



The European Agency for the Evaluation of Medicinal Products
Evaluation of Medicines for Human Use

London, 26 February 2004
CPMP/EWP/2158/99 draft

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

DRAFT

**POINTS TO CONSIDER ON THE CHOICE OF NON-INFERIORITY
MARGIN**

| | |
|---|---------------------------------|
| DISCUSSION IN THE EFFICACY WORKING PARTY | December 1999 – January 2004 |
| TRANSMISSION TO CPMP | February 2004 |
| RELEASE FOR CONSULTATION | February 2004 |
| DEADLINE FOR COMMENTS | May 2004 |

Note:

Any comments to this Points to Consider should be sent to the EMEA EWP Secretariat (Fax no. +44 20 7418 8613) by end of May 2004.

POINTS TO CONSIDER ON THE CHOICE OF NON-INFERIORITY MARGIN

INTRODUCTION

Many clinical trials comparing a product with a 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 new 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 product is not worse than the comparator by more than a specified, small amount. This amount is known as the non-inferiority margin, or delta (Δ).

In order to demonstrate non-inferiority, the preferred approach is to pre-specify a margin of non-inferiority in the protocol and then to show during analysis that a two-sided 95% confidence interval (or one-sided 97.5% interval) for the true difference between the two agents lies entirely on the positive side of the non-inferiority margin. Historically, it has been common to select as delta proportion of the difference between active comparator and placebo. Such an approach does not necessarily even ensure superiority over placebo and has no clinical rationale to support it. The choice of delta must always be justified on both clinical and statistical grounds. In each case it needs to be tailored specifically to the particular clinical context and no golden 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 PtC.

- 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 trial: trials with two arms, the test treatment and a comparator; and three-armed trials with the test treatment, 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 factors 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 factors of interest 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 although in some cases the numerical analogies may not hold. 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.

I. BACKGROUND

- I.1** The outcome of a non-inferiority trial is usually a two-sided 95% confidence interval, showing a credible range for the true difference between the test agent (test: T) and the 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 were positive and if this were all the evidence available, then it is more likely that the test agent is better than the comparator 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, 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.
- I.2** 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. What is not possible if the treatments truly are equal, is to design a study to rule out any degree of inferiority at all. This would require an infinitely large experiment.
- I.3** If T were truly superior to R, then it would be theoretically possible to design a study to rule out any specific degree of inferiority at a specified significance level. In this case, it is also possible to design a study to show the superiority of T to R, but this would necessarily have to be larger than a non-inferiority trial.
- I.4** 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.

II. GENERAL CONSIDERATIONS

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;

- 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.

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.

The selection of the non-inferiority margin is based upon a combination of statistical reasoning and clinical judgement. An appropriate selection should at the minimum provide assurance that the test drug has a clinically relevant effect greater than zero. Usually the primary focus of a trial is the relative effectiveness of the test and reference products, not only the demonstration of efficacy of the test product. In these cases an appropriate selection will, in addition to proving that the product has efficacy, also provide assurance that the test product is not substantially inferior to the reference, resulting in a tighter margin.

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 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 would not be 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.)

There are many conditions where established effective agents do not consistently demonstrate superiority in placebo controlled trials (e.g. depression, allergic rhinitis). In areas where this problem of lack of assay sensitivity arises, a non-inferiority trial, which does not also include a placebo arm, is not possible. See ICH E10 for a fuller discussion of assay sensitivity.

III. DEMONSTRATING EFFICACY

When data from trials designed to show superiority of a test product over placebo are being interpreted, an informal two-stage procedure is often employed involving the consideration of both statistical significance and clinical relevance. If the test product cannot demonstrate an advantage over placebo that is both statistically significant and clinically relevant it is unlikely that a positive decision will be reached regarding the efficacy of that treatment.

Similarly, 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 following sections use the methods generally used when interpreting data from superiority trials as guidance to help establish appropriate procedures for assessing the minimal requirements of data from non-inferiority trials.

III.1 Establishing superiority over placebo: statistical significance

For any product to be licensed for a specific indication, a requirement is that it should have some efficacy for the treatment of that condition. All further consideration is irrelevant if this is not the case. Licensing decisions are made on the basis of a consideration of risks and benefits. If there is no benefit this consideration cannot result in a favourable decision. To demonstrate efficacy the product should be able to demonstrate a statistically significant advantage over placebo. Note that satisfying this condition is necessary for the approval of a product but is never sufficient. This is the ‘statistical reasoning’ stage of the ICH E10 combination of ‘both statistical reasoning and clinical judgement’.

III.1.1 Placebo controlled superiority trial

In a placebo controlled trial, it would first be expected that the test product demonstrated a statistically significant advantage over placebo. This is generally done 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 should be above zero.

III.1.2 Three arm trials: test, reference and placebo

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 initial stage of the process, the performance of the reference arm is not the main consideration, although if the test and reference products both failed to demonstrate a statistically significant advantage over placebo, this could lead to concluding that the trial was insensitive, rather than that the test product was necessarily ineffective.

III.1.3 Two arm trials: test and reference

As there is no placebo arm in this type of trial, indirect comparisons to placebo must be used, which presents inherent difficulties. However, as with trials with a placebo arm, in a non-inferiority trial 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 detailed literature search should be conducted to identify all studies relevant to the comparison of the reference treatment with placebo in the condition being considered. These can be used for estimating the true difference between the reference and placebo in the intended patient population. If such estimation is not possible or does not lead to a satisfactory result, or the comparator did not consistently demonstrate superiority over placebo, the assay sensitivity of a non-inferiority study using this comparator may be questioned and only superiority of the test product would be interpretable.

There are several issues regarding the literature search that will then 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 effect size 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 more 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 an appropriate downward adjustment of the estimate for the treatment difference should be considered.

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 the historical data.

An ‘historical’ confidence interval compares the reference product with placebo (R vs. P). The planned trial comparing the test and reference products will also produce a confidence interval (T vs. R). If these intervals were combined, an ‘indirect’ confidence interval comparing the test product and placebo could be obtained (T vs. P). Delta can then be defined as the lower bound of T vs. R that ensures that the lower bound of the indirect confidence interval of T vs. 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.

III.2 Establishing superiority over placebo: clinical relevance

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

III.2.1 Placebo controlled superiority trial

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

III.2.2 Three arm trials: test, reference and placebo

As in a superiority trial clinical judgement is applied to assess whether the observed effect size is clinically relevant. Here, 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 the setting of the trial, the reference product effect size seen in this trial can provide some context for the effect size seen on 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 effect size is clinically relevant (assuming there are no new safety problems).

Similarly, the reference product effect size can be used to calibrate our assessment of relevance for the test product. If the reference product observed effect is on the smaller end of the range expected from experience, then smaller effects for the test product may be considered clinically relevant. Without this context the same observed reference effect may have been thought irrelevant.

If the reference product has not achieved statistical significance over placebo, or has performed very differently to how experience would lead us to expect, there would be questions 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

III.2.3 Two arm trials: test and reference

Again, as with trials including a placebo arm, once statistically significant efficacy has been established using the lower bound of a confidence interval, clinical relevance can be considered using the point estimate. The point estimate from the already calculated, indirect confidence interval for T vs. P, should be noted and the clinical relevance of that difference considered.

Detailed justification should be provided of why the observed difference is considered to be clinically beneficial. This should include reference to other trials in the same therapeutic area where clinically significant results were seen, and just as importantly trials where the results were not considered clinically useful. Consideration of the discussion in section III.1.3 will be useful here.

As in the previous section the performance of 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 then 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.

III.3 Conditions where superiority over placebo has not been reliably established

In some disease areas placebo controlled trials may be considered unethical, yet the comparator has not reliably demonstrated efficacy over placebo. Examples include some oncology settings and some orphan drugs. In such situations it will be difficult to specify delta using the considerations outlined above. The best efforts should still be made to produce an indirect confidence interval for the new product against placebo using whatever data exists for the reference. If there are no data, the reference could be treated as placebo, i.e. we would essentially be looking for superiority over the reference.

It is likely that indirect confidence intervals constructed in these circumstances would fail to demonstrate indirect superiority over placebo, but in such conditions this might not necessarily preclude a positive decision. The delta derived from these indirect, historical confidence intervals should still be used. This will allow decisions to be made acknowledging the limitations of the data for demonstrating efficacy. 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 define a larger delta and then claim 'non-inferiority'. In some areas decisions are made based upon only small amounts of data without the demonstration of efficacy being clear, but it 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. A declaration of 'non-

inferiority' should not be used to mask the fact that superiority over placebo has not been convincingly demonstrated.

IV. ESTABLISHING 'NON-INFERIORITY'

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

In this respect it is important to define objectives before starting the trial. Trials are generally called 'non-inferiority' if they are not aiming to show superiority over the reference. However 'demonstrating non-inferiority' is not considered to be a suitable 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.

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, especially if a three-arm trial has been conducted. The choice of delta for such an objective cannot be obtained only by 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 this value 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 on the reference product and its competitors. 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 licensing, 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 there are differences between the anticipated safety profiles of the two products a larger delta is likely to be acceptable, as some loss of efficacy might be accepted in exchange for the safety benefits. Practitioners in this case might be asked what observed efficacy difference would persuade them to switch to the new treatment.

This section has only considered some of the aims of a non-inferiority trial, and some of the possible solutions for selecting delta. The main point is that the aim of the trial should be precisely defined. Following that, a choice of 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 as least as small as the value deriving 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 and the approach employed for defining delta.

V. EXTREME AREAS WHERE IT IS DIFFICULT TO JUSTIFY ANY NON-INFERIORITY MARGIN

V.1 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 and in some cases it may not be possible at all. Discussion of the number of extra deaths that are acceptable is morally very difficult. However it is clear that not allowing a non-inferiority margin would not be in the best interests of public health. Unless a statistically significant difference has been found between the 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 actually accepting some, because we are continuing to use the currently licensed product. It is important therefore that non-inferiority trials should still be possible in these areas. 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.

The essential requirement in this situation is to provide enough information so the users of the product are confident about using it as an alternative to products already available. When making this decision a user would look at more than just the lower bound of the confidence interval, so it is important not to be over-restrictive when interpreting the data.

In determining the level of evidence necessary for such trials many things need to be considered, but it is important to keep the question that is being answered focussed. 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 smaller than an acceptable 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 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 confidence interval for the difference between the treatments, where the point estimate falls 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.

V.2 It is interesting 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 would have established superiority at the 5% level of significance. For each particular choice of delta 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 border on zero. For example if delta were specified as -5 , it might be that an 85% confidence interval (by definition narrower than a 95% interval), would touch zero when the 95% interval touched -5 . Hence 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 acceptable to specify a less stringent significance level, weighing up the increased risk of a false positive result against the risk of rejecting a drug with a small but useful advantage. It might be more acceptable, and easier from a moral 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 useful 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 to reduce the chance of a false negative, noting that the optimal balance between the two might be reconsidered in the extreme situation.

V.3 The objectives and the hypothesis of the trial should be clearly stated in advance. If the hypothesis is that the test product has a certain 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. If a reduced level of significance is being used this should also be stated. The justification for and implication of the choices should be discussed. The choice of significance level might be justified by surveying practitioners, as discussed in section IV.

V.4 It will not be possible to power these trials if it is assumed there is no difference in efficacy between the treatments. If an advantage is expected for the new product this should be used in the power calculations.

V.5 In some cases a safety advantage might be great enough to make some loss of efficacy acceptable. This should also be considered when acceptable outcomes are being considered. If this is the case a point estimate for the difference between treatments that is on the wrong side of zero might be acceptable. In this situation it would be advisable to have two primary endpoints, one representing efficacy, the other safety.

VI. PROPORTIONS

In clinical trials where the primary endpoint is summarised as a proportion the non-inferiority margin could be expressed as either an absolute or a relative difference. The company should justify whichever is used, based on clinical considerations.

The choice of non-inferiority margin that is acceptable when using an absolute difference varies with the response rates observed. For example a trial may select a value of delta assuming the reference product will have a response rate of around 50%. However if it transpired that the observed response rate was nearer to the extremes, the chosen delta might then seem too large.

The use of a relative, rather than absolute, difference is associated with a similar, but opposite, problem. A chosen delta on the odds ratio scale may seem appropriate when the reference treatment is expected to have response rates near 0% or 100%, but might seem too large if the observed rate is less extreme.

It is therefore recommended that in trials using proportions consideration should be given to the measurement scale. A combined non-inferiority criterion could be specified where the margin is pre-specified as both a relative and an absolute difference using the assumed reference rate, and that which emerges as the most conservative from the final data is used for the analysis.

Of course, if the observed data on the reference lead to doubts about the performance of the reference arm in the trial, then the whole validity of the trial may be questioned.

The ideas in this section could also be applied to situations not involving proportions where there is a choice of scale.

VII. CONCLUSIONS

- The minimal requirement is that the new treatment would have been successful if a placebo-controlled trial had been run. This means that the treatment should demonstrate, either directly or indirectly, both a statistically and clinically significant advantage over placebo.
- Often, for the determination of an acceptable non-inferiority margin a detailed literature search will need to be conducted to identify all studies that bear relevant information on the comparison of the reference with placebo in the considered indication. There are critical issues with historical data that need to be considered: Are the data still relevant? Is there publication bias, and if there is, can it be removed? Is the same population studied and have effects remained constant over time?
- When the results of a non-inferiority trial deviate from the pre-planned assumptions, the specified non-inferiority margin may become inappropriate. If the performance of the reference treatment deviates from the expected results, then the validity of the trial may be called into question.
- In many situations more than superiority over placebo is required, but it is important that the requirements are not so stringent that treatments with small but clinically useful advantages cannot demonstrate their value.
- When the treatment is used for prevention of death or in other cases of severe morbidity and there is no second chance for treatment it can be very difficult to justify a non-inferiority margin of any size and in some cases it may not be possible at all. In such critical areas, the point estimate for the difference between test product and active comparator should fall on the positive side of zero and the associated *P*-value should be small enough for the difference to be convincing.
- The objectives of a trial should be stated very clearly and precisely. Rather than just stating that the aim is to demonstrate ‘non-inferiority’, a more precise statement should be made. Following that a choice of delta should be made, backed up by evidence, based upon the precise objectives. The conclusions of the trial should then reflect these precisely defined objectives and the method used to select delta.
- The choice of delta must always be justified on both clinical and statistical grounds. In each case it needs to be tailored specifically to the particular clinical context and no golden rule can be provided that covers all clinical situations. At the time of assessment of an MAA decisions will be based upon considerations of risk-benefit.

