONS.....	10

INTRODUCTION

Many clinical trials comparing a test product with an active comparator are designed as non-inferiority trials. The term 'non-inferiority' is now well established, but if taken literally could be misleading. The objective of a non-inferiority trial is sometimes stated as being to demonstrate that the test product is not inferior to the comparator. However, only a superiority trial can demonstrate this. In fact a non-inferiority trial aims to demonstrate that the test product is not worse than the comparator by more than a pre-specified, small amount. This amount is known as the non-inferiority margin, or delta (Δ).

There are many situations where a non-inferiority trial might be performed as opposed to, or in addition to, a superiority trial over placebo. These include:

- Applications based upon essential similarity in areas where bioequivalence studies are not possible, e.g. modified release products or topical preparations;
- Products with a potential safety advantage over the standard might require an efficacy comparison to the standard to allow a risk-benefit assessment to be made;
- Cases where a direct comparison against the active comparator is needed to help assess risk-benefit;
- Cases where no important loss of efficacy compared to the active comparator would be acceptable;
- Disease areas where the use of a placebo arm is not possible and an active control trial is used to demonstrate the efficacy of the test product.

In the final 4 situations above a non-inferiority trial would not be necessary if superiority could be shown over the reference product.

In order to demonstrate non-inferiority, the recommended approach is to pre-specify a margin of non-inferiority in the protocol. After study completion, a two-sided 95% confidence interval (or one-sided 97.5% interval) for the true difference between the two agents will be constructed. This interval should lie entirely on the positive side of the non-inferiority margin. The choice of delta must always be justified on both clinical and statistical grounds. It always needs to be tailored specifically to the particular clinical context and no rule can be provided that covers all clinical situations. However, certain principles can be used to provide general guidance.

The following regulatory guidelines make reference to the choice of the margin of non-inferiority or equivalence. They should be read in conjunction with this Guideline.

- ICH Note for Guidance E9 (Statistical Principles for Clinical Trials);
- ICH Note for Guidance E10 (Choice of Control Group);
- CPMP Points to Consider on Switching Between Superiority and Non-inferiority.

In these documents the discussion of how to choose the margin of non-inferiority is limited. They do, however, make detailed comments regarding the design and conduct of studies designed to demonstrate non-inferiority. Such issues are extremely important, and if a trial has not been conducted to an appropriately high standard, the choice of delta can become an irrelevant issue.

This document will consider two types of non-inferiority trials: trials with two arms, the test product and a comparator; and three-armed trials with the test product, an active comparator and placebo.

There are many aspects of the performance of an experimental product to consider. Broadly these relate to efficacy and safety, but each of these broad categories can be broken down for an individual product into many points of interest. A clinical trial or clinical programme may plan to show non-inferiority for certain variables while superiority may be the objective for others. In this document the terms 'non-inferiority' and 'superiority' are used relating to single endpoints and not to the product profile as a whole.

It is assumed throughout that the effect of the treatments can be measured and that the measurements make it possible to distinguish between desired (positive) and undesired (negative) effects. It is further assumed that large positive values in the measured variable point to large positive effects.

The majority of the document uses the example of the absolute difference between treatments to illustrate the ideas. The discussion is also applicable to studies considering a relative effect with a few modifications. For example in a trial considering relative effects, no difference between treatments is reflected by a point estimate of one, as opposed to a difference of zero.

Efficacy parameters are used to illustrate the methods mentioned in this document, although non-inferiority margins can be defined for safety parameters as well. However, a lot of the discussion for efficacy end-points will not apply to safety trials, most notably the whole of section III.

1. BACKGROUND

The outcome of a non-inferiority trial is usually assessed by a two-sided 95% confidence interval, showing a credible range for the true difference between the test product (test: T) and the active comparator (reference: R). There are two aspects of the results that should attract particular attention. One is the point estimate of the difference, i.e. the observed difference between test and reference. The other is the lower limit of the confidence interval. The point estimate represents the best estimate of the true difference, so that if it is positive and if this is all the evidence available, it is more likely that the test product is better than the reference and vice-versa. The lower limit of the confidence interval, on the other hand, represents a lower bound and is usually interpreted as the degree of inferiority to the reference that can be excluded based on the data presented. Of course this is not an actual lower bound and the magnitude of inferiority could be greater. However it is generally considered that the chance of the true difference being worse than that suggested by this bound is acceptably small.

If T and R were equally efficacious, then the point estimate for the difference would have a 50% chance of being positive and a 50% chance of being negative, regardless of sample size. Hence the point estimate alone is not sufficient as an indicator of relative efficacy. The lower confidence limit for the difference, in the situation of true equality, would be expected to move closer to zero as the sample size increased, thus making it theoretically possible to rule out any desired degree of possible inferiority by using sufficiently large samples. However, if the treatments truly are equally efficacious, it is not possible to design a study to rule out all degrees of inferiority as this would require an infinitely large experiment.

Thus, it should be made clear at the outset that if no degree of possible inferiority of T to R is acceptable, then the development of products with equal efficacy to a comparator by means of non-inferiority trials would become impossible.

2. GENERAL CONSIDERATIONS

- The selection of the non-inferiority margin is based upon a combination of statistical reasoning and clinical judgement.
- A three-armed trial with test, reference and placebo allows some within-trial validation of the choice of non-inferiority margin and is therefore the recommended design; it should be used wherever possible.
- An appropriate choice of margin should provide assurance that the test drug has a clinically relevant effect greater than zero. This aspect of the choice of margin is discussed in section III.
- Usually the primary focus of a non-inferiority trial is the relative efficacy of the test and reference products, not simply demonstration that the test product has an effect. In these cases an appropriate choice of margin will, in addition to proving that the product has an effect, also provide assurance that the test product is not substantially inferior to the reference, resulting in a tighter margin. This aspect of the choice of margin is discussed in section IV.

- For the majority of non-inferiority trials it must be demonstrated that the margin satisfies both the requirements of section III and section IV. The choice of non-inferiority margin should be justified in the study protocol, and the justification should address the considerations of both sections.
- It is not appropriate to define the non-inferiority margin as a proportion of the difference between active comparator and placebo. Such ideas were formulated with the aim of ensuring that the test product was superior to (a putative) placebo; however they may not achieve this purpose. If the reference product has a large advantage over placebo this does not mean that large differences are unimportant, it just means that the reference product is very efficacious.
- It is not appropriate to use effect size (treatment difference divided by standard deviation) as justification for the choice of non-inferiority margin. This statistic provides information on how difficult a difference would be to detect, but does not help justify the clinical relevance of the difference, and does not ensure that the test product is superior to placebo.
- The choice of margin should be independent of considerations of power. It should be based upon the clinical and statistical principles noted in later sections of this document and not upon issues of sample size, as the size of the clinically important difference is not altered by the size of the study. A small study is not a justification for a wider non-inferiority margin.
- If an appropriate margin has been chosen, a confidence interval that lies entirely between $-\Delta$ and 0 (i.e. the test product is inferior to the reference, but not more than Δ worse) is still adequate to demonstrate non-inferiority. If this outcome does not seem acceptable, this demonstrates that Δ has not been chosen appropriately. (See also section V for discussion of situations where it is difficult to justify any amount of inferiority.)
- It is not possible to perform a non-inferiority trial in all situations. The decision to perform a non-inferiority trial should be justified considering both the therapeutic area and the profile of the reference product.
- There are many conditions where established effective agents do not consistently demonstrate superiority in placebo controlled trials (e.g. depression or allergic rhinitis). In areas where this lack of sensitivity exists, a non-inferiority trial which does not also include a placebo arm is not appropriate. See ICH E10 for a fuller discussion of assay sensitivity.
- If the performance of the reference product in a trial is very different from what was assumed when defining the non-inferiority margin then the chosen margin may no longer be appropriate. The implications of this should be considered at the planning stage.

3. DEMONSTRATING EFFICACY

A minimal requirement for the decision making process involved in interpreting data from a non-inferiority trial is that we must be confident that the test product would have been shown to be efficacious if a placebo controlled trial had been performed. The discussion in this section takes the methods commonly used when interpreting data from superiority trials as a model for assessing the minimal requirements for data from non-inferiority trials.

When data from trials designed to show superiority of a test product over placebo are being interpreted, an informal two-stage procedure is employed involving the consideration of both statistical significance and clinical relevance. The same two-stage procedure can be used for interpreting non-inferiority trials. In a superiority trial, it would first be expected that the test product demonstrated a statistically significant advantage over placebo. This relates to the ‘statistical reasoning’ stage of the ICH E10 combination of ‘both statistical reasoning and clinical judgement’. Statistical significance is generally assessed using the two-sided 0.05 level of significance (or one-sided 0.025). An alternative way of stating this requirement is that the lower bound of the two-sided 95% confidence interval (or one-sided 97.5% interval) for the difference between active and placebo should be above zero.

The next step in interpreting a superiority trial is to consider whether the difference from placebo is clinically relevant. This is the 'clinical judgement' stage of the ICH E10 combination of 'both statistical reasoning and clinical judgement'.

Establishing a clinically relevant benefit over placebo is accomplished by considering the point estimate of the difference between the test product and placebo and assessing its clinical relevance, either using the original scale or by considering responder rates. This is not primarily a statistical issue, but does require an intelligent combination of clinical thinking and data comprehension. Statistical significance has already been demonstrated, so the existence of an effect is considered to be established. A judgement must be made regarding whether the difference seen is clinically useful. This judgement is usually made in the context of the safety profile via an assessment of benefit/risk.

3.1 Three arm trials: test, reference and placebo

This trial design makes it possible to provide a direct demonstration of the superiority of the test and reference products over placebo. As such, it is not necessary to define a value for delta to establish that the test product has efficacy, however the following considerations should be addressed in the interpretation of the trial data.

As in the placebo controlled superiority trial, the test product must demonstrate a statistically significant advantage over placebo. The lower bound of the 95% confidence interval for the difference between the test product and placebo should be above zero.

At this stage the performance of the reference arm is not the main consideration, although if the test and reference products both fail to demonstrate a statistically significant advantage over placebo this could suggest that the trial is insensitive, or lacks assay sensitivity.

As in a superiority trial, clinical judgement is then applied to assess whether the observed difference from placebo is clinically relevant. The existence of the reference arm can assist in making this judgement. If the reference product is a licensed agent that is known to regularly produce a clinically relevant effect in trials of this type, the reference product difference from placebo seen in this trial can be used to help assess the clinical relevance of the difference between placebo and the test product. For example, if the test arm has performed better than the reference arm in the trial, it seems reasonable to assume that the test product's benefit is clinically relevant.

If the reference product has not demonstrated statistical significance over placebo, or has performed very differently to how experience would lead us to expect, questions could be raised about the performance of the reference product in this trial. In this situation the results from the reference arm could not provide a context, and any positive results from the test drug would have to stand alone. Possible reasons for the unexpected results from the reference treatment should be discussed.

3.2 Two arm trials: test and reference

As there is no placebo arm in this type of trial, indirect comparisons to placebo via previous studies comparing reference to placebo must be used to establish that the product has efficacy. This presents inherent difficulties, and necessitates that a non-inferiority margin is defined. However, the lower-bound of a 95% confidence interval can still be used to establish an efficacy advantage over placebo. The term 'putative placebo' is often used in this situation where no placebo has actually been used.

A systematic review should be conducted to identify studies relevant to the comparison of the reference treatment with placebo in the condition being considered. These can be used for estimating the difference between the reference and placebo in the intended patient population. If such estimation is not possible, or if the comparator did not consistently demonstrate superiority over placebo in adequately powered trials, the sensitivity of a non-inferiority study using this comparator may be questioned and only superiority of the test product to a comparator (active or placebo) would be interpretable. If the reference product is part of a class where individual products are all felt to be equally effective and safe it might be acceptable to use the overall class difference from placebo.

There are several issues regarding the literature search that will need to be discussed by the applicant. (See also ICH E10 for further discussion):

1. Selection bias. The criteria used for selecting which of the available studies to include should be thoroughly documented so that it is clear that, as far as is possible, an unbiased selection of studies was made.
2. Constancy of trial design and clinical practice over time. Some of the studies may be of little relevance because clinical practice may have changed, or the criteria or methods for measuring the reference product's effect have been modified. Consideration should be given to the design of the current trial in comparison to the previous trials regarding changes that may affect treatment outcome. Examples include entry criteria, method of diagnosis, concomitant treatments allowed, dosing regime of reference product, endpoints measured, timing of assessments, etc. If possible the design of the current trial should closely match the well-designed previous trials comparing the reference with placebo. If there are unavoidable differences in trial design the implications of this should be carefully considered, and it may not be possible to formulate a non-inferiority margin.
3. Constancy of effects over time. Consideration should be given to changes in the treatment difference seen over time. For example in some conditions event rates may have decreased over time because of general improvements in healthcare. In this situation it might be appropriate to include only the more recent studies in the calculations. If constancy of effect from recent trials to the current trial cannot be assured then a conservative approach to selecting a margin should be considered.
4. Publication bias. It may be that studies with a 'positive' outcome are more likely to be published than those with disappointing results. If it seems possible that such publication bias exists, a conservative approach should be taken in producing the historical confidence interval for reference versus placebo.

If good historical data are available, several methods exist that can be used to provide a non-inferiority margin. Common to all methods is an attempt to combine the variability and size of effect from the historical data with those expected from the current trial. Also common to all the methods are the weaknesses inherent in using historical data.

The 'historical' confidence interval compares the reference product with placebo (r minus p). The planned trial comparing the test and reference products will also produce a confidence interval (for t minus r). If these intervals are combined, an indirect confidence interval comparing the test product and placebo can be obtained (t minus p). Delta can be defined as the lower bound of t minus r that ensures that the lower bound of the indirect confidence interval of t minus p will be above zero. As the comparison is indirect it might be wise to be conservative and select some value smaller than that suggested by this indirect calculation.

In a submission the applicant should present both the direct confidence interval T minus R and the indirect interval T minus P .

Once statistically significant efficacy has been established clinical relevance should be considered. The point estimate from the indirect confidence interval for t minus p should be noted and the clinical relevance of that difference considered.

As with 3-arm trials the reference arm can supply some context. If the test product outperforms the reference, this provides some assurance of the clinical relevance of the difference. However, if at the time of analysis the data lead to doubts about the performance of the reference arm in the trial, the non-inferiority margin selected may seem inappropriate and the validity of the trial may be questioned. In the absence of a placebo arm it is more difficult to validate the performance of the reference treatment and historical data will be necessary to show that the performance in this trial is consistent with expectations.

Justification of whether the observed difference is considered to be clinically beneficial should include reference to other trials in the same therapeutic area where clinically relevant results were seen, and just as importantly trials where the results were not considered clinically relevant.

3.3 Conditions where superiority over placebo has not been reliably established

In some disease areas placebo controlled trials may be considered unethical, yet no available comparator has reliably demonstrated efficacy over placebo. Examples include some oncology settings and some orphan indications. In such situations it will be difficult to specify delta using the considerations outlined above, however the best efforts should still be made to produce an indirect confidence interval for the new product against placebo using whatever data exist for the reference. If there are no data, the reference could be treated as placebo and efficacy could be established by demonstrating direct superiority over the reference.

It is likely that indirect confidence intervals constructed in these circumstances would fail to demonstrate superiority over placebo, but in such conditions this might not necessarily preclude a positive opinion. The delta derived using the methods of section III.2 should not be discarded, as this is the value which signifies indirect superiority over placebo. Using this value means that any decision (whether positive or negative) will be made acknowledging the limitations of the data for demonstrating efficacy, i.e. we cannot be sure the test treatment is superior to placebo. It would not be good practice to define an arbitrary achievable delta and use that to claim non-inferiority. Such an approach would create a false impression of the confidence we can have in the efficacy of the product. It is important that the basis upon which a decision is being made is clear, whether that basis is weak or strong.

Similarly there may be situations where following the guidance of previous sections will lead to a small value of delta which cannot be excluded with a feasibly sized trial. Again it is not good practice to simply define a larger delta and then claim non-inferiority. Rather, it should be acknowledged that it has not been possible to demonstrate superiority to placebo. In some areas decisions are made based upon only small amounts of data without the demonstration of efficacy being clear, but it is important that everybody is aware that this is what is being done, and that decisions are made with full awareness of the risks being taken.

4. ESTABLISHING ACCEPTABLE EFFICACY RELATIVE TO THE ACTIVE COMPARATOR

Establishing that the new active compound would have been successful in a placebo-controlled trial is necessary but it will not usually be sufficient. The comparison between test and reference will often be of importance in its own right.

In this respect it is important to define objectives before starting the trial. The design and analysis of a non-inferiority trial should reflect the question the trial is aiming to address. There are many different reasons why such a trial might be conducted, and the objective for running the trial should influence the choice of delta. Trials are generally labelled non-inferiority trials if they are not aiming to show superiority over the reference. However, 'demonstrating non-inferiority' is not considered to be a sufficiently detailed objective for a trial. A lot of clarity can be gained if more precise aims are described.

If the only objective is to show indirect superiority over placebo, this should be stated and delta can then be chosen using the methods of section III alone.

Alternatively the aim may be to provide data to show that there is no important loss of efficacy if the test product is used instead of the reference. This is probably the most common aim of non-inferiority trials. The choice of delta for such an objective cannot be obtained by only looking at past trials of the comparator against placebo. Ideas such as choosing delta to be a percentage of the expected difference between active and placebo have been advocated, but this is not considered an acceptable justification for the choice. Such ideas were principally formulated to ensure that the reference product was superior to placebo, but this has already been addressed in section III of this document. To adequately choose delta an informed decision must be taken, supported by evidence of what is considered an unimportant difference in the particular disease area.

If there are already many treatments being used interchangeably for the disease under consideration a possible approach might be to consider the information available from all of them. From this a delta may be constructed which summarises the information known about the relative efficacy of these products, and the new trial can be designed to provide a similar level of knowledge of the relative

efficacy of the new product. This approach will not be possible if the market currently has only one product. In this situation, considering who will have to be persuaded to use the product after marketing authorisation, a possibility might be to survey practitioners on the range of differences that they consider to be unimportant, and choose delta based upon a summary statistic of the responses. Any such survey should be phrased in a way that does not bias respondents towards nominating large values.

In the situation where the test product is anticipated to have a safety advantage over the reference it is likely that a larger delta could be justified as some loss of efficacy might be accepted in exchange for the safety benefits, although it would still be expected that superior efficacy to placebo should be demonstrated. In such situations it may be useful to specify co-primary endpoints, one to demonstrate superiority in terms of the safety endpoint, the other non-inferiority on the efficacy endpoint. Other circumstances which, may warrant such consideration include a more convenient route of administration, more convenient posology, superiority on a secondary efficacy endpoint, etc.

This section has only considered some of the aims of a non-inferiority trial, and some of the possible approaches to selecting delta. The main point is that the aim of the trial should be precisely defined. Following that, a choice for delta should be made, supported by evidence, based upon the precise objectives. This evidence will not solely come from past trials of the comparator against placebo. Of course the final choice must always be at least as small as the value derived from the considerations of section III. The conclusions of the trial should not be that 'non-inferiority' has been demonstrated, but some more precise statement reflecting the objectives of the trial.

If the performance of the active comparator in the trial is very different to what was anticipated a priori there may be difficulty in interpreting the meaning of the differences between test and reference and the pre-defined delta may no longer seem appropriate. In this situation it may not be possible to draw positive conclusions from the trial. A three-arm trial including placebo provides a degree of protection against this problem as the relative effect of the active comparator compared to placebo is directly estimated in the trial.

5. AREAS WHERE IT IS DIFFICULT TO JUSTIFY A NON-INFERIORITY MARGIN OF ANY SIZE

Where the treatment under consideration is used for the prevention of death or irreversible morbidity and there is no second chance for treatment it can be very difficult to justify a non-inferiority margin of any size. Discussion of the number of extra deaths that are acceptable is ethically very difficult. However it is not in the best interest of public health to reject all choices of margin. Unless a statistically significant difference has been found between treatments, the confidence interval for the difference will not only indicate that the test product has a possible inferiority to the reference, but it will also show that it has possible superiority. Hence even if we think we are not prepared to accept any possible level of inferiority we are accepting some, by continuing to use the currently authorised product. It is important therefore that non-inferiority trials should still be possible in these areas. This section discusses some approaches to help facilitate this.

5.1 Superiority using an increased significance level

As noted in section I.4, allowing no non-inferiority margin ($\Delta=0$, essentially a superiority trial) prevents equally efficacious products from producing positive trial data, except by chance. Even products with small but clinically useful advantages would only consistently demonstrate their benefit in huge trials.

Focusing, for now, on the main efficacy endpoint only, in this particular situation there are three main reasons why a non-inferiority trial might be run rather than a trial designed to show superiority over the reference:

1. The products truly are equally efficacious, leaving a non-inferiority trial as the only option.
2. The test product has a small advantage that would require such a large trial to detect as to be impractical.
3. The product has a disadvantage, but that disadvantage is smaller than a proposed non-inferiority margin.

Obviously it is important for public health that products falling into category 2 are able to pass the tests set up more often than they fail. However it would be better, in these extreme conditions, if those in category 3 did not succeed. Any requirements set up must find this balance. The success or failure of products in category 1 is less important from a public health perspective.

This leads us to conclude that in such critical areas, a point estimate on the wrong side of zero can rarely be acceptable. With such data we are more likely to be in category 3 than category 2.

It is helpful here to parallel the setting of a non-inferiority margin with the idea of significance testing. If the 95% confidence interval were entirely above zero, we have established superiority at the 5% level of significance. For each particular value for the lower bound of our 95% confidence interval there is another confidence interval with some other coverage probability that for the same data would have a lower bound of exactly zero. For example with a data-set where the lower bound of an 85% confidence interval (by definition narrower than a 95% interval) touches zero, it might be that the 95% interval touches -5 . If delta had been defined to be -5 then achieving non-inferiority in this example would correspond to having demonstrated superiority at the 15% level of significance. Hence we can parallel running a non-inferiority trial to running a superiority trial at a less stringent significance level.

It might be an acceptable approach, in extreme situations, to run a superiority trial using a less stringent significance level than $P=0.05$, weighing up the increased risk of a false positive result against the risk of rejecting a drug with a valuable efficacy advantage. It might be more acceptable, and easier from an ethical perspective, to specify a level of confidence we require in the superiority of a drug, than to specify an extra number of deaths that is of no clinical importance.

It is recognised that a certain delta is implicitly hidden in the increased alpha level, and that this approach does not mean that all possible inferiority is ruled out. However, it is important to remember that, although 95% confidence and 5% significance have become commonly accepted, there is still a possibility of false positive results even using this significance level. Here we are merely increasing the chance of a false positive superiority claim to reduce the chance of a false negative, noting that the optimal balance between the two might be different in extreme situations.

If this approach is used the objectives and the hypothesis of the trial should be clearly stated in advance. If the hypothesis is that the test product has an advantage, this should be stated. A trial where the results support a sound pre-specified hypothesis is more persuasive than one where the results are surprising.

It would not be acceptable to switch to this approach retrospectively upon seeing study results which failed to achieve significance at conventional levels. The plan and justification for using an increased significance level would need to be clearly stated in the study protocol.

If this approach is to be used it is strongly recommended that scientific advice or protocol assistance be sought on whether the CHMP consider it to be appropriate for the particular case under study.

5.2 Products with an advantage in another aspect

In some cases if the product has an advantage in another important facet of its profile, it may be possible to define a non-inferiority margin for efficacy. If this is the case a point estimate for the difference between treatments that is on the wrong side of zero could be acceptable. In this situation it would be advisable to have two primary endpoints and plan to show non-inferiority for efficacy and superiority for the other important factor.

6. CONCLUSIONS

- The selection of the non-inferiority margin is based upon a combination of statistical reasoning and clinical judgement.

- A three-armed trial with test, reference and placebo allows some within-trial validation of the choice of non-inferiority margin and is therefore associated with fewer difficulties. This is the recommended design and should be used wherever possible.
- An appropriate choice of margin should provide assurance that the test drug has a clinically relevant superiority over placebo. This aspect of the choice of margin is discussed in section III.
- Usually the primary focus of a non-inferiority trial is the relative efficacy of the test and reference products, not simply demonstration that the test product has an effect.. In these cases an appropriate choice of margin will, in addition to proving that the product has an effect, also provide assurance that the test product is not substantially inferior to the reference, resulting in a tighter margin. This aspect of the choice of margin is discussed in section IV.
- For the majority of non-inferiority trials it must be demonstrated that the margin satisfies both the requirements of section III and section IV. The choice of non-inferiority margin should be justified in the study protocol, and the justification should address the considerations of both sections.
- It is not possible to perform a non-inferiority trial in all situations. The decision to perform a non-inferiority trial should be justified considering both the therapeutic area and the profile of the reference product.
- It may be possible to justify a wider non-inferiority margin for efficacy if the product has an advantage in some other aspect of its profile. This margin should not, however, be so wide that superiority to placebo is left in doubt.
- In some extreme situations it may be acceptable to run a superiority trial specifying a significance level greater than 0.05 as an alternative to defining a non-inferiority margin.