INTERNATIONAL SOCIETY FOR CLINICAL BIOSTATISTICS

Chairman. Sub-Committee on Statistics in Regulatory Affairs:
Jorgen Seldrup, PhD, Director, Quintiles Biostatistics Europe
4 route de la Riviere, Parc Club des Tanneries - Lingolsheim, BP306
67832 Tanneries Cedex, France

Permanent Office, P.O. Box 130, DK-3460 Birkerod, Denmark
Phone: +45 4567 2279 Fax: +45 7022 1571 E-mail: Office@ISCB.info
Home page http://www.ISCB.info

Comments regarding CPMP/EWP/2158/99 draft: "Points to consider on the choice of non-inferiority margin", London, 26 February 2004.

General

We welcome the discussion in the introductory, background and general considerations sections. However, inevitably throughout there is as much discussion of the non-inferiority trial *design* as there is of the choice of the margin that maybe including the word 'design' in the title would better reflect the balance of discussion.

It is felt that the emphasis on the point estimate in respect of assessing the clinical relevance (e.g. in III.2.1 "Establishing a clinically relevant benefit over placebo is generally accomplished by considering the point estimate...") is at best misleading. It may even be incorrect to suggest that this is what the clinician uses to assess clinical significance. The point estimate of an effect is an 'uncertain' estimate of the true effect and the clinician in assessing relevance of an efficacy result would do better in reflecting on the smallest effect as indicated by the confidence interval.

Although the PtC is a useful guidance in general terms, disease-specific guidelines discussing the practical choice of delta would be of even more use. Failing that for the moment some real examples inserted in the present PtC would be helpful.

It might also be helpful to have some comments specifically on composite and surrogate endpoint given that they might generate additional problems. Consideration should also be given to other scales than "original scale ... or responder scale" (e.g. 'standardised difference', i.e. Cohen effect size). The binary case is clearly difficult but one possibility here might be to select a margin on the log-odds scale that also guaranties a maximum possible difference on the probability scale.

The PtC uses a mixture of terms - product, agent, treatment, arm - in the introduction and elsewhere. We suggest that a consistent term be used - for instance, treatment and treatment group.

The document would benefit by adding an index and a list of definitions (e.g. 'effect size', used in many places in the document, may mean different things to different people; in many sample size contexts it is defined as a 'standardised' term – mean over standard deviation). More generally we feel that it would benefit from being made shorter and more precise.

Specific Points

1. INTRODUCTION – 6th paragraph
   The expressions "factors of interest", "certain variables" and "single factors" could have been better chosen as they may be confused with the statistical term 'factor'.

2. I BACKGROUND
   I.1 The use of the word "outcome" to describe the use of a confidence interval (CI) seems inappropriate. The CI is used to describe the outcome but is not in itself the outcome.
   I.2-I.3 These two sections describe the situations of T = R and T > R, respectively. More discussion of T < R than is found in I.4 may be interesting.

3. III DEMONSTRATING EFFICACY
   This is the central part of the PtC and it needs much more clarity in order to provide useful guidance. The discussion of the indirect CI in III.1.3 is particularly unclear. The prominence of 'historical' data in the discussion of need for statistical justification is almost contradictory when at the same time it is acknowledged that the literature is biased towards 'positive' publication. It is well understood that agreeing on a non-inferiority margin will be extremely difficult in many situations. However, un-thoughtful use of 'statistical data' should not be the answer. One should also be careful to not unreservedly accept 'historical' data as clinically relevant just because statistical significance has been shown.

   III.1.2 Some discussion of the case where the three-armed placebo and reference controlled study results in T > P but R = P would be interesting.

   III.2.2 Last sentence in the first paragraph ends: "... 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)". The effect size is still relevant; it is the risk/benefit which may be unacceptable.

4. IV ESTABLISHING 'NON-INFERIORITY'
   In the second paragraph it is stated that "...'demonstrating non-inferiority' is not considered to be a suitable objective for a trial". In the same vein it would not be

'a suitable objective' to demonstrate superiority. We believe it is a question not of a 'suitable objective' but of an insufficiently detailed description of the objective.

In the fourth paragraph it is said that choosing delta as a percentage of the expected difference "is not considered an acceptable ... choice". This may be too strong. In some situations it may be the 'right' (or only) choice meaningful to statistician and clinician.

In the fifth paragraph (last sentence) it is stated that "any such survey should be phrased in a way that does not bias respondents towards nominating large values"; 'inappropriate values' may be a better term – large values (given a reference) may be acceptable in a given situation.

In the sixth paragraph differences between safety profiles are singled out as acceptable reasons for a larger delta. There may also be other circumstances which may warrant such consideration, e.g. improved regimen, more convenient route of administration, aspects which would improve compliance.

The seventh paragraph finishes by stating that "the final choice must always be at ("as" is a typo) least as small as the value from the considerations of section III". Considering the unclear statements in that section, this statement is not very helpful.

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

Prevention of death and irreversible morbidity are singled out, as situations where any size non-inferiority margin is ethically difficult to justify. It should not be overlooked that superiority trials in the situation discussed (test v. Reference) may be equally 'un-ethical'. Such trials may require the 'placebo add on' type of trial.

V.1 The two paragraphs following the three reasons for why a non-inferiority trial might be run should be dropped. The reasoning is unconvincing, e.g. products of category 2 (small advantage) might always be in the public interest, especially in extreme situations where it is difficult to justify any non-inferiority margin. In addition, the statements in these paragraphs are contradictory to those in V.5.

V.2 2nd paragraph. The extra number of deaths that are accepted by the suggested approach (regarding the conventional lower 95% confidence limit) is simply masked by the decreased confidence level. This means that a certain delta is implicitly hidden in the (increased) alpha level, and – more important – this delta will implicitly depend on the selected sample size. Therefore, the interpretation becomes even more difficult, and it lacks transparency. We therefore suggest that it should be discarded.

6. VI PROPORTIONS

It is not clear why the acceptability of the margin depends upon the observed responses (second paragraph). We suspect this is an untested supposition and therefore should not appear in a guideline.

The PtC should explain why the relative difference is associated with a similar but opposite problem to that of the absolute difference (3rd paragraph). If one method maintains the difference in proportions (absolute difference) whilst the other effectively adjusts the difference in proportions, such that it is narrower when approaching 0 or 1, then how can both methods have problems. If the fixed difference is indeed a problem then surely the relative difference (Odds Ratio) must be the solution.

The difference in proportions is not amenable to covariate adjustment - although some limited methods do exist. However, covariate adjustment is required for stratified designs (as per the CPMP PtC on Covariate adjustment and ICH E9). It should be stated that the margin and model formulation should be consistent and that careful thought should be given to stratified designs. It should be highlighted that the odds ratio is more amenable to covariate adjustment.

In the 4th paragraph a combined non-inferiority criterion is discussed. It is unclear how the suggested criterion of specifying the margin on two scales and choosing the most conservative would perform in practice. In any case, it may be impractical and should not be necessary in confirmatory trials.