Main Paper

Pharmaceutical Statistics

(wileyonlinelibrary.com) DOI: 10.1002/pst.508

Published online in Wiley Online Library

The draft FDA guideline on non-inferiority clinical trials: a critical review from European pharmaceutical industry statisticians

Bernhard Huitfeldt,[a]* and Jürgen Hummel,[b] on behalf of European Federation of Statisticians in the Pharmaceutical Industry (EFSPI)

The European Federation of Statisticians in the Pharmaceutical Industry (EFSPI) engages more than 2000 statisticians through its ten national organizations. Amongst other things, EFSPI is involved in reviewing regulatory guidelines under development, including the draft FDA guideline on non-inferiority clinical trials. This review resulted in several critical comments relating to as follows: (i) the lack of one single standard for proving efficacy of new drugs implied by the guideline; (ii) the problems with the suggested 'fraction of effect to be preserved'; (iii) the formulation of the primary hypothesis in a non-inferiority trial aiming at indirectly demonstrating a new drug is superior to placebo; and (iv) the preference in the guideline for the fixed-margin method over the synthesis method in the analysis. The presumed implications of this guideline, if implemented as is, are (i) increased confusion of how efficacy could be demonstrated when placebo control is not available, (ii) more complicated communication between pharmaceutical industry and FDA because of the apparent disagreements on fundamental statistical matters, and (iii) illogical consequences in the approval process because of which order drugs are approved rather than how they fulfill the regulatory requirements. We believe that the area is not yet ready for such a prescriptive regulatory guidance and that further research and experience are required until the methodology can be finally agreed. A strategy needs to be developed by regulatory agencies together with drug industry and academia for a long term solution for this topic. **Copyright © 2011 John Wiley & Sons, Ltd.**

Keywords: FDA guideline; non-inferiority; synthesis method; fixed-margin method; EFSPI

1. INTRODUCTION

The environment for undertaking clinical trials in drug development is gradually changing. One important element of this change is the perception of the role of placebo, which is part of the current clinical trial paradigm for proving efficacy of a new drug treatment. In light of the continuing increase of new effective treatments, the use of placebo in clinical trials is being challenged by the medical community including ethics committees and patient advocacy groups. Regulatory authorities, on the other hand, require placebo-controlled trials for a new drug to be approved unless it is impossible for ethical or other reasons to conduct such trials. It is, therefore, of paramount importance to find a balance in this dilemma, which offers a reasonable environment for drug industries to develop new drugs combined with regulatory requirements that allow authorities to perform their legitimate task to protect the public from inefficient and unsafe drugs. It is against this background that the new US Food and Drug Administration (FDA) draft guideline for non-inferiority (NI) clinical trials should be seen. The final version of this guideline will set the conditions for drug development for a long time, which is the reason why it has attracted so much interest in the clinical trial community.

This paper gives a critical review of the draft guideline based on consolidated comments from the national organizations of the European Federation of Statisticians in the Pharmaceutical Industry (EFSPI).

2. CURRENT REGULATORY DOCUMENTS FOR NON-INFERIORITY STUDIES

Currently, there are two regulatory guidelines that directly address the issue of demonstrating efficacy by using an active control design. The first one was developed by the International Conference on Harmonization (ICH) entitled 'Choice of Control Group and Related Issues in Clinical Trials' (ICH-E10) and was published in 2000 [1]. The second one was developed by the European Medicines Agency (EMA) and was called 'Guideline on the choice of the non-inferiority margin' published in the series of 'points to consider' documents. This guideline came into effect in 2006 [2]. More recently a reflection paper has been issued from EMA addressing a related question, namely the need for active

[a] *BH Statistical Consulting, Stenåsvägen 2, 155 33 Nykvarn, Sweden*

[b] *PPD Inc, Fleming House, Phoenix Crescent, Strathclyde Business Park, Bellshill, North Lanarkshire ML4 3NJ, UK*

**Correspondence to: Bernhard Huitfeldt,, BH Statistical Consulting, Nykvarn, Sweden.*
E-mail: bernhard.huitfeldt@ebox.tninet.se

This article is published in Pharmaceutical Statistics as a special issue on ICH.E10: The First Ten Years, edited by Steven A. Julious, Medical Statistics Group, Health Services Research, University of Sheffield, Regent Court, 30 Regent Street, Sheffield, S1 4DA, England.

Copyright © 2011 John Wiley & Sons, Ltd.

control, where use of a placebo control is deemed ethical [3]. In the following sections these guidance documents are briefly summarized.

2.1. ICH-E10 Choice of Control Group in Clinical Trial and Related Issues

This guideline gives a broad description of design issues in controlled clinical trials. In particular, different purposes of clinical trials are discussed distinguishing between those that provide evidence of efficacy by using a direct or indirect comparison with placebo, and those that provide relative efficacy and safety information about a test compared with an active control. It is mentioned that an indirect comparison with placebo can be achieved by showing superiority to an active control or by a non-inferiority study, where a test treatment is shown to be non-inferior to an active control previously shown to be superior to placebo. However, no specific guidance is given on the choice of non-inferiority margin. The draft FDA guideline on non-inferiority clinical trials attempts to fill that void.

In the ICH-E10, the concept of assay sensitivity was introduced to indicate whether a clinical trial was able to distinguish an effective treatment from a less effective or ineffective treatment. Constancy, the other key assumption about non-inferiority trials, which relates to the effect of the active control being constant over time, was also discussed even though the very term constancy was never used.

Judging by the way the guideline has been written, it seems clear that harmonization across regions (USA, Europe, and Japan) had been very difficult to achieve, and this is illustrated in the sentence 'the guideline does not address the regulatory requirements in any region, but describes what trials using each design can demonstrate'. In spite that, ICH-E10 was an important step in the development of this area and has inspired many statisticians to conduct further research for improving the methodology.

2.2. European Medicines Agency guideline on choice of non-inferiority margin

The EMA guideline is more focussed on the actual choice of the non-inferiority margin compared with the ICH-E10, which has a broader scope. It defines this margin as a pre-specified small amount (delta), which is used to demonstrate that the test product is not worse than the comparator by more than this amount. The EMA guideline is concerned with both the absolute efficacy of the test treatment in relation to placebo and the relative efficacy of the test treatment and the active control, whereas in the ICH-E10 the relative efficacy is emphasized to a lesser degree. But similarly to ICH-E10, no specific advice is given for the choice of margin. Instead, the guideline provides general advice on how studies should be designed to be suitable for a non-inferiority study and considerations for using historical data as a basis for the choice of margin. The general spirit of the EMA and ICH guidelines is very similar with a few exceptions, which will be detailed in Section 5 of this paper.

2.3. European Medicines Agency reflection paper on three-armed trials

In this reflection paper, the importance of including a placebo control in pivotal trials used to support marketing authorization applications is emphasized, whenever this is ethical and feasible. But it is also stated that three-armed trials with experimental medicine, placebo and active control represent a scientific gold standard and that there are multiple reasons to support their use in drug development. The need for active control is not a formal requirement but is strongly recommended for a proper evaluation of the benefit/risk of a new experimental medicine. The following two situations are mentioned when this is particularly relevant:

1. The experimental medicine might be associated with serious safety concerns or
2. The treatment with a medicine with inferior efficacy might cause serious harm for the patients.

It is emphasized that the comparison with an active control should normally be direct, that is, within one and the same trial but that there are circumstances where an indirect comparison might be sufficiently reliable. The issue about direct or indirect comparisons is closely related to the issue of an indirect comparison of an experimental medicine to placebo in a non-inferiority trial. However, in this reflection paper, there is no discussion about how such an indirect comparison should be conducted, for example, in terms of determination of a non-inferiority margin.

3. DRAFT FOOD AND DRUG ADMINISTRATION GUIDANCE ON NON-INFERIORITY CLINICAL TRIALS: MAIN CONCEPTS

The FDA has publicly expressed its expectations regarding the choice of non-inferiority margins for some time, for example, in a presentation from February 2002 [4]. Thus, the pharmaceutical industry was eagerly awaiting the draft guidance on non-inferiority clinical trial [5], which was released for comments on 2 March 2010 [6]. The draft guidance document was highly detailed, filling 2433 lines over 63 pages.

The guidance can be grouped into four main parts. The first part covers a general discussion of regulatory, study design, scientific and statistical issues, whilst the second part provides details on statistical approaches to determine the non-inferiority margin. This is supplemented in the third part by commonly asked questions about NI studies. In the fourth part, five case studies are presented to illustrate different aspects of the process of choosing a NI margin, of the application of a method of NI analysis, and other considerations relevant to conduct and interpretation of results from a NI study. The FDA guidance regards the determination of indirect efficacy over placebo as the main objective for a non-inferiority study (see Section 4 for more details).

In this section, we introduce the main concepts used in the draft guidance, particularly those to which our comments in Section 4 relate.

3.1. Non-inferiority margin: role of M_1 and M_2

The definition of the non-inferiority margin is based on two values (M_1 and M_2), which need to be chosen before the non-inferiority study is performed.

The draft guidance defines M_1 as the 'entire effect of the active control assumed to be present in the NI study' and specifies its derivation as the 'upper bound of the 95% two-sided confidence interval (CI) for control–treatment (C-T)', acknowledging that this

Copyright © 2011 John Wiley & Sons, Ltd.

leads to a 'conservative' value. The choice of M_1 is to be based on the following:

1. Treatment effect estimated using historical evidence of sensitivity to drug effects (HESDE) with active control drug, using a meta-analysis
2. Assessment of likelihood that the current effect of the active control is similar to the past effect (constancy assumption). Using HESDE as the basis of M_1 for the NI study is only appropriate if the constancy assumption holds.
3. Assessment of the quality of the NI trial to be performed

M_2, the non-inferiority margin, is defined as 'the largest clinically acceptable difference (degree of inferiority) of the test drug compared with the active control'. The derivation is to be based on the largest loss of effect that would be clinically acceptable, therefore, quantifying how much of the effect of the active control needs to be preserved. Its choice is a 'matter of clinical judgement', but 'M_2 can never be greater than M_1'.

Figure 1, reproduced from Figure 3 in the draft FDA guidance, illustrates four different examples and their proposed interpretation:

1. 'Test drug is effective (NI-demonstrated)'.
2. 'Upper bound of 95% CI $>$ M_2, indicating unacceptable loss of the control effect'.
3. 'Upper bound of 95% CI $<$ M_1, but it is slightly $>$ M_2. Judgment could lead to conclusion of effectiveness'.
4. 'No evidence of effectiveness for test drug'.

3.2. Analysis methodology: fixed margin and synthesis method

The draft guidance specifies two different analysis approaches for a non-inferiority study: the fixed margin method and the synthesis method.

In the fixed margin method, M_1 is based on the effect of the active comparator in previous studies. The NI margin is prespecified, usually chosen to be smaller than M_1 (i.e., M_2), and the NI study is successful if it rules out inferiority of test drug to the control drug by the NI margin. The name 'fixed margin method' comes from the use of past studies to derive a single fixed value for M_1.

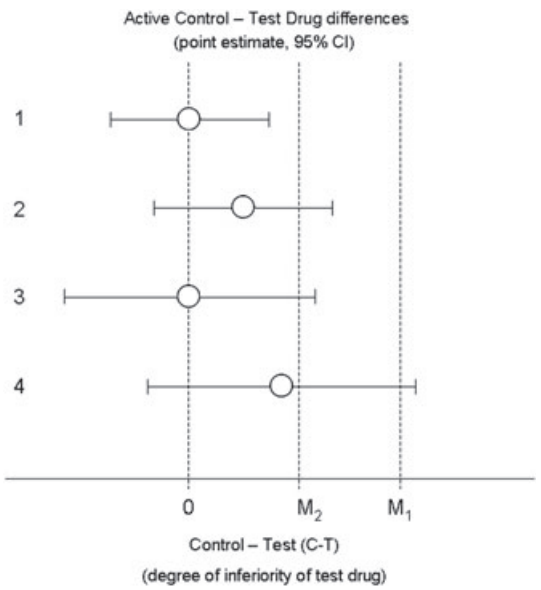

Figure 1. Active Control–Test Drug differences.

The synthesis method, in contrast, combines (synthesizes) within one analysis an estimate of the treatment drug effect relative to the control drug (from the NI trial) and an estimate of control drug relative to placebo (from a meta-analysis of historical trials). It does so by treating both data sources as if they came from the same trial, thus estimating what the placebo effect would have been had a placebo treatment been included in the NI study. One CI is produced for the superiority hypothesis without specifying a fixed NI margin.

We also want to highlight that the FDA guidance explicitly expresses its preference, stating that 'the fixed margin approach is preferable'.

4. EUROPEAN FEDERATION OF STATISTICIANS IN THE PHARMACEUTICAL INDUSTRY COMMENTS ON DRAFT FOOD AND DRUG ADMINISTRATION GUIDANCE

4.1. Collating of comments

European Federation of Statisticians in the Pharmaceutical Industry (EFSPI) is a federation of national associations of biostatistics or medical statistics from ten European countries. For more information about EFSPI, see http://www.efspi.org/. It was formed in 1990 and is engaged in statistical aspects of research, development, production, and surveillance of drugs and medical devices. Its constitutional objectives include promoting professional standards of statistics and the standing of the statistical profession of importance for the European pharmaceutical industry. Through its member associations, EFSPI engages over 2000 statisticians.

An important task of the federation is to be actively involved in the development and review of regulatory guidelines, such as those from ICH, FDA, and EMA. In particular, EFSPI was involved in the review of ICH-E9 (Statistical Principles for Clinical Trials) and ICH-E10 (The Choice of Control Group in Clinical Trials), and more recently also in the EMA guidelines about missing data and investigation of bioequivalence. Regarding ICH-E10, EFSPI wrote a critical position paper describing shortcomings of this guideline in providing harmonized guidance in a disputed area [7]. This engagement started EFSPI's interest in the issue of proving efficacy in active control studies, in the face of increasing difficulty in using placebo in such trials. The FDA guidance on non-inferiority clinical trials can be seen as a continuation of the ICH-E10 guideline, with the inclusion of the development of methods and regulatory practices that have taken place since then. It was therefore entirely natural for EFSPI to engage in the review of this new FDA guideline.

EFSPI has a well-established process for collating comments on draft regulatory guidelines. The EFSPI council appoints a rapporteur, who coordinates a group of individuals representing the member organizations. These individuals collate review comments from their respective member organization, which are subsequently consolidated to become the official EFSPI comments after thorough discussion and agreement in the coordinating group. The consolidated comments are finally approved by the EFSPI council before they are submitted.

This process was applied for the review of the current FDA guideline on non-inferiority clinical trials. The complete document including all comments from EFSPI, major and minor, can be found on the EFSPI website [8].

Copyright © 2011 John Wiley & Sons, Ltd.

The document is organized into a number of general comments (referring to concepts or issues appearing in several places of the guideline) and a series of specific comments (regarding particular sections, generally with a specific proposed change). In the following, we present the major features of the EFSPI comments as they were submitted to the FDA.

4.2. Single standard for proving efficacy

The true objective of an active-controlled efficacy trial is to show that the drug is efficacious, that is, would have been superior to placebo if a placebo controlled trial could have been conducted. The use of an active controlled trial in combination with historical data, to indirectly demonstrate superiority to placebo, should not be used as a basis to require an arbitrarily higher standard for proving efficacy (in this case by introducing an additional fixed margin M_2, which addresses the relative efficacy of the test drug in relation to the control). This important principle has not been applied in the draft guidance. Thus, we are concerned about the recommendation in the guidance that suggests different standards for proving efficacy depending on whether placebo or an active comparator is used. Arguments for a single standard of evidence for deciding whether a pharmaceutical treatment has demonstrated sufficient efficacy have also been presented in [9].

We acknowledge that there are weaknesses caused by an indirect comparison, that is, the assumptions of assay sensitivity and constancy. Thus, imposing some degree of conservativeness may well be motivated, for example, through some kind of discounting, for example in the determination of M_1. However, the primary purpose should still be to establish efficacy over placebo. The examination of relative efficacy for the new drug versus the control should not be an integrated part of analyzing the primary objective, which is to prove superiority of the test drug over placebo.

We suggest therefore that it should suffice to require meeting the margin M_1 for demonstrating efficacy of a test drug should a fixed-margin approach be used. This would be in line with the usual requirements in placebo-controlled trials and it would eliminate the uncomfortable need for the subjective and probably in many cases not well-understood decision on a fraction of effect of an active control to be preserved (M_2).

4.3. Fraction of effect to be preserved

The guideline introduces a second fixed margin requirement (M_2), which is defined as a fraction of the active control effect to be preserved. Thus, when comparing a new test drug with an active control, the requirement implies that this more stringent margin (M_2) is also excluded. This has two problems.

First, the requirement seems to be based on the trial design (i.e., a non-inferiority design) rather than, more appropriately, on the existence of an effective therapy for the condition being studied. The document encourages the use of placebo-controlled trials to demonstrate a treatment effect when ethically feasible. Given this, it does not seem logical to require a certain fraction of effect to be preserved only when a non-inferiority design is chosen, but not when a placebo-controlled design is chosen. Our opinion is that the existence of an effective therapy, not the trial design, should determine whether preservation of effect is required.

The second problem regarding the implementation of this requirement is that it bases the conclusions regarding a clinically meaningful effect on the lower end of the confidence bound for effect to be preserved, rather than on the point estimate of the difference between the test drug and the active control. This is inconsistent with the customary approach for judging whether a treatment effect is clinically meaningful, and also may lead to serious logical inconsistencies in approval decisions. It may even prevent truly superior new drugs (or truly effective drugs with improved safety) to be approved. This has also been convincingly demonstrated in [9].

4.4. Statement of primary hypothesis

One fundamental issue with the draft guidance concerns the statement of the primary hypothesis to be tested and the associated type-1 error rate to be controlled. According to the guidance, the margin M_1 should be chosen so that, when a difference between the treatment and control of M_1 or greater can be ruled out, one can conclude that the treatment difference is greater than zero (i.e., treatment is superior to a placebo). In other words, M_1 is simply a nuisance parameter in this situation. However, the document states that the primary hypothesis to be tested, and the associated type-1 error rate to be controlled, should be with respect to the difference between treatment and control of M_1 or greater. Because the primary aim is to demonstrate that the treatment is superior to placebo, the most appropriate null hypothesis should be that the treatment has no effect, and the type-1 error rate of interest should be the probability of declaring an ineffective treatment to be effective. This probability can be controlled through appropriate choice of M_1, or through other statistical approaches such as the synthesis method. Stating the hypothesis in terms of M_1, which is simply a nuisance parameter, adds confusion throughout the document.

4.5. Synthesis method versus fixed margin method

The document describes two methods of analysis, the fixed-margin method and the synthesis method, in an inconsistent way. In some places, the document correctly refers to the synthesis method as an alternative to the fixed-margin approach that differs only in the way the variance terms (from the historical data and the non-inferiority trial) are pooled. The way the variance terms are pooled in the synthesis method implies that this method is more efficient than the fixed-margin method. Sometimes, the document refers to the inefficiency of the fixed-margin method as a form of discounting that provides some additional assurance in the presence of concern over the constancy assumption. However, in other places it incorrectly refers to fundamental problems with the synthesis method that are not shared by the fixed-margin method. In fact, there are no such problems; in some cases, this would become much clearer if the issue with regard to the statement of the null hypothesis were fixed, as described in Section 4.4.

There are incorrect assertions that the fixed-margin method is not affected by the constancy assumption. There are also incorrect statements regarding advantages with respect to sample size calculation and the need for clinical judgment regarding efficacy preserved. Finally, the document rejects the use of the synthesis method for M_1, but potentially allows its use when evaluating M_2. We propose that the most efficient method should be used for both analyses, that is, the synthesis method.

The preferred analysis method in the guideline for the fixed margin approach is the so-called '95–95' method [10], justifying it because of its conservativeness when demonstrating efficacy

Copyright © 2011 John Wiley & Sons, Ltd.

of a new drug over placebo. However, this method suffers from several drawbacks. It is, by definition, a fixed margin approach, which does not treat the historical estimate of efficacy of the active control over placebo as a random variable, which indeed it is. Further, because the method is based on the most unfavorable end of a 95% confidence interval, it will lead to an overly conservative outcome. The synthesis approach, accounting for the variability in both the current and the historical studies, should be recommended instead.

4.6. Other comments

The guidance document is quite extensive, and the general impression is that it is overly wordy and unnecessarily complicated. The suggestion is to make it more condensed by primarily removing repetitions. One suggestion could be to define all key concepts and terms in one place and then use cross-referencing when these terms appear later in the document. For example, the term assay sensitivity is defined in four places with slightly different wordings, which is rather confusing:

1. 'could have distinguished an effective from an ineffective drug'
2. 'control drug had at least the effect it was expected to have'
3. 'ability of the trial to have detected a difference between treatments of a specified size, M_1 (the entire assumed treatment effect of the active control in the NI trial), if such a difference were present'
4. 'active control would have had an effect of at least M_1'

The word 'effectiveness' is used inappropriately throughout the document instead of efficacy, which is assessed in a study using a non-inferiority design. Effectiveness is usually defined as the extent to which an intervention does more good than harm when provided under usual circumstances of health care practice whereas efficacy relates to the extent to which an intervention has a positive effect on the disease under clinical trial conditions [11].

5. EUROPEAN MEDICINES AGENCY AND DRAFT FOOD AND DRUG ADMINISTRATION GUIDELINES: SIMILARITIES AND DIFFERENCES

Many pharmaceutical companies aim to perform their drug development globally. As this is greatly facilitated by consistent regulatory requirements, we compare the EMA guideline with the draft FDA guideline. Both documents are conceptually similar, but use different terminology: 'Demonstrating efficacy' (EMA) corresponds to meeting margin M_1 (FDA) and 'Establishing acceptable relative efficacy to active comparator' (EMA) corresponds to meeting margin M_2 (FDA).

There are some important differences between the two guidelines, though. The EMA guideline asks the sponsor to specify the goal of the study, for which it offers two options (indirect superiority over placebo or no important loss of efficacy). The FDA guidance only accepts the goal of no important loss of efficacy. As such, the FDA guidance is more stringent; if a trial satisfies the FDA guidance, then it also satisfies the EMA guidance.

Another difference relates to the derivation of M_2. The FDA guidance suggests that M_2 is defined as a fraction of the active control effect to be preserved. The EMA guideline states that a non-inferiority margin as a proportion of the active versus placebo difference is deemed inappropriate, both for studies where the purpose is to indirectly prove a new drug is superior to placebo and for studies where relative efficacy is the primary purpose.

6. IMPLICATIONS OF DRAFT FOOD AND DRUG ADMINISTRATION GUIDANCE

In the same way, as ICH-E10 addressed a scientifically immature area at the time it was written, the FDA draft guideline on non-inferiority clinical trials also tries to regulate an area for which neither theory nor practice has been stabilized. If the final version of the guideline will look more or less the same as the draft version, then there are reasons to be very concerned about the implications for drug development in the future. We have a general feeling that the ambition to ensure a conservative assessment of new drugs, that is, focusing on protecting the type-1 error, leads to several elements in the draft guidance that aim to achieve conservativeness in a way that is very confusing and unpredictable for the pharmaceutical industry. The implication is that many truly effective treatments will never reach the market because too little consideration is given to the type-2 error in this context.

There is also genuine disagreement with regard to a number of fundamental statistical issues between statisticians working in the pharmaceutical industry and the authors of the draft guideline. This is not because of the fact that the proposed conservative methodology generally results in greater difficulties to get new drugs approved, but, more importantly, that this methodology is not founded on sound statistical theory, which have been described both in this paper and in the Pharmaceutical Innovation Steering Committee (PISC) Expert Team paper [9]. The communication between FDA and industry will definitely be more complicated if there are disagreements that are not based on our different roles in drug development but rather on different opinions on the subject matter. A similar disagreement existed when FDA launched the idea of individual bioequivalence [12], which was never enforced because of strong opposition from the statistical community in drug industry.

Another implication of the FDA draft guideline concerns the logical inconsistency associated with the preservation of effect criterion, that is, the use of M_2 in a fixed-margin approach. As has been illustrated in [9], depending on which drug is approved first a new experimental medicine may not be approvable even though it is superior to placebo and with numerically better efficacy compared with the drug already approved.

7. CONCLUSIONS AND SUGGESTED ACTIONS

We welcome the draft guidance on non-inferiority clinical trials, as it documents current FDA's expectations for such trials. However, we are sincerely concerned about some of the requirements stated within the guideline. In our opinion, they do not represent a fair balance between the need for the regulatory agency to protect the public from ineffective new drugs and the desire for the industry to offer the public new effective drugs. This is reflected in many measures suggested to protect the type-1 error without a similar appreciation of the impact on the type-2 error. Some of these requirements are also not founded on sound statistical theory, for example, formulating the primary hypothesis as a non-inferiority hypothesis when the primary scientific question is concerned with superiority of the test drug over placebo. Another example is the preference for the fixed-margin method over the

Copyright © 2011 John Wiley & Sons, Ltd.

more efficient synthesis method. The use of the synthesis method would eliminate the need for determining a non-inferiority margin, which is partly based on a subjective decision.

We do not think that this area is quite ready for a regulatory guideline with such prescriptive requirements of the nature seen in the draft FDA guideline. We would suggest that this document instead is regarded as a position statement providing the current FDA opinion, with an acknowledgement that further research and experience are required until the methodology can be finally agreed. Furthermore, we would welcome an invitation by both the US and EU regulatory agencies to all interested parties, including industry and academia, to develop a strategy for the development of a long term solution for this topic. This should also include a broader discussion on how efficacy of new drugs should be demonstrated in the future when fewer and fewer relevant trials will be available that have adequately demonstrated the efficacy of standard treatments against placebo.

Acknowledgements

The authors wish to thank and acknowledge the role of all members of the EFSPI review committee for the FDA draft guideline on non-inferiority trials, which included Egbert Biesheuvel (MSD, Holland), Chrissie Fletcher (Amgen, UK), Lars Frison (AstraZeneca, Sweden), Toni Sarapohja (OrionPharma, Finland) and Kristian Windfeld (Lundbeck, Denmark).

REFERENCES

[1] International Conference on Harmonisation (ICH). *Note for Guidance on Choice for Control Group in Clinical Trials*. EMA: London, 2001.

[2] European Medicines Agency (EMA). *Guideline on the choice of non-inferiority margin*. EMA: London, 2006.

[3] European Medicines Agency (EMA). *Reflection paper on the need for active control in therapeutic areas where use of placebo is deemed ethical and one or more established medicines are available*. EMA: London, 2010.

[4] Temple RJ. Active Control Non-Inferiority Studies: Theory, Assay Sensitivity, Choice of Margin. *Presentation given on 19 February 2002*. Available at: http://www.fda.gov/ohrms/dockets/ac/02/slides/3837s1_02_Temple.ppt (accessed 06 April 2011).

[5] US Food and Drug Administration (FDA). *Guidance for Industry-Non-Inferiority Clinical Trials*. FDA: Rockville, 2010.

[6] *FDA News Drug Daily Bulletin* 2 March 2010; **2**(42).

[7] Huitfeldt B, Danielson L, Ebbutt A, Schmidt K. Choice of control in clinical trials–issues and implications of ICH E10. *Drug Information Journal* 2001; **35**:1147–1156.

[8] EFSPI comments on draft FDA. Guidance for Industry – Non-Inferiority Clinical Trials. Available at: http://www.efspi.org/aadocs/efspi_comments_fda_ni_final.pdf (accessed 10 May 2011).

[9] Peterson P, Carroll K, Chuang-Stein C, Ho Y-Y, Jiang Q, Gang L, Sanchez M, Sax R, Wang Y-C, Snapinn S. PISC expert team white paper: Toward a consistent standard of evidence when evaluating the efficacy of en experimental treatment from a randomized active-controlled trial. *Statistics in Biopharmaceutical Research* 2010; **2**:522–531. DOI: 10.1198/sbr.2010.09016.

[10] Rothmann M, Li N, Chen G, Chi GYH, Temple R, Tsou H-H. Design and analysis of non-inferiority mortality trial in oncology. *Statistics in Medicine* 2003; **22**:239–264.

[11] Eichler H-G, Bloechl-Daum B, Abadie E, Barnett D, König F, Pearson S. Relative efficacy of drugs: an emerging issue between regulatory agencies and third-party payers. *Nature Reviews* 2010; **9**:277–291.

[12] US Food and Drug Administration (FDA). *Guidance for Industry-Statistical approaches to establishing bioequivalence*. FDA: Rockville, 2001.

Copyright © 2011 John Wiley & Sons, Ltd.