Ref. 9

Guidance for Industry

E 10 Choice of Control Group and Related Issues in Clinical Trials

U.S. Department of Health and Human Services
Food and Drug Administration
Center for Drug Evaluation and Research (CDER)
Center for Biologics Evaluation and Research (CBER)

May 2001
ICH

Guidance for Industry

E 10 Choice of Control Group and Related Issues in Clinical Trials

Additional copies are available from:

Office of Training and Communication
Division of Drug Information, HFD-240
Center for Drug Evaluation and Research
Food and Drug Administration
5600 Fishers Lane
Rockville, MD 20857
(Tel) 301-827-4573
http://www.fda.gov/cder/guidance/index.htm

or

Office of Communication, Training and
Manufacturers Assistance, HFM-40
Center for Biologics Evaluation and Research
Food and Drug Administration
1401 Rockville Pike, Rockville, MD 20852-1448
http://www.fda.gov/cber/guidelines.htm
Fax: 1-888-CBERFAX or 301-827-3844
Phone: the Voice Information System at 800-835-4709 or 301-827-1800

U.S. Department of Health and Human Services
Food and Drug Administration
Center for Drug Evaluation and Research (CDER)
Center for Biologics Evaluation and Research (CBER)

May 2001
ICH

TABLE OF CONTENTS

I. INTRODUCTION (1.0) 1

A. GENERAL SCHEME AND PURPOSE OF GUIDANCE (1.1) 2
B. PURPOSE OF CONTROL GROUP (1.2) 3
1. *Randomization (1.2.1)* 3
2. *Blinding (1.2.2)* 4
C. TYPES OF CONTROLS (1.3) 4
1. *Placebo Concurrent Control (1.3.1)* 5
2. *No-treatment Concurrent Control (1.3.2)* 5
3. *Dose-response Concurrent Control (1.3.3)* 5
4. *Active (Positive) Concurrent Control (1.3.4)* 6
5. *External Control (Including Historical Control) (1.3.5)* 6
6. *Multiple Control Groups (1.3.6)* 6
D. PURPOSES OF CLINICAL TRIALS AND RELATED ISSUES (1.4) 6
1. *Evidence of Efficacy (1.4.1)* 7
2. *Comparative Efficacy and Safety (1.4.2)* 7
3. *Fairness of Comparisons (1.4.3)* 7
a. Dose (1.4.3.1) 8
b. Patient Population (1.4.3.2) 8
c. Selection and Timing of Endpoints (1.4.3.3) 8
E. ASSAY SENSITIVITY (1.5) 9
1. *Assay Sensitivity in Non-inferiority or Equivalence Trials (1.5.1)* 9
a. Historical Evidence of Sensitivity to Drugs Effects and Choosing the Non-inferiority Margin (1.5.1.1) 10
b. Appropriate Trial Conduct (1.5.1.2) 12
2. *Assay Sensitivity in Trials Intended to Demonstrate Superiority (1.5.2)* 13

II. DETAILED CONSIDERATION OF TYPES OF CONTROL (2.0) 14

A. PLACEBO CONTROL (2.1) 14
1. *Description (See Section 1.3.1) (2.1.1)* 14
2. *Ability to Minimize Bias (2.1.2)* 14
3. *Ethical Issues (2.1.3)* 14
4. *Usefulness of Placebo-controlled Trials and Validity of Inference in Particular Situations (2.1.4)* 15
5. *Modifications of Design and Combinations with Other Controls That Can Resolve Ethical, Practical, or Inferential Issues (2.1.5)* 16
a. Additional Control Groups (2.1.5.1) 16
b. Other Modifications of Study Design (2.1.5.2) 17
6. *Advantages of Placebo-controlled Trials (2.1.6)* 19
a. Ability to Demonstrate Efficacy (2.1.6.1) 19
b. Measures *Absolute* Efficacy and Safety (2.1.6.2) 19
c. Efficiency (2.1.6.3) 19
d. Minimizing the Effect of Subject and Investigator Expectations (2.1.6.4) 20
7. *Disadvantages of Placebo-controlled Trials (2.1.7)* 20
a. Ethical Concerns (See Sections 2.1.3 and 2.1.4) (2.1.7.1) 20
b. Patient and Physician Practical Concerns (2.1.7.2) 20
c. Generalizability (2.1.7.3) 20
d. No Comparative Information (2.1.7.4) 21
B. NO-TREATMENT CONCURRENT CONTROL (SEE SECTION 1.3.2) (2.2) 21
C. DOSE-RESPONSE CONCURRENT CONTROL (SEE SECTION 1.3.3) (2.3) 21
1. *Description (2.3.1)* 21
2. *Ability to Minimize Bias (2.3.2)* 22
3. *Ethical Issues (2.3.3)* 22
4. *Usefulness of Dose-response Studies and Validity of Inference in Particular Situations (2.3.4)* 22
5. *Modifications of Design and Combinations with Other Controls That Can Resolve Ethical, Practical, or Inferential Problems (2.3.5)* 22

6. *Advantages of Dose-response Trials (2.3.6)* *22*
a. Efficiency (2.3.6.1) 22
b. Possible Ethical Advantage (2.3.6.2) 23
7. *Disadvantages of Dose-response Study (2.3.7)* *23*
D. ACTIVE CONTROL (SEE SECTION 1.3.4) (2.4) 23
1. *Description (2.4.1)* *23*
2. *Ability to Minimize Bias (2.4.2)* *24*
3. *Ethical Issues (2.4.3)* *24*
4. *Usefulness of Active Control Trials; Validity of Inference in Particular Situations (2.4.4)* *24*
5. *Modifications of Design and Combinations with Other Controls That Can Resolve Ethical, Practical, or Inferential Issues (2.4.5)* *25*
6. *Advantages of Active Control Trials (2.4.6)* *25*
a. Ethical and Practical Advantages (2.4.6.1) 25
b. Information Content (2.4.6.2) 25
7. *Disadvantages of Active Control Trials (2.4.7)* *25*
a. Information Content (2.4.7.1) 25
b. Large Sample Size (2.4.7.2) 25
E. EXTERNAL CONTROL (INCLUDING HISTORICAL CONTROL, SEE SECTION 1.3.5) (2.5) 26
1. *Description (2.5.1)* *26*
2. *Ability to Minimize Bias (2.5.2)* *26*
3. *Ethical Issues (2.5.3)* *27*
4. *Usefulness of Externally Controlled Trials; Validity of Inference in Particular Situations (2.5.4)* *28*
5. *Modifications of Design and Combinations with Other Controls That Can Resolve Ethical, Practical or Inferential Problems (2.5.5)* *28*
6. *Advantages of Externally Controlled Trials (2.5.6)* *29*
7. *Disadvantages of Externally Controlled Trials (2.5.7)* *29*

III. CHOOSING THE CONCURRENT CONTROL GROUP (3.0) 29

Guidance for Industry[1]

E10 Choice of Control Group and Related Issues in Clinical Trials

This guidance represents the Food and Drug Administration's (FDA's) current thinking on this topic. It does not create or confer any rights for or on any person and does not operate to bind FDA or the public. An alternative approach may be used if such approach satisfies the requirements of the applicable statutes and regulations.

This guidance is intended to assist applicants in choosing a control group for clinical trials intended to demonstrate the efficacy of a treatment. The guidance also discusses related trial design and conduct issues and describes what trials using each design can demonstrate. This guidance does not address the regulatory requirements of any region.

I. INTRODUCTION (1.0)[2]

The choice of control group is always a critical decision in designing a clinical trial. That choice affects the inferences that can be drawn from the trial, the ethical acceptability of the trial, the degree to which bias in conducting and analyzing the study can be minimized, the types of subjects that can be recruited and the pace of recruitment, the kind of endpoints that can be studied, the public and scientific credibility of the results, the acceptability of the results by regulatory authorities, and many other features of the study, its conduct, and its interpretation.

---

[1] This guidance was developed within the Expert Working Group (Efficacy) of the International Conference on Harmonisation of Technical Requirements for Registration of Pharmaceuticals for Human Use (ICH) and has been subject to consultation by the regulatory parties, in accordance with the ICH process. This document has been endorsed by the ICH Steering Committee at *Step 4* of the ICH process, July 2000. At *Step 4* of the process, the final draft is recommended for adoption to the regulatory bodies of the European Union, Japan, and the United States.

[2] Arabic numbers reflect the organizational breakdown in the document endorsed by the ICH Steering Committee at *Step 4* of the ICH process, July 2000.

A. General Scheme and Purpose of Guidance (1.1)

The purpose of this guideance is to describe the general principles involved in choosing a control group for clinical trials intended to demonstrate the efficacy of a treatment and to discuss related trial design and conduct issues. This guidance does not address the regulatory requirements in any region, but describes what trials using each design can demonstrate. The general principles described in this guidance are relevant to any controlled trial but the choice of control group is of particularly critical importance to clinical trials carried out during drug development to demonstrate efficacy. The choice of the control group should be considered in the context of available standard therapies, the adequacy of the evidence to support the chosen design, and ethical considerations.

This guidance first describes the purpose of the control group and the types of control groups commonly employed to demonstrate efficacy. It then discusses the critical design and interpretation issues associated with the use of an active control trial to demonstrate efficacy by showing non-inferiority or equivalence to the control (Section 1.5). There are circumstances in which a finding of non-inferiority cannot be interpreted as evidence of efficacy. Specifically, for a finding of non-inferiority to be interpreted as showing efficacy, the trial needs to have had the ability to distinguish effective from less effective or ineffective treatments.

The guidance then describes trials using each kind of control group in more detail (see sections 2.0-2.5.7) and considers, for each:

- Its ability to minimize bias
- Ethical and practical issues associated with its use
- Its usefulness and the quality of inference in particular situations
- Modifications of study design or combinations with other controls that can resolve ethical, practical, or inferential concerns
- Its overall advantages and disadvantages

Several other ICH guidances are particularly relevant to this guidance:

- E3: Structure and Content of Clinical Study Reports
- E4: Dose-Response Information to Support Drug Registration
- E5: Ethnic Factors
- E6: Good Clinical Practice: Consolidated Guideline
- E8: General Considerations for Clinical Trials
- E9: Statistical Principles for Clinical Trials

Although trials using any of the control groups described and discussed in this guidance may be useful and acceptable in clinical trials that serve as the basis for marketing approval in at least some circumstances, they are not equally appropriate or useful in every case. The general approach to selecting the type of control is outlined in Section 3.0, Figure 1, and Table 1.

Although this guidance is focused primarily on clinical trials intended to assess the efficacy of a treatment, many of the considerations discussed also apply to the assessment of specific safety hypotheses and to safety or efficacy comparisons of two treatments.

B. Purpose of Control Group (1.2)

Control groups have one major purpose: to allow discrimination of patient outcomes (for example, changes in symptoms, signs, or other morbidity) caused by the test treatment from outcomes caused by other factors, such as the natural progression of the disease, observer or patient expectations, or other treatment. The control group experience tells us what would have happened to patients if they had not received the test treatment or if they had received a different treatment known to be effective.

If the course of a disease were uniform in a given patient population, or predictable from patient characteristics such that outcome could be predicted reliably for any given subject or group of subjects, results of treatment could simply be compared with the known outcome without treatment. For example, one could assume that pain would have persisted for a defined time, blood pressure would not have changed, depression would have lasted for a defined time, tumors would have progressed, or the mortality after an acute infarction would have been the same as previously seen. In unusual cases, the course of illness is in fact predictable in a defined population and it may be possible to use a similar group of patients previously studied as a historical control (see section 1.3.5). In most situations, however, a concurrent control group is needed because it is not possible to predict outcome with adequate accuracy or certainty.

A concurrent control group is one chosen from the same population as the test group and treated in a defined way as part of the same trial that studies the test treatment, and over the same period of time. The test and control groups should be similar with regard to all baseline and on-treatment variables that could influence outcome, except for the study treatment. Failure to achieve this similarity can introduce a bias into the study. Bias here (and as used in ICH E9) means the systematic tendency of any aspects of the design, conduct, analysis, and interpretation of the results of clinical trials to make the estimate of a treatment effect deviate from its true value. Randomization and blinding are the two techniques usually used to minimize the chance of such bias and to ensure that the test treatment and control groups are similar at the start of the study and are treated similarly in the course of the study (see ICH E9). Whether a trial design includes these features is a critical determinant of its quality and persuasiveness.

1. Randomization (1.2.1)

Assurance that subject populations are similar in test and control groups is best attained by randomly dividing a single sample population into groups that receive the test or control treatments. Randomization avoids systematic differences between groups with respect to known or unknown baseline variables that could affect outcome. Inability to eliminate systematic differences between treatment groups is a major problem of studies without a concurrent randomized control (see external control trials, section 1.3.5). Randomization also provides a sound basis for statistical inference.

2. *Blinding (1.2.2)*

The groups should not only be similar at baseline, but should be treated and observed similarly during the trial, except for receiving the test and control drug. Clinical trials are often *double-blind* (or *double-masked*), meaning that both subjects and investigators, as well as sponsor or investigator staff involved in the treatment or clinical evaluation of subjects, are unaware of each subject's assigned treatment. Blinding is intended to minimize the potential biases resulting from differences in management, treatment, or assessment of patients, or interpretation of results that could arise as a result of subject or investigator knowledge of the assigned treatment. For example:

- Subjects on active drug might report more favorable outcomes because they expect a benefit or might be more likely to stay in a study if they knew they were on active drug.
- Observers might be less likely to identify and report treatment responses in a no-treatment group or might be more sensitive to a favorable outcome or adverse event in patients receiving active drug.
- Knowledge of treatment assignment could affect vigor of attempts to obtain on-study or follow-up data.
- Knowledge of treatment assignment could affect decisions about whether a subject should remain on treatment or receive concomitant medications or other ancillary therapy.
- Knowledge of treatment assignment could affect decisions as to whether a given subject's results should be included in an analysis.
- Knowledge of treatment assignment could affect choice of statistical analysis.

Blinding is intended to ensure that subjective assessments and decisions are not affected by knowledge of treatment assignment.

C. Types of Controls (1.3)

Control groups in clinical trials can be classified on the basis of two critical attributes: (1) the type of treatment used and (2) the method of determining who will be in the control group. The type of control treatment may be any of the following four: (1) placebo, (2) no treatment, (3) different dose or regimen of the study treatment, or (4) a different active treatment. The principal methods of determining who will be in the control group are by randomization or by selection of a control population separate from the population treated in the trial (external or historical control). This document categorizes control groups into five types. The first four are concurrently controlled (the control group and test groups are chosen from the same population and treated concurrently), usually with random assignment to treatment; they are distinguished by the type of control treatment (listed above) used. External (historical) control groups, regardless of the comparator treatment, are considered together as the fifth type because of serious concerns about the ability of such trials to ensure comparability of test and control groups and their ability to minimize important biases, making this design usable only in unusual circumstances.

It is increasingly common to carry out studies that have more than one type of control group. Each type of control group is appropriate in some circumstances, but none is usable or adequate in every situation. The five types of control group are:

1. *Placebo Concurrent Control (1.3.1)*

In a placebo-controlled trial, subjects are randomly assigned to a test treatment or to an identical-appearing treatment that does not contain the test drug. The treatments may be titrated to effect or tolerance, or may be given at one or more fixed doses. Such trials are almost always double-blind. The name of the control suggests that its purpose is to control for *placebo* effect (improvement in a subject resulting from thinking that he or she is taking a drug), but that is not its only or major benefit. Rather, the placebo control design, by allowing blinding and randomization and including a group that receives an inert treatment, controls for all potential influences on the actual or apparent course of the disease other than those arising from the pharmacologic action of the test drug. These influences include spontaneous change (natural history of the disease and regression to the mean), subject or investigator expectations, the effect of being in a trial, use of other therapy, and subjective elements of diagnosis or assessment. Placebo-controlled trials seek to show a difference between treatments when they are studying effectiveness, but may also seek to show lack of difference (of specified size) in evaluating a safety measurement. In that case, the question of whether the trial could have shown such a difference if there had been one is critical (see section 1.5).

The use of a placebo control group does not imply that the control group is untreated. In many placebo-controlled trials, the new treatment and placebo are each added to a common standard therapy (so-called *add-on studies*, see section 2.1.5.2.1).

2. *No-treatment Concurrent Control (1.3.2)*

In a no treatment-controlled trial, subjects are randomly assigned to test treatment or to no (i.e., absence of) study treatment. The principal difference between this design and a placebo-controlled trial is that subjects and investigators are not blind to treatment assignment. Because of the advantages of double-blind designs, this design is likely to be needed and suitable only when it is difficult or impossible to double-blind (e.g., treatments with easily recognized toxicity) and only when there is reasonable confidence that study endpoints are objective and that the results of the trial are unlikely to be influenced by the factors listed in section 1.2.2. Note that it is often possible to have a blinded evaluator carry out endpoint assessment, even if the overall trial is not double-blind. This is a valuable approach and should always be considered in trials that cannot be blinded, but it does not solve the other problems associated with knowing the treatment assignment (see section 1.2.2).

3. *Dose-response Concurrent Control (1.3.3)*

In a randomized, fixed-dose, dose-response trial, subjects are randomized to one of several fixed-dose groups. Subjects may either be placed on their fixed dose initially or be raised to that dose gradually, but the intended comparison is between the groups on their final dose. Dose-response

trials are usually double-blind. They may include a placebo (zero dose) and/or active control. In a concentration-controlled trial, treatment groups are titrated to several fixed-concentration windows; this type of trial is conceptually similar to a fixed-dose, dose-response trial. In a regimen-controlled trial subjects are randomized to two or more regimens of the study drug (e.g., once vs. twice daily, 3 days vs. 7 days).

4. Active (Positive) Concurrent Control (1.3.4)

In an active control (or positive control) trial, subjects are randomly assigned to the test treatment or to an active control treatment. Such trials are usually double-blind, but this is not always possible; many oncology trials, for example, are considered difficult or impossible to blind (see section 1.3.2) because of different regimens, different routes of administration, and different toxicities. Active control trials can have two distinct objectives with respect to showing efficacy: (1) to show efficacy of the test treatment by showing it is as good as a known effective treatment or (2) to show efficacy by showing superiority of the test treatment to the active control. They may also be used with the primary objective of comparing the efficacy and/or safety of the two treatments (see section 1.4). Whether the purpose of the trial is to show efficacy of the new treatment or to compare two treatments, the question of whether the trial would be capable of distinguishing effective from less effective or ineffective treatments is critical (see section 1.5).

5. External Control (Including Historical Control) (1.3.5)

An externally controlled trial compares a group of subjects receiving the test treatment with a group of patients external to the study, rather than to an internal control group consisting of patients from the same population assigned to a different treatment. The external control can be a group of patients treated at an earlier time (historical control) or a group treated during the same time period but in another setting. The external control may be defined (a specific group of patients) or non-defined (a comparator group based on general medical knowledge of outcome). Use of this latter comparator is particularly treacherous (such trials are usually considered uncontrolled) because general impressions are so often inaccurate. So-called baseline-controlled studies, in which subjects' status on therapy is compared with status before therapy (e.g., blood pressure, tumor size), have no internal control and are thus uncontrolled or externally controlled (see section 2.5).

6. Multiple Control Groups (1.3.6)

As will be described further below (see section 1.5.1), it is often possible and advantageous to use more than one kind of control in a single study, e.g., use of both an active control and placebo. Similarly, trials can use several doses of test drug and several doses of an active control, with or without placebo. This design may be useful for active drug comparisons where the relative potency of the two drugs is not well established, or where the purpose of the trial is to establish relative potency.

D. Purposes of Clinical Trials and Related Issues (1.4)

Two purposes of clinical trials should be distinguished: (1) assessment of the efficacy and/or safety of a treatment and (2) assessment of the relative (comparative) efficacy, safety, risk/benefit relationship or utility of two treatments.

1. *Evidence of Efficacy (1.4.1)*

A trial using any of the control types may demonstrate efficacy of the test treatment by showing that it is superior to the control (placebo, no treatment, and low dose of test drug, active drug). An active control trial may, in addition, demonstrate efficacy in some cases by showing the new treatment to be similar in efficacy to a known effective treatment. This similarity establishes the efficacy of the test treatment, however, only if it can be assumed that the active control was effective under the conditions of the trial, as two treatments would also look similar if neither were effective in the trial (see section 1.5).

Clinical trials designed to demonstrate efficacy of a new drug by showing that it is similar in efficacy to a standard agent have been called *equivalence* trials. Most of these are actually non-inferiority trials, attempting to show that the new drug is not less effective than the control by more than a defined amount, generally called the margin.

2. *Comparative Efficacy and Safety (1.4.2)*

In some cases, the focus of the trial is on the comparison of one treatment with another treatment, not the efficacy of the test drug per se. Depending on the therapeutic area, these trials may be seen as providing information that is important for relative risk/benefit assessment. The active comparator(s) should be acceptable to the region for which the data are intended. It is not necessary to demonstrate superiority to the active comparator, and, depending on the situation, it may not be necessary to show non-inferiority. For example, a less effective treatment could have safety advantages and thus be considered useful.

Even though the primary focus of such a trial is the comparison of treatments, rather than demonstration of efficacy, the cautions described for conducting and interpreting non-inferiority trials need to be taken into account (see section 1.5). Specifically, the ability of the comparative trial to detect a difference between treatments when one exists needs to be established because a trial incapable of distinguishing between treatments that are in fact different cannot provide useful comparative information.

3. *Fairness of Comparisons (1.4.3)*

For the comparative trial to be informative concerning relative safety and/or efficacy, the trial needs to be fair; i.e., the conditions of the trial should not inappropriately favor one treatment over the other. In practice, an active control equivalence or non-inferiority trial offered as evidence of efficacy also almost always needs to provide a fair effectiveness comparison with the control, because any doubt as to whether the control in the study had its usual effect would undermine assurance that the trial had assay sensitivity (see section 1.5). Among aspects of trial

design that could unfairly favor one treatment are choice of dose or patient population and selection and timing of endpoints.

a. Dose (1.4.3.1)

In comparing the test drug with an active control, it is important to choose an appropriate dose and dose regimen of the control and test drugs. In examining the results of a comparison of two treatments, it is important to consider whether an apparently less effective treatment has been used at too low a dose or whether the apparently less well-tolerated treatment has been used at too high a dose. In some cases, to show superior efficacy or safety convincingly it will be necessary to study several doses of the control and perhaps several doses of the test treatment.

b. Patient Population (1.4.3.2)

Selection of subjects for an active control trial can affect outcome; the population studied should be carefully considered in evaluating what the trial has shown. For example, if many subjects in a trial have previously failed to respond to the control treatment, there would be a bias in favor of the new treatment. The results of such a trial could not be generalized to the entire population of previously untreated patients. A finding of superiority of the new treatment, however, still would be evidence of the efficacy of the new treatment in the population studied. In fact, a trial of a new treatment in apparent nonresponders to another treatment, in which the nonresponders are randomized to either the new or failed treatment (so long as this does not place the patients at risk), can provide a demonstration of the value of the new treatment in such nonresponders, a clinically valuable observation.

Similarly, it is sometimes possible to identify patient subsets more or less likely to have a favorable response or to have an adverse response to a particular drug. For example, blacks usually respond poorly to the blood pressure effects of beta blockers and angiotensin- converting enzyme inhibitors, so that a comparison of a new antihypertensive with these drugs in these patients would tend to show superiority of the new drug. It would not be appropriate to conclude that the new drug is generally superior. Again, however, a planned trial in a subgroup, with recognition of its limitations and of what conclusion can properly be drawn, could be informative.

c. Selection and Timing of Endpoints (1.4.3.3)

When two treatments are used for the same disease or condition, they may differentially affect various outcomes of interest in that disease, particularly if they represent different classes or modalities of treatment. Therefore, when comparing them in a clinical trial, the choice and timing of endpoints may favor one treatment or the other. For example, thrombolytics in patients with acute myocardial infarction can reduce mortality but increase hemorrhagic stroke risk. If a new, more pharmacologically active, thrombolytic were compared with an older thrombolytic, the more active treatment might look better if the endpoint were mortality, but worse if the endpoint were a composite of mortality and disabling stroke. Similarly, in comparing two analgesics in the management of dental pain, assigning a particularly heavy weight to pain at

early time points would favor the drug with more rapid onset of effect, while assigning more weight to later time points would favor a drug with a longer duration of effect.

E. Assay Sensitivity (1.5)

Assay sensitivity is a property of a clinical trial defined as the ability to distinguish an effective treatment from a less effective or ineffective treatment. Assay sensitivity is important in any trial but has different implications for trials intended to show differences between treatments (superiority trials) and trials intended to show non-inferiority. If a trial intended to demonstrate efficacy by showing superiority of a test treatment to control lacks assay sensitivity, it will fail to show that the test treatment is superior and will fail to lead to a conclusion of efficacy. In contrast, if a trial is intended to demonstrate efficacy by showing a test treatment to be non-inferior to an active control, but lacks assay sensitivity, the trial may find an ineffective treatment to be non-inferior and could lead to an erroneous conclusion of efficacy.

When two treatments within a trial are shown to have different efficacy (i.e., when one treatment is superior), that finding itself demonstrates that the trial had assay sensitivity. In contrast, a successful non-inferiority trial (i.e., one that has shown non-inferiority), or an unsuccessful superiority trial, generally does not contain such direct evidence of assay sensitivity.

1. Assay Sensitivity in Non-inferiority or Equivalence Trials (1.5.1)

The presence of assay sensitivity in a non-inferiority or equivalence trial may be deduced from two determinations:

Historical evidence of sensitivity to drug effects, i.e., that similarly designed trials in the past regularly distinguished effective treatments from less effective or ineffective treatments and

Appropriate trial conduct, i.e., that the conduct of the trial did not undermine its ability to distinguish effective treatments from less effective or ineffective treatments.

Historical evidence of sensitivity to drug effects can, and should, be evaluated before beginning a non-inferiority trial. Specifically, it should be determined that, in the specific therapeutic area under study, appropriately designed and conducted trials that used a specific active treatment, or other treatments with similar effects, reliably showed an effect. Optimally, this is demonstrated by finding that the active treatment intended for use as the active control was reliably found superior to placebo. If this is the case, there is historical evidence of sensitivity to drug effects for similarly designed active control trials (see section 1.5.1.1).

Appropriateness of trial conduct can only be fully evaluated after the active control non-inferiority trial is completed. Not only should the design of the non-inferiority trial be similar to that of previous trials used to determine historical evidence of sensitivity to drug effects (e.g., entry criteria, allowable concomitant therapy); but, in addition, the actual study population entered, the concomitant therapies actually used, etc., should be assessed to ensure that conduct of the study was, in fact, similar to the previous trials. The trial should also be conducted with

high quality (e.g. good compliance, few losses to follow-up). Together with historical evidence of sensitivity to drug effects, appropriate trial conduct (section 1.5.1.2) provides assurance of assay sensitivity in the new active control trial.

The design and conduct of a non-inferiority trial thus involve four critical steps:

Determining that historical evidence of sensitivity to drug effects exists. Without this determination, demonstration of efficacy from a showing of non-inferiority is not possible and should not be attempted.

Designing a trial. Important details of the trial design, e.g., study population, concomitant therapy, endpoints, run-in periods, should adhere closely to the design of the trials used to determine that historical evidence of sensitivity to drug effects exists

Setting a margin. An acceptable non-inferiority margin should be defined, taking into account the historical data and relevant clinical and statistical considerations.

Conducting the trial. The trial conduct should also adhere closely to that of the historical trials and should be of high quality.

a. Historical Evidence of Sensitivity to Drug Effects and Choosing the Non-inferiority Margin (1.5.1.1)

As noted earlier, most active control equivalence trials are really non-inferiority trials intended to establish the efficacy of a new treatment. Analysis of the results of non-inferiority trials is discussed in ICH guidances E9 and E3. Briefly, in such a trial, test and known effective treatments are compared. Prior to the trial, an equivalence or non-inferiority margin, sometimes called *delta*, is selected. This margin is the degree of inferiority of the test treatments to the control that the trial will attempt to exclude statistically. If the confidence interval for the difference between the test and control treatments excludes a degree of inferiority of the test treatment as large as, or larger than, the margin, the test treatment can be declared non-inferior; if the confidence interval includes a difference as large as the margin, the test treatment cannot be declared non-inferior.

The margin chosen for a non-inferiority trial cannot be greater than the *smallest effect size that the active drug would be reliably expected to have* compared with placebo in the setting of the planned trial. If a difference between active control and the new drug favors the control by as much as or more than this margin, the new drug might have no effect at all. Identification of the smallest effect size that the active drug would be reliably expected to have is only possible when there is historical evidence of sensitivity to drug effects and, indeed, identification of the margin is based upon that evidence. The margin generally is identified based on past experience in placebo-controlled trials of adequate design under conditions similar to those planned for the new trial, but could also be supported by dose response or active control superiority studies. Regardless of the control groups used in those earlier studies, the value of interest in determining the margin is the measure of superiority of the active treatment to its control, not uncontrolled

measures such as change from baseline. Note that exactly how to calculate the margin is not described in this document, and there is little published experience on how to do this.

The determination of the margin in a non-inferiority trial is based on both statistical reasoning and clinical judgment, should reflect uncertainties in the evidence on which the choice is based, and should be suitably conservative. If this is done properly, a finding that the confidence interval for the difference between new drug and the active control excludes a suitably chosen margin provides assurance that the test drug has an effect greater than zero. In practice, the non-inferiority margin chosen usually will be smaller than that suggested by the smallest expected effect size of the active control because of interest in ensuring that some clinically acceptable effect size (or fraction of the control drug effect) was maintained. For example, it would not generally be considered sufficient in a mortality non-inferiority study to ensure that the test treatment had an effect greater than zero; retention of some substantial fraction of the mortality effect of the control would usually be sought. This would also be true in a trial whose primary focus is the relative effectiveness of a test drug and active control (see section 1.4.2), where it would be usual to seek assurance that the test and control drug were quite similar, not simply that the new drug had any effect at all.

The fact that the choice of the margin to be excluded is based on historical evidence gives the non-inferiority trial an element in common with a historically controlled (externally controlled) trial. The non-inferiority trial design is appropriate and reliable only when the historical estimate of drug effect size can be well supported by reference to the results of previous studies of the control drug. These studies should lead to the conclusion that the active control can consistently be distinguished from placebo in appropriately sized trials of design similar to the proposed trial and should identify an effect size that represents the smallest effect that the control can reliably be expected to have. If placebo-controlled trials of a design similar to the one proposed more than occasionally show no difference between the proposed active control and placebo, and this cannot be explained by some characteristic of the study, only superiority of the test drug would be interpretable.

Whether there is historical evidence of sensitivity to drug effects in any given case is to some degree a matter of judgment. In some cases sensitivity to drug effects is clear from the consistency of results of prior placebo-controlled trials or is obvious because the outcome of treated and untreated disease is very different. For example, in many infectious diseases cure rates on effective treatment far exceed the spontaneous cure rates over the course of a short-term study. There are many conditions, however, in which drugs considered effective cannot regularly be shown superior to placebo in well-controlled trials; and one therefore cannot reliably determine a minimum effect the drug will have in the setting of a specific trial. Such conditions tend to include those in which there is substantial improvement and variability in placebo groups, and/or in which the effects of therapy are small or variable, such as depression, anxiety, dementia, angina, symptomatic congestive heart failure, seasonal allergies, and symptomatic gastroesophageal reflux disease.

In all these cases, there is no doubt that the standard treatments are effective because there are many well-controlled trials of each of these drugs that have shown an effect. Based on available

experience, however, it would be difficult to describe trial conditions in which the drug would reliably have at least a minimum effect (i.e., conditions in which there is historical evidence of sensitivity to drug effects) and that, therefore, could be used to identify an appropriate margin. In some cases, the experience on which the historical evidence of sensitivity to drug effects is based may be of questionable relevance, e.g., if standards of treatment and diagnosis have changed substantially over time (for an example, see section 2.1.7.1). If someone proposing to use an active-control or non-inferiority design cannot provide sufficient support for historical evidence of the sensitivity to drug effects of the study with the chosen non-inferiority margin, a finding of non-inferiority cannot be considered informative with respect to efficacy.

As noted, a determination regarding historical evidence of sensitivity to drug effects applies only to trials of a specific design. For a planned non-inferiority trial to be similarly sensitive to drug effects, it is essential that the trial have critical design characteristics similar to those of the historical trials. These design characteristics include, for example, the entry criteria (severity of medical condition, concomitant illness, method of diagnosis), dose and regimen of control drug, concomitant treatments used, the endpoint measured and timing of assessments, and the use of a washout period to exclude selected patients. When differences in study design characteristics are unavoidable or desirable (e.g. because of technological or therapeutic advances), the implications of any differences for the determination of the presence of historical evidence of sensitivity to drug effects and for choice of margin should be carefully considered.

b. Appropriate Trial Conduct (1.5.1.2)

Even where there is historical evidence of sensitivity to drug effects and the new study is similar in design to the past studies, assay sensitivity can be undermined by the actual conduct of the trial. To ensure assay sensitivity of a trial, its conduct should be of high quality and the patients actually enrolled, the treatments (other than the test treatment) actually given, and the assessments actually made should be similar to those of the trials on which the determination of historical sensitivity to drug effects was based.

There are many factors in the conduct of a trial that can reduce the observed difference between an effective treatment and a less effective or ineffective treatment and therefore may reduce a trial's assay sensitivity, such as:

1. Poor compliance with therapy
2. Poor responsiveness of the enrolled study population to drug effects
3. Use of concomitant non-protocol medication or other treatment that interferes with the test drug or that reduces the extent of the potential response
4. An enrolled population that tends to improve spontaneously, leaving no room for further drug-induced improvement
5. Poorly applied diagnostic criteria (patients lacking the disease to be studied)
6. Biased assessment of endpoint because of knowledge that all patients are receiving a potentially active drug, e.g., a tendency to read blood pressure responses as normalized, potentially reducing the difference between test drug and control

Clinical researchers and trial sponsors intend to perform high-quality trials, and the availability of the Good Clinical Practices guidance (ICH E6) will continue to enhance trial quality. Nonetheless, it should be appreciated that in trials intended to show a difference between treatments there is a strong imperative to use a good trial design and minimize trial errors because many trial imperfections increase the likelihood of failing to show a difference between treatments when one exists. In placebo-controlled trials many efforts are made to improve compliance and increase the likelihood that the patient population will be responsive to drug effects to ensure that an effective treatment will be distinguished from placebo. Nonetheless, in many clinical settings, despite the strong stimulus and extensive efforts to ensure trial excellence and assay sensitivity, clinical trials are often unable to reliably distinguish effective drugs from placebo.

In contrast, in trials intended to show that there is not a difference of a particular size (non-inferiority) between two treatments, there may be a much weaker stimulus to engage in many of these efforts to ensure study quality that will help ensure that differences will be detected, i.e., that ensure assay sensitivity. The kinds of trial error that diminish observed differences between treatments (e.g., poor compliance, high placebo response, certain concomitant treatment, misclassification of outcomes) are of particular concern with respect to preservation of assay sensitivity. However, when it is believed that the new drug is actually superior to the control, there will be a strong stimulus to conduct a high quality trial so that the non-inferiority margin is more likely to be excluded. It should also be noted that some kinds of trial errors can increase variance, which would decrease the likelihood of showing non-inferiority by widening the confidence interval, so that a difference between treatment and control greater than the margin could not be excluded. There would therefore be a strong stimulus in non-inferiority trials to reduce such sources of variance as poor measurement technique.

As noted, to determine that a non-inferiority trial had appropriate trial conduct, its conduct should be reviewed not only for the presence of factors that might obscure differences between treatments but also for factors that might make the trial different from the trials that provided the basis for determining the non-inferiority margin. In particular, it should be determined whether any observed differences in the populations enrolled, the use of concomitant therapies, compliance with therapy, and the extent of, and reasons for, dropping out could adversely affect assay sensitivity. Even when the design and conduct of a trial appear to have been quite similar to those of the trials providing the basis for determining the non-inferiority margin, outcomes with the active control treatment that are visibly atypical (e.g., cure rate in an antibiotic trial that is unusually high or low) can indicate that important differences existed.

2. Assay Sensitivity in Trials Intended to Demonstrate Superiority (1.5.2)

The question of assay sensitivity, although particularly critical in non-inferiority trials, actually arises in any trial that fails to detect a difference between treatments, including a placebo-controlled trial and a dose-response trial. If a treatment fails to show superiority to placebo, for example, it means either that the treatment was ineffective or that the study as designed and conducted was not capable of distinguishing an effective treatment from placebo.

A useful approach to the assessment of assay sensitivity in active control trials and in placebo-controlled trials is the *three-arm trial*, including both placebo and a known active treatment, a trial design with several advantages. Such a trial measures effect size (test drug versus placebo) and allows comparison of test drug and active control in a setting where assay sensitivity is established by the active control versus placebo comparison. (See Section 2.1.5.1.1).

II. DETAILED CONSIDERATION OF TYPES OF CONTROL (2.0)

A. Placebo Control (2.1)

1. Description (See Section 1.3.1) (2.1.1)

In a placebo-controlled trial, subjects are assigned, almost always by randomization, to either a test treatment or to a placebo. A placebo is a *dummy* treatment that appears as identical as possible to the test treatment with respect to physical characteristics such as color, weight, taste and smell, but that does not contain the test drug. Some trials may study more than one dose of the test treatment or include both an active control and placebo. In these cases, it may be easier for the investigator to use more than one placebo (*double-dummy*) than to try to make all treatments look the same. The use of placebo facilitates, and is almost always accompanied by, double-blinding (or double-masking). The difference in outcome between the active treatment and placebo groups is the measure of treatment effect under the conditions of the trial. Within this general description there are a wide variety of designs that can be used successfully: Parallel or crossover designs (see ICH E9), single fixed dose or titration in the active drug group, several fixed doses. Several designs meriting special attention will be described below. Note that not every study that includes a placebo is a placebo-controlled study. For example, an active control study could use a placebo for each drug (double-dummy) to facilitate blinding; this is still an active control trial, not a placebo-controlled trial. A placebo-controlled trial is one in which treatment with a placebo is compared with treatment with a test drug.

It should also be noted that not all placebos are completely inactive. For example, some vehicle controls used in studies of topical skin preparations may have beneficial activity. This does not impair the ability of the design to measure the specific effect of the test agent. Special problems arise when the chosen vehicle control may have harmful effects. In this case a *no treatment* arm would allow the measurement of the total effect of the test agent plus its vehicle.

2. Ability to Minimize Bias (2.1.2)

The placebo-controlled trial, using randomization and blinding, generally minimizes subject and investigator bias. Such trials, however, are not impervious to blind-breaking through recognition of pharmacologic effects of one treatment; blinded outcome assessment can enhance bias reduction in such cases. This concern may be particularly relevant in crossover studies.

3. Ethical Issues (2.1.3)

When a new treatment is tested for a condition for which no effective treatment is known, there is usually no ethical problem with a study comparing the new treatment to placebo. Use of a placebo control may raise problems of ethics, acceptability, and feasibility, however, when an effective treatment is available for the condition under study in a proposed trial. In cases where an available treatment is known to prevent serious harm, such as death or irreversible morbidity in the study population, it is generally inappropriate to use a placebo control. There are occasional exceptions, however, such as cases in which standard therapy has toxicity so severe that many patients have refused to receive it.

In other situations, when there is no serious harm, it is generally considered ethical to ask patients to participate in a placebo-controlled trial, even if they may experience discomfort as a result, provided the setting is noncoercive and patients are fully informed about available therapies and the consequences of delaying treatment. Such trials, however, even if ethical, may pose important practical problems. For example, deferred treatment of pain or other symptoms may be unacceptable to patients or physicians and they may not want to participate in a trial that requires this. Whether a particular placebo controlled trial of a new agent will be acceptable to subjects and investigators when there is known effective therapy is a matter of investigator, patient, and institutional review board (IRB)/ independent ethics committee (IEC) judgment, and acceptability may differ among ICH regions. Acceptability could depend on the specific design of the trial and the patient population chosen, as will be discussed below (see section 2.1.5).

Whether a particular placebo-controlled trial is ethical may in some cases depend on what is believed to have been clinically demonstrated under the particular circumstances of the trial. For example, a short term placebo-controlled trial of a new antihypertensive agent in patients with mild essential hypertension and no end-organ disease might be considered generally acceptable, while a longer trial, or one that included sicker patients, probably would not be.

It should be emphasized that use of a placebo or no-treatment control does not imply that the patient does not get any treatment at all. For example, in an oncology trial, when no active drug is approved, patients in both the placebo or no-treatment group and the test drug group will receive needed palliative treatment, such as analgesics, and best supportive care. Many placebo-controlled trials are conducted as *add-on trials*, where all patients receive a specified standard therapy or therapy left to the choice of the treating physician or institution (see section 2.1.5.2.1).

4. Usefulness of Placebo-controlled Trials and Validity of Inference in Particular Situations (2.1.4)

When used to show effectiveness of a treatment, the placebo-controlled trial is as free of assumptions and reliance on external (extra-study) information as it is possible to be. Most problems in the design or conduct of a trial increase the likelihood of failure to demonstrate a treatment difference (and thereby establish efficacy), so that the trial contains built-in incentives for trial excellence. Even when the primary purpose of a trial is comparison of two active agents or assessment of dose-response, the addition of a placebo provides an internal standard that enhances the inferences that can be drawn from the other comparisons.

Placebo-controlled trials also provide the maximum ability to distinguish adverse effects caused by a drug from those resulting from underlying disease or intercurrent illness. Note, however, that when used to show similarity of two treatments, for example, to show that a drug does not have a particular adverse effect by showing similar rates of the event in drug-treated and placebo-treated patients, placebo-controlled trials have the same assay sensitivity problem as any equivalence or non-inferiority trial (see section 1.5.1). To interpret the result, one must know that if the study drug had caused an adverse event, the event would have been observed. Ordinarily, such a study should include an active control treatment that does cause the adverse event in question, but in some cases it may be possible to conclude that a study has assay sensitivity to such an effect by documenting *historical sensitivity to adverse drug effects* for a particular study design.

5. Modifications of Design and Combinations with Other Controls That Can Resolve Ethical, Practical, or Inferential Issues (2.1.5)

It is often possible to address the ethical or practical limitations of placebo-controlled trials by using modified study designs that still retain the inferential advantages of these trials. In addition, placebo-controlled trials can be made more informative by including additional treatment groups, such as multiple doses of the test agent or a known active control treatment.

a. Additional Control Groups (2.1.5.1)

i. Three-arm Trial; Placebo and Active Control (2.1.5.1.1)

As noted in section 1.5.1, three-arm trials including an active control as well as a placebo-control group can readily assess whether a failure to distinguish test treatment from placebo implies ineffectiveness of the test treatment or is simply the result of a trial that lacked the ability to identify an active drug. The comparison of placebo to standard drug in such a trial provides internal evidence of assay sensitivity. It is possible to make the active groups larger than the placebo group to improve the precision of the active drug comparison, if this is considered important. This may also make the trial more acceptable to patients and investigators, as there is less chance of being randomized to placebo.

ii. Additional Doses (2.1.5.1.2)

Randomization to several fixed doses of the test drug in addition to placebo allows assessment of dose-response and may be particularly useful in a comparative trial to ensure a fair comparison of treatments (see ICH E4: Dose-Response Information to Support Drug Registration).

iii. Factorial Designs (2.1.5.1.3)

Factorial designs may be used to explore several doses of the investigational drug as monotherapy and in combination with several doses of another agent proposed for use in combination with it. A single study of this type can define the properties of a wide array of combinations. Such studies are common in the evaluation of new antihypertensive therapies, but

can be considered in a variety of settings where more than one treatment is used simultaneously. For example, the independent additive effects of aspirin and streptokinase in preventing mortality after a heart attack were shown in such a trial.

b. Other Modifications of Study Design (2.1.5.2)

i. Add on Study, Placebo-Controlled; Replacement Study. (2.1.5.2.1)

An *add-on* study is a placebo-controlled trial of a new agent conducted in people also receiving standard treatment. Such studies are particularly important when available treatment is known to decrease mortality or irreversible morbidity, and when a non-inferiority trial with standard treatment as the active control cannot be carried out or would be difficult to interpret (see section 1.5). It is common to study anticancer, antiepileptic, and heart failure drugs this way. This design is useful only when standard treatment is not fully effective (which, however, is almost always the case), and it has the advantage of providing evidence of improved clinical outcomes (rather than mere non-inferiority). Efficacy is, of course, established by such studies only for the combination treatment, and the dose in a monotherapy situation might be different from the dose found to be effective in combination. In general, this approach is likely to succeed only when the new and standard treatments possess different pharmacologic mechanisms, although there are exceptions. For example, combination treatments for people with AIDS may show a beneficial effect of pharmacologically related drugs because of delays in development of resistance.

A variation of this design that can sometimes give information on monotherapy, and that is particularly applicable in the setting of chronic disease, is the replacement study, in which the new drug or placebo is added by random assignment to conventional treatment given at an effective dose and the conventional treatment is then withdrawn, usually by tapering. The ability to maintain the subjects' baseline status is then observed in the drug and placebo groups using predefined success criteria. This approach has been used to study steroid-sparing substitutions in steroid-dependent patients, avoiding initial steroid withdrawal and the recrudescence of symptoms in a washout period. The approach has also been used to study antiepileptic drug monotherapy.

ii. Early *Escape*; Rescue Treatment (2.1.5.2.2)

It is possible to design a study to plan for *early escape* from ineffective therapy. Early escape refers to prompt removal of subjects whose clinical status worsens or fails to improve to a defined level (blood pressure not controlled by a prespecified time, seizure rate greater than some prescribed value, blood pressure rising to a certain level, angina frequency above a defined level, liver enzymes failing to normalize by a preset time in patients with hepatitis), who have a single event that treatment was intended to prevent (first recurrence of unstable angina, grand mal seizure, paroxysmal supraventricular arrhythmia), or who otherwise require rescue treatment. In such cases, the need to change treatment becomes a study endpoint. The criteria for deciding whether these endpoints have occurred should be well specified, and the timing of measurements should ensure that patients would not remain untreated with an active drug while their disease is poorly controlled. The primary difficulty with this trial design is that it may give information

only on short-term effectiveness. The randomized withdrawal trial (see section 2.1.5.2.4), however, which can also incorporate early-escape features, can give information on long-term effectiveness.

iii. Limited Placebo Period (2.1.5.2.3)

In a situation where long-term placebo treatment would not be acceptable, the use of a placebo group for a short period at the beginning of an active control trial could establish assay sensitivity (at least for short-term effects). The trial would then continue without the placebo group.

iv. Randomized Withdrawal (2.1.5.2.4)

In a randomized withdrawal trial, subjects receiving a test treatment for a specified time are randomly assigned to continued treatment with the test treatment or to placebo (i.e., withdrawal of active therapy). Subjects for such a trial could be derived from an organized open single-arm study, from an existing clinical cohort (but usually with a protocol-specified *wash-in* phase to establish the initial on-therapy baseline), from the active arm of a controlled trial, or from one or both arms of an active control trial. Any difference that emerges between the group receiving continued treatment and the group randomized to placebo would demonstrate the effect of the active treatment. The pre-randomization observation period on treatment can be of any length; this approach can therefore be used to study long-term persistence of effectiveness when long-term placebo treatment would not be acceptable. The post-withdrawal observation period could be of fixed duration or could use early escape or time to event (e.g., relapse of depression) approaches. As with the early-escape design, careful attention should be paid to procedures for monitoring patients and assessing study endpoints to ensure that patients failing on an assigned treatment are identified rapidly.

The randomized withdrawal approach is useful in several situations. First, it may be suitable for drugs that appear to resolve an episode of recurring illness (e.g., antidepressants), in which case the withdrawal study is, in effect, a relapse-prevention study. Second, it may be used for drugs that suppress a symptom or sign (chronic pain, hypertension, and angina), but where a long-term placebo-controlled trial would be difficult; in this case, the study can establish long-term efficacy. Third, the design is particularly useful in determining how long a therapy should be continued (e.g., post-infarction treatments with a beta-blocker).

The general advantage of randomized withdrawal designs, when used with an early-escape endpoint, such as return of symptoms, is that the period of placebo exposure with poor response that a patient would have to undergo is short.

This type of design can address dosing issues. After all patients have received an initial fixed dose, they could be randomly assigned in the withdrawal phase to several different doses (as well as placebo), a particularly useful approach when there is reason to think the initial and maintenance doses might be different, either on pharmacodynamic grounds or because there is substantial accumulation of active drug resulting from a long half life of parent drug or active metabolite. Note that the randomized withdrawal design could be used to assess dose-response

after an initial placebo-controlled titration study (See ICH E4). The titration study is an efficient design for establishing effectiveness, but does not give good dose-response information in many cases. The randomized withdrawal phase, with responders randomly assigned to several fixed doses and placebo, will permit dose-response to be studied rigorously while allowing the efficiency of the titration design to be used in the initial phase of the trial.

In using randomized withdrawal designs, it is important to appreciate the possibility of withdrawal phenomena, suggesting the wisdom of relatively slow tapering. A patient may develop tolerance to a drug such that no benefit is being accrued, but the drug's withdrawal may lead to disease exacerbation, resulting in an erroneous conclusion of persisting efficacy. It is also important to realize that treatment effects observed in these trials may be larger than those seen in an unselected population because randomized withdrawal studies are *enriched* with responders and exclude people who cannot tolerate the drug. This phenomenon results when the trial explicitly includes only subjects who appear to have responded to the drug or includes only people who have completed a previous phase of study (which is often an indicator of a good response and always indicates ability to tolerate the drug). In the case of studies intended to determine how long a therapy should be continued, such entry criteria provide the study population and comparison of interest.

v. Other Design Considerations (2.1.5.2.5)

In any placebo-controlled study, unbalanced randomization (e.g., 2:1, study drug to placebo) may enhance the safety database and may also make the study more attractive to patients and/or investigators.

6. Advantages of Placebo-controlled Trials (2.1.6)

a. Ability to Demonstrate Efficacy (2.1.6.1)

Like other superiority trials, a placebo-controlled trial contains internal evidence of assay sensitivity. When a difference is demonstrated it is interpretable without reference to external findings.

b. Measures *Absolute* Efficacy and Safety (2.1.6.2)

The placebo-controlled trial measures the total pharmacologically mediated effect of treatment. In contrast, an active control trial or a dose-comparison trial measures the effect relative to another treatment. The placebo-controlled trial also allows a distinction between adverse events due to the drug and those due to the underlying disease or *background noise*. The absolute effect size information is valuable in a three-group trial (test, placebo, active), even if the primary purpose of the trial is the test versus active control comparison.

c. Efficiency (2.1.6.3)

Placebo-controlled trials are efficient in that they can detect treatment effects with a smaller sample size than any other type of concurrently controlled study.

d. Minimizing the Effect of Subject and Investigator Expectations (2.1.6.4)

Use of a blinded placebo control may decrease the amount of improvement resulting from subject or investigator expectations because both are aware that some subjects will receive no active drug. This may increase the ability of the study to detect true drug effects.

7. Disadvantages of Placebo-controlled Trials (2.1.7)

a. Ethical Concerns (See Sections 2.1.3 and 2.1.4) (2.1.7.1)

When effective therapy that is known to prevent death or irreversible morbidity exists for a particular population, that population cannot usually be ethically studied in placebo-controlled trials; the particular conditions and populations for which this is true may be controversial. Ethical concerns may also direct studies toward less ill subjects or toward examination of short-term endpoints when long-term outcomes are of greater interest. Where a placebo-controlled trial is unethical and an active control trial would not be credible, it may be very difficult to study new drugs at all. For example, it would not be considered ethical to carry out a placebo-controlled trial of a thrombolytic agent in patients with acute myocardial infarction. Yet it would be difficult in the current environment to establish a valid non-inferiority margin based on historical data because of the emergence of acute revascularization procedures that might alter the size of the benefits of the thrombolytics. The designs described in section 2.1.5 may be useful in some of these cases.

b. Patient and Physician Practical Concerns (2.1.7.2)

Physicians and/or patients may be reluctant to accept the possibility that the patient will be assigned to the placebo treatment, even if there is general agreement that withholding or delaying treatment will not result in harm. Subjects who sense they are not improving may withdraw from treatment because they attribute lack of effect to having been treated with placebo, complicating the analysis of the study. With care, however, withdrawal for lack of effectiveness can sometimes be used as a study endpoint. Although this may provide some information on drug effectiveness, such information is less precise than actual information on clinical status in subjects receiving their assigned treatment.

c. Generalizability (2.1.7.3)

It is sometimes argued that any controlled trial, but especially a placebo-controlled trial, represents an artificial environment that gives results different from true *real world* effectiveness. If study populations are unrepresentative in placebo-controlled trials because of ethical or practical concerns, questions about the generalizability of study results can arise. For example, protocol, investigator, or patient choice from placebo-controlled trials may exclude patients with more serious disease. In some cases, only a limited number of patients or centers may be willing

to participate in studies. Whether these concerns actually (as opposed to theoretically) limit generalizability has not been established.

d. No Comparative Information (2.1.7.4)

Placebo-controlled trials lacking an active control give little useful information about comparative effectiveness, information that is of interest and importance in many circumstances. Such information cannot reliably be obtained from cross-study comparisons, as the conditions of the studies may have been quite different.

B. No-treatment Concurrent Control (See Section 1.3.2) (2.2)

The randomized no-treatment control is similar in its general properties and its advantages and disadvantages to the placebo-controlled trial. Unlike the placebo-controlled trial, however, it cannot be fully blinded, and this can affect all aspects of the trial, including subject retention, patient management, and all aspects of observation (see section 1.2.2). This design is appropriate in circumstances where a placebo-controlled trial would be performed, except that blinding is not feasible or practical. When this design is used, it is desirable to have critical decisions, such as eligibility and endpoint determination or changes in management, made by an observer blinded to treatment assignment. Decisions related to data analysis, such as inclusion of patients in analysis sets, should also be made by individuals without access to treatment assignment. See ICH E9 for further discussion.

C. Dose-response Concurrent Control (See Section 1.3.3) (2.3)

1. Description (2.3.1)

A dose-response study is one in which subjects are randomly assigned to two or more dosage groups, with or without a placebo group. Dose-response studies are carried out to establish the relation between dose and efficacy and adverse effects and/or to demonstrate efficacy. The first use is considered in ICH E4; the use to demonstrate efficacy is the subject of this guidance. Evidence of efficacy could be based on significant differences in pair-wise comparisons between dosage groups or between dosage groups and placebo, or on evidence of a significant positive trend with increasing dose, even if no two groups are significantly different. In the latter case, however, further study may be needed to assess the effectiveness of the low doses. As noted in ICH E9, the particular approach for the primary efficacy analysis should be prespecified.

Studies in which the treatment groups vary in regimen raise many of the same considerations as dose-response trials. Since the use of regimen-controlled trials to establish efficacy is uncommon, the current discussion is focussed on dose-response trials.

There are several advantages to inclusion of a placebo (zero-dose) group in a dose-response study. First, it avoids studies that are uninterpretable because all doses produce similar effects so that one cannot assess whether all doses are equally effective or equally ineffective. Second, the placebo group permits an estimate of the total pharmacologically mediated effect of treatment,

although the estimate may not be very precise if the dosing groups are relatively small. Third, as the drug-placebo difference is generally larger than inter-dose differences, use of placebo may permit smaller sample sizes. The size of various dose groups need not be identical; e.g., larger samples could be used to give more precise information about the effect of smaller doses or be used to increase the power of the study to show a clear effect of what is expected to be the optimal dose. Dose-response studies can include one or more doses of an active control treatment. Randomized withdrawal designs can also assign subjects to multiple dosage levels.

2. *Ability to Minimize Bias (2.3.2)*

If the dose-response study is blinded, it shares with other randomized and blinded designs an ability to minimize subject and investigator bias. When a drug has pharmacologic effects that could break the blind for some patients or investigators, it may be easier to preserve blinding in a dose-response study than in a placebo-controlled trial. Masking treatments may necessitate multiple dummies or preparation of several different doses that look alike.

3. *Ethical Issues (2.3.3)*

The ethical and practical concerns related to a dose-response study are similar to those affecting placebo-controlled trials. Where there is therapy known to be effective in preventing death or irreversible morbidity, it is no more ethically acceptable to randomize deliberately to subeffective control therapy than it is to randomize to placebo. Where therapy is directed at less serious conditions or where the toxicity of the therapy is substantial relative to its benefits, dose-response studies that use lower, potentially less effective and less toxic doses or placebo may be acceptable to patients and investigators.

4. *Usefulness of Dose-response Studies and Validity of Inference in Particular Situations (2.3.4)*

In general, a blinded dose-response study is useful for the determination of efficacy and safety in situations where a placebo-controlled trial would be useful and has similar credibility (see section 2.1.4).

5. *Modifications of Design and Combinations with Other Controls That Can Resolve Ethical, Practical, or Inferential Problems (2.3.5)*

In general, the sorts of modification made to placebo-controlled studies to mitigate ethical, practical, or inferential problems are also applicable to dose-response studies (see section 2.1.5).

6. *Advantages of Dose-response Trials (2.3.6)*

a. Efficiency (2.3.6.1)

Although a comparison of a large, fully effective dose to placebo may be maximally efficient for showing efficacy, this design may produce unacceptable toxicity and gives no dose-response

information. When the dose-response is monotonic, the dose-response trial is reasonably efficient in showing efficacy and also yields dose-response information. If the optimally effective dose is not known, it may be more prudent to study a range of doses than to choose a single dose that may prove to be suboptimal or to have unacceptable adverse effects.

b. Possible Ethical Advantage (2.3.6.2)

In some cases, notably those in which there is likely to be dose-related efficacy and dose-related important toxicity, the dose-response study may represent a difference-showing trial that can be ethically or practically conducted, even where a placebo-controlled trial could not be, because there is reason for patients and investigators to accept lesser effectiveness in return for greater safety.

7. Disadvantages of Dose-response Study (2.3.7)

A potential problem that should be recognized is that a positive dose-response trend (i.e., a significant correlation between the dose and the efficacy outcome), without significant pair-wise differences, can establish efficacy (see 2.3.1), but may leave uncertainty as to which doses (other than the largest) are actually effective. Of course, a single-dose study poses a similar problem with respect to doses below the one studied, giving no information at all about such doses.

It should also be appreciated that it is not uncommon to show no difference between doses in a dose-response study; if there is no placebo group this is usually an uninformative outcome.

If the therapeutic range is not known at all, the design may be inefficient, as many patients may be assigned to sub-therapeutic or supratherapeutic doses.

Dose-response designs may be less efficient than placebo-controlled titration designs for showing the presence of a drug effect; they do, however, in most cases provide better dose-response information (see ICH E4).

D. Active Control (See Section 1.3.4) (2.4)

1. Description (2.4.1)

An active control (positive control) trial is one in which an investigational drug is compared with a known active drug. Such trials are randomized and usually double blind. The most crucial design question is whether the trial is intended to show a difference between the two treatments or to show non-inferiority or equivalence. A sponsor intending to demonstrate effectiveness by means of a trial showing non-inferiority of the test drug to a standard agent needs to address the issue of the assay sensitivity of the trial, as discussed in section 1.5. In a non-inferiority or equivalence trial, the active control treatment needs to be of established efficacy at the dose used and under the conditions of the study (see ICH E9: Statistical Principles for Clinical Trials). In general, this means it should be a drug acceptable in the region to which the studies will be submitted for the same indication at the dose being studied. A superiority study favoring the test

drug, on the other hand, is readily interpretable as evidence of efficacy, even if the dose of active control is too low or the active control is of uncertain benefit (but not if it could be harmful). Such a result, however--superiority in the trial of the test agent to the control--is interpretable as true superiority of the test treatment to the control treatment only when the active control is used in appropriate patients at an appropriate dose and schedule (see section 1.4.3). Lack of appropriate use of the control treatment would also make the study unusable as a non-inferiority study, if superiority of the test drug is not shown, because assay sensitivity of the study would not be ensured (see section 1.5.2).

2. Ability to Minimize Bias (2.4.2)

A randomized and blinded active control trial generally minimizes subject and investigator bias, but a note of caution is warranted. In a non-inferiority trial, investigators and subjects know that all subjects are getting active drug, although they do not know which one. This could lead to a tendency toward categorizing borderline cases as successes in partially subjective evaluations, e.g., in an antidepressant study, which could decrease observed treatment differences and increase the likelihood that a finding of non-inferiority would not represent evidence of effectiveness.

3. Ethical Issues (2.4.3)

Active control trials are generally considered to pose fewer ethical and practical problems than placebo-controlled trials because all subjects receive active treatment. It should be appreciated, however, that subjects receiving a new treatment are not receiving standard therapy (just as a placebo control group is not) and may be receiving an ineffective or harmful drug. This is an important matter if the active control therapy is known to improve survival or decrease the occurrence of irreversible morbidity, i.e. conditions in which a placebo or no treatment control would be unacceptable. There should therefore be a sound rationale for the test treatment. If there is no strong reason to expect the new drug to be at least as good as the standard, an add-on study (see section 2.1.5.2.1) may be more appropriate, if the conditions allow such a design.

4. Usefulness of Active Control Trials; Validity of Inference in Particular Situations (2.4.4)

When a new treatment shows an advantage over an active control, the study is readily interpreted as showing efficacy, just as any other superiority trial is, assuming that the active control is not actually harmful. When an active control trial is used to show efficacy by demonstrating non-inferiority, there is the special consideration of assay sensitivity, which is considered above in section 1.5. The active control trial can also be used to assess comparative efficacy if assay sensitivity is established.

5. *Modifications of Design and Combinations with Other Controls That Can Resolve Ethical, Practical, or Inferential Issues (2.4.5)*

As discussed earlier (section 2.1.5), active control trials can include a placebo group, multiple-dose groups of the test drug, and/or other dose groups of the active control. Comparative dose-response studies, in which there are several doses of both test and active control, are typical in analgesic trials. The doses in active control trials can be fixed or titrated, and both cross-over and parallel designs can be used. The assay sensitivity of a non-inferiority trial can sometimes be supported by a randomized placebo-controlled withdrawal phase at the end (see section 2.1.5.2.4) or by an initial short period of comparison to placebo (see section 2.1.5.2.3). Active control superiority studies in selected populations (nonresponders to other therapy or to the active control) can be very useful and are generally easy to interpret, although the results may not be generalizable.

6. *Advantages of Active Control Trials (2.4.6)*

a. Ethical and Practical Advantages (2.4.6.1)

The active control design, whether intended to show non-inferiority or equivalence or superiority, reduces ethical concerns that arise from failure to use drugs with documented important health benefits. It also addresses patient and physician concerns about failure to use documented effective therapy. Recruitment and IRB/IEC approval may be facilitated, and it may be possible to study larger samples. There may be fewer withdrawals due to lack of effectiveness.

b. Information Content (2.4.6.2)

Where superiority to an active treatment is shown, active control studies are readily interpretable regarding evidence of efficacy. The larger sample sizes needed are sometimes more achievable and acceptable in active control trials and can provide more safety information. Active control trials also can, if properly designed, provide information about relative efficacy.

7. *Disadvantages of Active Control Trials (2.4.7)*

a. Information Content (2.4.7.1)

See section 1.5 for discussion of the problem of assay sensitivity and the ability of the trial to support an efficacy conclusion in non-inferiority or equivalence trials. Even when assay sensitivity is supported and the study is suitable for detecting efficacy, there is no direct assessment of effect size, and there is also greater difficulty in quantitating safety outcomes.

b. Large Sample Size (2.4.7.2)

Generally, the non-inferiority margin to be excluded is chosen conservatively in order to be reasonably sure that the margin is not greater than the smallest effect size that the active control would reliably be expected to have. In addition, because there will usually be an intent to rule

out loss of more than some reasonable fraction (see section 1.5.1) of the control drug effect, a still smaller non-inferiority margin is often used. Because the choice of the margin will therefore be conservative, sample sizes may be very large. In an active control superiority trial, the expected difference between two drugs is always smaller than the expected difference between test drug and placebo, again leading to large sample sizes.

E. External Control (Including Historical Control, See Section 1.3.5) (2.5)

1. Description (2.5.1)

An externally controlled trial is one in which the control group consists of patients who are not part of the same randomized study as the group receiving the investigational agent; i.e., there is no concurrently randomized control group. The control group is thus not derived from exactly the same population as the treated population. Usually, the control group is a well-documented population of patients observed at an earlier time (historical control), but it could be a group at another institution observed contemporaneously, or even a group at the same institution but outside the study. An external control study could be a superiority study (e.g. comparison with an untreated group) or a non-inferiority study. Sometimes certain patients from a larger external experience are selected as a control group on the basis of particular characteristics that make them similar to the treatment group; there may even be an attempt to match particular control and treated patients.

In so-called baseline-controlled studies, the patient's state over time is compared with their baseline state. Although these studies are sometimes thought to use *the patient as his own control*, they do not in fact have an internal control. Rather, changes from baseline are compared with an estimate of what would have happened to the patients in the absence of treatment with the test drug. Both baseline-controlled trials and trials that use a more complicated sequential on-off-on (drug, placebo, drug) design, but that do not include a concurrently randomized control group, are of this type. As noted, in these trials the observed changes from baseline or between study periods are always compared, at least implicitly, to some estimate of what would have happened without the intervention. Such estimates are generally made on the basis of general knowledge, without reference to a specific control population. Although in some cases this is plainly reasonable, e.g., when the effect is dramatic, occurs rapidly following treatment, and is unlikely to have occurred spontaneously (e.g., general anesthesia, cardioversion, measurable tumor shrinkage), in most cases it is not so obvious and a specific historical experience should be sought. Designers and analysts of such trials need to be aware of the limitations of this type of study and should be prepared to justify its use.

2. Ability to Minimize Bias (2.5.2)

Inability to control bias is the major and well-recognized limitation of externally controlled trials and is sufficient in many cases to make the design unsuitable. It is always difficult, and in many cases impossible, to establish comparability of the treatment and control groups and thus to fulfill the major purpose of a control group (see section 1.2). The groups can be dissimilar with respect to a wide range of factors, other than use of the study treatment, that could affect outcome,

including demographic characteristics, diagnostic criteria, stage or severity of disease, concomitant treatments, and observational conditions (such as methods of assessing outcome, investigator expectations). Such dissimilarities can include important but unrecognized prognostic factors that have not been measured. Blinding and randomization are not available to minimize bias when external controls are used. It is well documented that untreated historical-control groups tend to have worse outcomes than an apparently similarly chosen control group in a randomized study, possibly reflecting a selection bias. Control groups in a randomized study need to meet certain criteria to be entered into the study, criteria that are generally more stringent and identify a less sick population than is typical of external control groups. An external control group is often identified retrospectively, leading to potential bias in its selection. A consequence of the recognized inability to control bias is that the potential persuasiveness of findings from externally controlled trials depends on obtaining much more extreme levels of statistical significance and much larger estimated differences between treatments than would be considered necessary in concurrently controlled trials.

The inability to control bias restricts use of the external control design to situations in which the effect of treatment is dramatic and the usual course of the disease highly predictable. In addition, use of external controls should be limited to cases in which the endpoints are objective and the impact of baseline and treatment variables on the endpoint is well characterized.

As noted, the lack of randomization and blinding, and the resultant problems with lack of assurance of comparability of test group and control group, make the possibility of substantial bias inherent in this design and impossible to quantitate. Nonetheless, some approaches to design and conduct of externally controlled trials could lead them to be more persuasive and potentially less biased. A control group should be chosen for which there is detailed information, including, where pertinent, individual patient data regarding demographics, baseline status, concomitant therapy, and course on study. The control patients should be as similar as possible to the population expected to receive the test drug in the study and should have been treated in a similar setting and in a similar manner, except with respect to the study therapy. Study observations should use timing and methodology similar to those used in the control patients. To reduce selection bias, selection of the control group should be made before performing comparative analyses; this may not always be feasible, as outcomes from these control groups may have been published. Any matching on selection criteria or adjustments made to account for population differences should be specified prior to selection of the control and performance of the study. Where no obvious single optimal external control exists, it may be advisable to study multiple external controls, providing that the analytic plan specifies conservatively how each will be used in drawing inferences (e.g., study group should be substantially superior to the most favorable control to conclude efficacy). In some cases, it may be useful to have an independent set of reviewers reassess endpoints in the control group and in the test group in a blinded manner according to common criteria.

3. Ethical Issues (2.5.3)

When a drug is intended to treat a serious illness for which there is no satisfactory treatment, especially if the new drug is seen as promising on the basis of theoretical considerations, animal

data, or early human experience, there may be understandable reluctance to perform a comparative study with a concurrent control group of patients who would not receive the new treatment. At the same time, it is not responsible or ethical to carry out studies that have no realistic chance of credibly showing the efficacy of the treatment. It should be appreciated that many promising therapies have had less dramatic effects than expected or have shown no efficacy at all when tested in controlled trials. Investigators may, in these situations, be faced with very difficult judgments. It may be tempting in exceptional cases to initiate an externally controlled trial, hoping for a convincingly dramatic effect, with a prompt switch to randomized trials if this does not materialize.

Alternatively, and generally preferably, in dealing with serious illnesses for which there is no satisfactory treatment, but where the course of the disease cannot be reliably predicted, even the earliest studies should be randomized. This is usually possible when studies are carried out before there is an impression that the therapy is effective. Studies can be monitored by independent data monitoring committees so that dramatic benefit can be detected early. The concurrently controlled trial can detect extreme effects very rapidly and, in addition, can detect modest, but still valuable, effects that would not be credibly demonstrated by an externally controlled trial.

4. Usefulness of Externally Controlled Trials; Validity of Inference in Particular Situations (2.5.4)

An externally controlled trial should generally be considered only when prior belief in the superiority of the test therapy to all available alternatives is so strong that alternative designs appear unacceptable and the disease or condition to be treated has a well-documented, highly predictable course. It is often possible, even in these cases, to use alternative, randomized, concurrently controlled designs (see section 2.1.5).

Externally controlled trials are most likely to be persuasive when the study endpoint is objective, when the outcome on treatment is markedly different from that of the external control and a high level of statistical significance for the treatment-control comparison is attained, when the covariates influencing outcome of the disease are well characterized, and when the control closely resembles the study group in all known relevant baseline, treatment (other than study drug), and observational variables. Even in such cases, however, there are documented examples of erroneous conclusions arising from such trials.

When an external control trial is considered, appropriate attention to design and conduct may help reduce bias (see section 2.5.2).

5. Modifications of Design and Combinations with Other Controls That Can Resolve Ethical, Practical or Inferential Problems (2.5.5)

The external control design can incorporate elements of randomization and blinding through use of a randomized, placebo controlled withdrawal phase, often with early escape provisions, as described earlier (see section 2.1.5.2.4). The results of the initial period of treatment, in which

subjects who appear to respond are identified and maintained on therapy, are thus *validated* by a rigorous, largely assumption- and bias-free study.

6. Advantages of Externally Controlled Trials (2.5.6)

The main advantage of an externally controlled trial is that all patients can receive a promising drug, making the study more attractive to patients and physicians.
The design has some potential efficiencies because all patients are exposed to test drug, of particular importance in rare diseases. However, despite the use of a single treatment group in an externally controlled trial, the estimate of the external control group outcome always should be made conservatively, possibly leading to a larger sample size than would be needed in a placebo-controlled trial. Great caution (e.g., applying a more stringent significance level) is called for because there are likely to be both identified and unidentified or unmeasurable differences between the treatment and control groups, often favoring treatment.

7. Disadvantages of Externally Controlled Trials (2.5.7)

The externally controlled study cannot be blinded and is subject to patient, observer, and analyst bias; these are major disadvantages. It is possible to mitigate these problems to a degree, but even the steps suggested in section 2.5.2 cannot resolve such problems fully, as treatment assignment is not randomized and comparability of control and treatment groups at the start of treatment, and comparability of treatment of patients during the trial, cannot be ensured or well assessed. It is well documented that externally controlled trials tend to overestimate efficacy of test therapies. It should be recognized that tests of statistical significance carried out in such studies are less reliable than in randomized trials.

III. CHOOSING THE CONCURRENT CONTROL GROUP (3.0)

Table 1 describes the usefulness of specific types of control groups, and Figure 1 provides a decision tree for choosing among different types of control groups. Although the table and figure focus on the choice of control to demonstrate efficacy, some designs also allow comparisons of test and control agents. The choice of control can be affected by the availability of therapies and by medical practices in specific regions.

The potential usefulness of the principal types of control (placebo, active, and dose-response) in specific situations and for specific purposes is shown in Table 1. The table should be used with the text describing the details of specific circumstances in which potential usefulness can be realized. In all cases, it is presumed that studies are appropriately designed. External controls are a case so distinct that they are not included in the table.

In most cases, evidence of efficacy is most convincingly demonstrated by showing superiority to a concurrent control treatment. If a superiority trial is not feasible or is inappropriate for ethical or practical reasons, and if a defined treatment effect of the active control is regularly seen (e.g.,

as it is for antibiotics in most situations), a non-inferiority or equivalence trial can be used and can be persuasive.

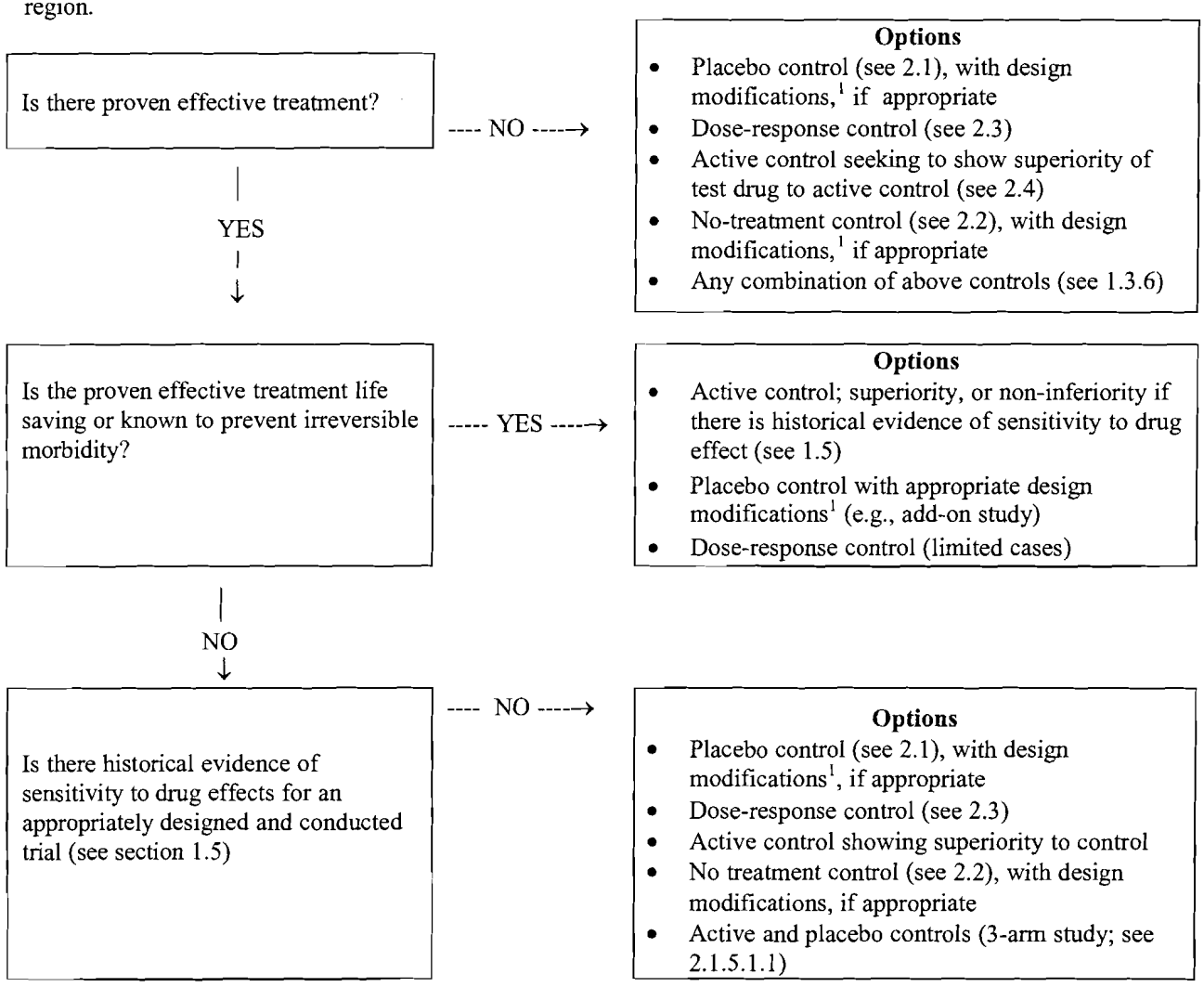
Table 1. Usefulness of Specific Concurrent Control Types in Various Situations

	Type of Control							
Trial Objective	Placebo	Active non-inferiority	Active Superiority	Dose Response (D/R)	Placebo + Active	Placebo + D/R	Active + D/R	Placebo + Active + D/R
Measure *Absolute* effect size	Y	N	N	N	Y	Y	N	Y
Show existence of effect	Y	P	Y	Y	Y	Y	Y	Y
Show Dose-Response relationship	N	N	N	Y	N	Y	Y	Y
Compare therapies	N	P	Y	N	Y	N	P	Y

Y=Yes, N=No, P=Possible, depending on whether there is historical evidence of sensitivity to drug effects

Figure 1: Choosing the Concurrent Control for Demonstrating Efficacy

This figure shows the basic logic for choosing the control group; the decision may depend on the available drugs or medical practices in the specific region.



↓
YES
----------→

Options

- Placebo control (see 2.1), with design modifications, if appropriate
- Dose-response control
- Active control showing superiority to control
- Active and placebo controls (3-arm study; see 2.1.5.1.1)
- Active control non-inferiority (see 1.5)

[1] Add-on, replacement, early escape, brief placebo period, and randomized withdrawal (see section 2.1.5.2).