

Placebo-Controlled Trials and Active-Control Trials in the Evaluation of New Treatments

Part 1: Ethical and Scientific Issues

Robert Temple, MD, and Susan S. Ellenberg, PhD

In recent years, several authors have argued that placebo-controlled trials are invariably unethical when known effective therapy is available for the condition being studied, regardless of the condition or the consequences of deferring treatment. Some have also disputed the value of placebo-controlled trials in such a setting, asserting that the comparison of new treatment with old treatment is sufficient to establish efficacy and is all that should be of interest. This article considers the ethical concerns about use of placebo controls and describes the limited ability of active-control equivalence (also known as noninferiority) trials to establish efficacy of new therapies in many medical contexts. The authors conclude that placebo-controlled trials are not uniformly

unethical when known effective therapies are available; rather, their acceptability is determined by whether the patient will be harmed by deferral of therapy. If patients are not harmed, such trials can ethically be carried out. Furthermore, active-control trials, although valuable, informative, and appropriate in many circumstances, often cannot provide reliable evidence of the effectiveness of a new therapy.

Ann Intern Med. 2000;133:455-463.

For author affiliations and current addresses, see end of text.

See editorial comment on pp 474-475.

Placebo-controlled trials are used extensively in the development of new pharmaceuticals. They are sometimes challenged as unethical in settings in which patients could be treated with an existing therapy (1–7). The issues of when placebo controls are ethically acceptable and when they are scientifically necessary are important and worthy of discussion.

The Ethics of Placebo Controls

The Declaration of Helsinki

The Declaration of Helsinki (8) is an international document that describes ethical principles for clinical investigation. Those who contend that placebo controls are unethical whenever known effective therapy exists for a condition usually cite the following sentence in the Declaration as support for that position: “In any medical study, every patient—including those of a control group, if any—should be assured of the best proven diagnostic and therapeutic method.”

We believe that an interpretation of this sentence as barring placebo controls whenever an effective treatment exists is untenable. First, the requirement that all patients receive the “best proven diagnostic and therapeutic method” would bar not only placebo-controlled trials but also active-control and historically controlled trials. When effective treatment exists, the patient receiving the investi-

gational treatment instead of the established therapy is clearly not getting the best proven treatment.

Second, it does not seem reasonable to consider as equivalent all failures to use known effective therapy. Historically, concerns about placebo use have usually arisen in the context of serious illness. There is universal agreement that use of placebo or otherwise untreated controls is almost always unethical when therapy shown to improve survival or decrease serious morbidity is available. But in cases in which the treatment does not affect the patient’s long-term health, an ethical imperative to use existing therapy is not plausible. Can it be, for example, that because topical minoxidil or oral finasteride can grow hair, a placebo-controlled trial of a new remedy for baldness is unethical? Is it really unethical to use placebos in short-term studies of drugs for allergic rhinitis, insomnia, anxiety, dermatoses, heartburn, or headaches in fully informed patients? We do not believe that there is a reasonable basis for arguing that such studies and many other placebo-controlled studies of symptom relief are unethical and that an informed patient cannot properly be asked to participate in them.

Third, there is good reason to doubt that the cited phrase was intended to discourage placebo-controlled trials. The phrase under discussion was not part of the original 1964 Declaration but was added in 1975 to reinforce the idea that the physician–patient relationship “must be respected just as it would be in a purely therapeutic situa-

tion not involving research objectives” (8). In the explanation accompanying the 1975 change, the issue of placebo-controlled trials was not even mentioned (9). The American Medical Association (10), the World Health Organization (11), and the Council for International Organizations of Medical Sciences (12) have rejected the position that the Declaration uniformly bars placebo-controlled trials when proven therapy is available.

Informed Consent in Placebo-Controlled Trials

Patients asked to participate in a placebo-controlled trial must be informed of the existence of any effective therapy, must be able to explore the consequences of deferring such therapy with the investigator, and must provide fully informed consent. Concern about whether consent to participate in trials is as informed as we would like to believe is valid, but these concerns apply as much to the patient’s decision to forgo known effective treatment and risk exposure to a potentially ineffective or even harmful new agent in an active-control trial as to a decision to accept possible persistence of symptoms in a placebo-controlled trial. Thus, this problem is not unique to placebo-controlled trials.

For the above reasons, we conclude that placebo-controlled trials may be ethically conducted even when effective therapy exists, as long as patients will not be harmed by participation and are fully informed about their alternatives. Although in many cases application of this standard will be fairly straightforward, in others it will not, and there may be debate about the consequences of deferring treatment (13).

Assessment of Effectiveness with Active-Control Trials

Clinical trials that, because of deficiencies in study design or conduct, are unlikely to provide scientifically valid and clinically meaningful results raise their own ethical concerns (12, 14). The remainder of this paper will address the inability of commonly proposed alternatives to placebo-controlled trials to evaluate the effectiveness of new treatments in many medical settings.

Active-Control Equivalence Trials (Noninferiority Trials)

The ability to conduct a placebo-controlled trial ethically in a given situation does not necessarily mean that

placebo-controlled trials should be carried out when effective therapy exists. Patients and physicians might still prefer a trial in which every participant is given an active treatment. What remains to be examined is why placebo-controlled trials (or, more generally, trials intended to show an advantage of one treatment over another) are frequently needed to demonstrate the effectiveness of new treatments and often cannot be replaced by active-control trials showing that a new drug is equivalent or noninferior to a known effective agent. The limitations of active-control equivalence trials (ACETs) that are intended to show the effectiveness of a new drug have long been recognized and are well described (15–33) but are perhaps not as widely appreciated as they should be. A recent proposed international guideline on choice of control group addresses this issue in detail (33).

The Fundamental Problem: Need for Assay Sensitivity

There are two distinct ways to show that a new therapy is effective. One can show that the new therapy is superior to a control treatment, or one can show that the new therapy is equivalent to or not worse by some defined amount than a known effective treatment. Each method can be valid, but each requires entirely different inferential approaches. A well-designed study that shows superiority of a treatment to a control (placebo or active therapy) provides strong evidence of the effectiveness of the new treatment, limited only by the statistical uncertainty of the result. No information external to the trial is needed to support the conclusion of effectiveness. In contrast, a study that successfully shows “equivalence”—that is, little difference between a new drug and known active treatment—does not by itself demonstrate that the new treatment is effective. “Equivalence” could mean that the treatments were both effective in the study, but it could also mean that both treatments were ineffective in the study. To conclude from an ACET that a new treatment is effective on the basis of its similarity to the active control, one must make the critical (and untestable within the study) assumption that the active control had an effect in that particular study. In other words, one must assume that if a placebo group had been included, the placebo would have been inferior to the active control (15–33). Support for this assumption must come from sources external to the trial. Although it might appear reasonable to expect a known

Table. Results of Six Trials Comparing Nomifensine, Imipramine, and Placebo*

Study	Common Baseline Score on the Hamilton Depression Scale	Four-Week Adjusted Score on the Hamilton Depression Scale (Number of Participants)			P Value†
		Nomifensine	Imipramine	Placebo	
R301	23.9	13.4 (33)	12.8 (33)	14.8 (36)	0.78
G305	26.0	13.0 (39)	13.4 (30)	13.9 (36)	0.86
C311(1)	28.1	19.4 (11)	20.3 (11)	18.9 (13)	0.81
V311(2)	29.6	7.3 (7)	9.5 (8)	23.5 (7)	0.63‡
F313	37.6	21.9 (7)	21.9 (8)	22.0 (8)	1.0
K317	26.1	11.2 (37)	10.8 (32)	10.5 (36)	0.85

* Results shown are those reported in the review by Leber (18).

† Two-tailed P value for nomifensine versus imipramine.

‡ $P < 0.001$ for nomifensine versus placebo and imipramine versus placebo.

active agent to be superior to placebo in any given appropriately designed trial, experience has shown that this is not the case for many types of drugs.

The ability of a study to distinguish between active and inactive treatments is termed *assay sensitivity*. If assay sensitivity cannot be assumed, then even if the new and standard treatments appear virtually identical and the confidence interval for their comparison is exquisitely narrow, the study cannot demonstrate effectiveness of the new drug. (Note that in practice, ACETs are not designed simply to show lack of a statistically significant difference between treatments. Rather, such trials are designed to show noninferiority—that the new treatment is not inferior to the control by more than a specified margin. This approach is described in the Appendix.)

The best evidence that an active drug would have an effect superior to that of placebo in a given study would be a series of trials of similar design in which the active drug has reliably outperformed placebo. The ACET thus requires information external to the trial (the information about past placebo-controlled studies of the active control) to interpret the results. In this respect, an ACET is similar to a historically controlled trial. In some settings, such as highly responsive cancers, most infectious diseases, and some cardiovascular conditions, such external information is available and ACETs can and do provide a valid and reliable basis for evaluating new treatments. In many cases, however, the historically based assumption of assay sensitivity cannot be made; for many types of effective drugs, studies of apparently adequate size and design do not regularly distinguish drugs from placebo (16–18, 25, 34). More than 20 years ago, Lasagna (19) described this difficulty particularly well (reflecting long recognition of the problem among analgesiologists):

... a comparison between new drug and standard ... is convincing only when the new remedy is superior to standard treatment. If it is inferior, or even indistinguishable from a standard remedy, the results are not readily interpretable. In the absence of placebo controls, one does not know if the “inferior” new medicine has any efficacy at all, and “equivalent” performance may reflect simply a patient population that cannot distinguish between two active treatments that differ considerably from each other, or between active drug and placebo. Certain clinical conditions, such as serious depressive states, are notoriously difficult to evaluate because of the delay in drug effects and the high rate of spontaneous improvement, and even known remedies are not readily distinguished from placebo in controlled trials.

The problem is well recognized in studies of antidepressant drugs (18, 32). In practice, many such studies include three arms—new drug, active control, and placebo—to provide clear evidence of effectiveness (new drug vs. placebo) and an internal standard (active control vs. placebo). This design allows a clear distinction (particularly valuable to a drug manufacturer) between a *drug* that does not work (the standard agent is superior to placebo but the new drug is not) and a *study* that does not work (neither the standard drug nor the new drug is superior to placebo).

The assay sensitivity problem was illustrated by Leber (18), who examined the results of all three-arm studies comparing nomifensine (an effective but toxic antidepressant), imipramine (a standard tricyclic antidepressant shown to be superior to placebo in dozens of clinical trials), and placebo. The results of the studies are shown in the Table. No study found a difference between nomifensine

and imipramine on the Hamilton depression scale (a standard measure of depression), but the changes from baseline with both drugs were substantial and seemed clinically meaningful. Examination of the placebo results, however, shows similar changes. Only one of the six studies—the smallest one—found any significant difference between either of the two active drugs and placebo. None of the other five studies showed even a trend favoring either drug. These five studies appear to have lacked assay sensitivity; they could not distinguish active from inactive treatments. Although some of these studies were small, three studies with 30 or more patients per group were typical of studies that often did show effectiveness of imipramine or other antidepressants.

Although we cannot be certain of the reason for these outcomes, the most likely explanation is that differences in study samples, study designs, or study conduct affected the response to these antidepressants and thus the ability of the studies to identify effective therapy. It does not seem to be merely a matter of study size, however; many studies with 10 to 30 patients per group (including one of the six shown) detect effects of antidepressant drug effects, and many much larger studies of essentially the same design do not show even a favorable trend. Similar patterns, although not as extreme, have been seen with many recently developed antidepressants, such as fluoxetine (34). Overall, in recent experience at the U.S. Food and Drug Administration, about one third to one half of modern antidepressant trials do not distinguish a known effective drug from placebo (Laughren T. Unpublished observations).

One might speculate that variable results of trials of antidepressants are simply the consequence of modest effect sizes coupled with samples too small to overcome the inherent variability of the condition studied. Results, however, are consistent with effect sizes that vary greatly and unpredictably from study to study. With current knowledge, one cannot specify a particular study population, treatment protocol, or sample size that will regularly identify active agents.

Antidepressants are only one of many classes of drugs with assay sensitivity problems. Analgesics (35), anxiolytics, antihypertensives, hypnotics, antianginal agents, angiotensin-converting enzyme inhibitors for heart failure, postinfarction β -blockers (36), antihistamines, nonsteroidal asthma prophylaxis, motility-modifying drugs for gastroesophageal reflux disease, and many other effective

agents are often indistinguishable from placebo in well-designed and -conducted trials.

A recently published overview by Tramer and coworkers (37) of studies of ondansetron, a widely used and very effective antiemetic, provides a further example of this phenomenon. Although the totality of data clearly supports the efficacy of this agent, many placebo–ondansetron comparisons show no effect of the drug. It is notable that the incidence of nausea and vomiting varied greatly among the trials and in some cases was so low that it precluded any demonstration of efficacy. In a placebo-controlled study of an antiemetic, a low rate of nausea and vomiting in the placebo group would lead to a negative outcome—the drug could not appear superior to placebo, and the trial could not provide evidence of effectiveness. In contrast, an ACET (new drug vs. ondansetron) with a low rate of nausea and vomiting in both arms would not be unambiguously interpretable. If one assumed that the low rate in the active-control group reflected the known ability of ondansetron to reduce a rate of nausea and vomiting that would have been high in the absence of treatment, one would conclude that the new drug was also effective. But the article by Tramer and coworkers shows that such an assumption cannot be supported in many situations. Clearly, if many placebo-controlled studies of ondansetron showed no effect, a trial showing “equivalence” of a new agent to ondansetron could not be considered reliable evidence that the new agent was effective, unless one could identify a treatment setting (for example, a setting defined by the chemotherapy administered) in which ondansetron was regularly distinguishable from placebo.

In the cases described, the effectiveness of drugs that sometimes (or even often) fail to be proven superior to placebo is not in doubt; even if a drug is statistically significantly superior to placebo in only 50% of well-designed and well-conducted studies, that proportion is still vastly greater than the small fraction that would be expected to occur by chance if the drugs were ineffective. The problem may be a generally small response that varies among populations, insufficient adherence to therapy or use of concomitant medication, study samples that improve spontaneously (leaving no room for drug-induced improvement) or that are unresponsive to the drug, or some other reason not yet recognized. What all of these influences have in common is that they reduce or eliminate the drug–placebo difference, so that a study design and size adequate to de-

tect a larger effect will not detect the reduced effect. In each case, however, the problem is not identifiable a priori by examining the study; it is recognized only by the observed failure of the trial to distinguish the drug and placebo treatments.

Incentive To Minimize Errors Is Reduced in ACETs

Active-control equivalence trials present another problem that is difficult to quantitate or assess in any given study. Most imperfections in a clinical trial—patient non-adherence to treatment, use of concomitant therapy potentially affecting study outcome, inclusion of inappropriate patients (for example, those who lack the disease or those who experience spontaneous improvement), or administering the wrong treatments—tend to reduce observable differences between treatment groups, promoting the conclusion that the two treatments are indistinguishable. Study organizers seeking to demonstrate a difference between treatments have a powerful incentive to minimize such imperfections and to identify a population in which an effect could be demonstrated. This incentive is absent when the intent is to demonstrate *lack* of difference (17, 32). This is not to suggest that trial organizers deliberately make less of an effort to maintain study quality in ACETs than in placebo-controlled trials, any more than the practice of blinding investigators to treatment suggests that investigators are not to be trusted. It is important, however, to recognize the possible influence of the desired outcome on the conduct of clinical trials.

It is difficult in any given ACET to determine the extent to which the ability to show potential treatment differences has been diminished by deficiencies in study design and conduct. In such areas as treatment of depression, however, even placebo-controlled trials, in which the incentive to conduct an excellent study capable of showing a difference between treatments is maximal, often cannot distinguish effects of active drugs from those of placebo. Results of ACETs would be expected to be at least as variable as those of placebo-controlled trials in their ability to detect treatment differences. In considering how to conduct ACETs, this issue needs to be recognized. In addition, approaches to study interpretation usually thought of as conservative, such as intention-to-treat analyses, are no longer conservative when the objective of a trial is to show no difference between treatments (17, 24, 32).

Use of Active Controls

Active-control equivalence trials can be informative and have been used successfully and appropriately in many therapeutic areas in which assay sensitivity is not in doubt. These trials are often credible and have been widely used in such areas as treatment of cancer, infectious disease, and some cardiovascular conditions (for example, acute myocardial infarction treated with thrombolysis). In general, the larger the effect size, the less study-to-study variability in outcomes, and the fewer the instances of unexplained failure of the control agent to show superiority to placebo in well-controlled studies, the more persuasive is the case for using this design. Investigators who intend to perform an ACET will therefore need to review previous placebo-controlled trials of the control agent to see whether it can be persuasively shown that such information exists. The ACET should be as similar as possible to the past placebo-controlled trials with regard to patient selection, dose, end points, assessment procedures, use of concomitant therapy, and other pertinent study design characteristics (17, 20).

Given the inevitable residual uncertainty about the assay sensitivity of a trial that does not contain an internal standard, reliance on ACETs may also require more evidence of replicability than would be needed for trials intended to show differences. It should be appreciated, however, that even if assay sensitivity can be assumed, the effect that the active control can be presumed to have had under the study conditions will often be relatively small. In such cases, large sample sizes will be needed to provide the narrow confidence interval needed to ensure that the new drug is not inferior to the control by more than that amount. This issue is considered further in the Appendix.

Studying Relative Effectiveness

In some cases, a study may be intended to evaluate the comparative effectiveness of two known active treatments. In that case, too, the presence of assay sensitivity is essential to interpretation of the trial. If one cannot be confident that the trial could have distinguished active drug from placebo, one cannot be confident that it could have distinguished a more effective drug from a less effective drug. A three-arm study (new drug, placebo, and active control) is optimal because it can 1) assess assay sensitivity and, if assay sensitivity is confirmed, 2) measure the effect of the new drug and 3) compare the effects of the two active treatments.

Alternative Approaches

Not all placebo-controlled studies leave patients untreated. It is frequently possible to provide standard therapy while carrying out a superiority study—that is, a study intending to demonstrate an advantage of a treatment regimen over the control. Sometimes a new agent can be assessed by using an “add-on” study design in which all patients are given standard therapy and are randomly assigned to also receive either new agent or placebo. This design is common in trials of therapy for cancer, heart failure, and epilepsy, in which omitting standard therapy would generally be unacceptable. Such studies are not directly informative about a drug as monotherapy, but they do provide interpretable evidence of effectiveness in a well-defined setting and are particularly appropriate where clinical use of the new agent will largely be as added treatment. Moreover, if successful, they demonstrate the ability to provide benefit greater than the standard therapy alone, in contrast to the (usually) less clinically interesting demonstration that a new therapy is not worse than the standard. This design is not useful, however, if the new drug and standard therapy are pharmacologically similar.

Although we have argued that an informed patient may choose to accept pain or discomfort or to defer needed long-term therapy for a short time to participate in a placebo-controlled trial, we do not mean to suggest that indifference to patient discomfort is appropriate. Some study designs limit the duration of placebo exposure without compromising the rigor of the study. These include “early escape” designs and randomized withdrawal studies (31, 38). In an “early escape” study, patients are randomly assigned to receive new drug or placebo, but a well-defined treatment failure end point (such as persistence of symptoms or maintenance of elevated blood pressure at a specified time) is used as the basis for changing therapy in patients who are not benefiting from their initially assigned treatment. In a randomized withdrawal study, apparently responsive patients are given an investigational therapy for a period and are randomly assigned to receive placebo or to continue active therapy. The randomly assigned groups can be compared for a defined period or by using an “early escape” approach. This design was initially proposed by Amery (39) as a way of avoiding extended placebo treatment of patients with angina pectoris. A particular value of the randomized withdrawal study is that it demonstrates a persistent effect for durations that would be difficult to study in placebo-controlled trials.

Regulatory Status of Study Designs

Critics of placebo-controlled trials have often attributed their use to Food and Drug Administration practices that favor the smallest possible trials, seek to assess absolute efficacy, and ignore what they consider the more important clinical question of how a new drug compares with standard therapy (1–6). Although a broad range of trial designs can be used to demonstrate the effectiveness of a new drug (15), regulations describing adequate and well-controlled studies have since 1985 indicated concerns about the interpretation of ACETs, reflecting views expressed since the 1950s by numerous clinical and statistical researchers (15–33). Thus, where assay sensitivity cannot be established for an ACET, trials that show a difference between treatments (a placebo-controlled trial is only one such example) would be needed to demonstrate effectiveness. The basis for this requirement is not a preference for small trials (although efficiency is not a trivial matter) nor indifference to comparisons (although under law, a drug need not be superior to or even as good as other therapy to be approved), but rather the fundamental need for evidence of assay sensitivity to interpret an ACET as showing effectiveness of a new drug.

Conclusions

Placebo controls are clearly inappropriate for conditions in which delay or omission of available treatments would increase mortality or irreversible morbidity in the population to be studied. For conditions in which forgoing therapy imposes no important risk, however, the participation of patients in placebo-controlled trials seems appropriate and ethical, as long as patients are fully informed. Arguments to the contrary are not based on established ethical principles but rather rely on a literal reading of one passage in the Declaration of Helsinki that would also preclude the conduct of active-control trials, and even historically controlled trials, whenever effective treatment exists. It seems inconceivable that the authors of the 1975 revision intended such an outcome, and nothing in their explanation of the revision suggests they did (8). We therefore believe this interpretation is untenable.

If ACETs were always adequate substitutes for placebo-controlled trials, the ethical issue might not arise. Unfortunately, ACETs are often uninformative. They can neither demonstrate the effectiveness of a new agent nor provide a valid comparison to control therapy unless assay

sensitivity can be assured, which often cannot be accomplished without inclusion of a concurrent placebo group.

Appendix

Blackwelder (40) and others (20, 22, 24, 31) have pointed out that equivalence testing can be better described in most cases as a test of a one-sided hypothesis that the test drug is not inferior to the control by a defined amount, the “equivalence margin,” also called the noninferiority margin. The null and alternate hypotheses H_0 and H_a then become:

$$H_0 = E_S - E_T \geq \Delta$$

$$H_a = E_S - E_T \leq \Delta$$

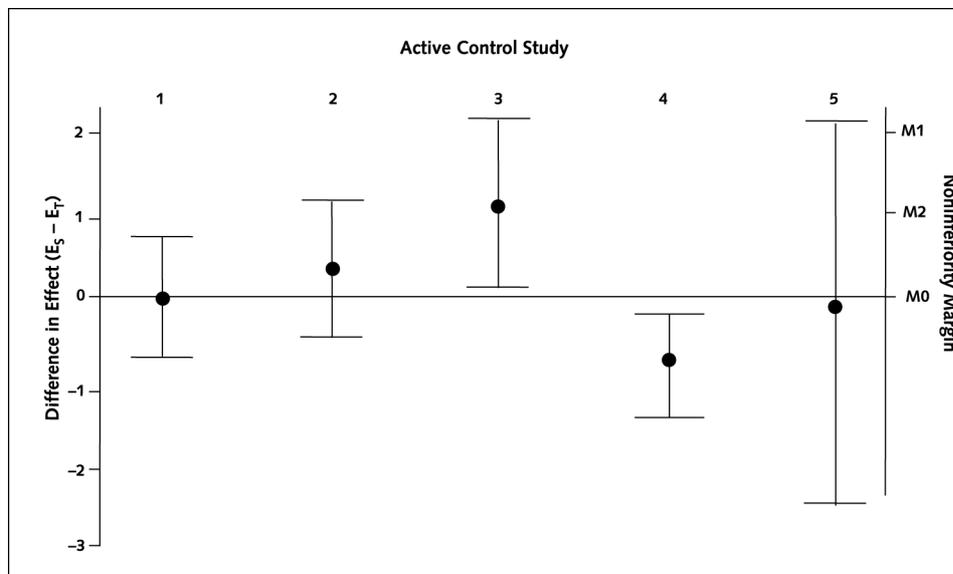
where E_S and E_T are the effects of the standard and test drugs, respectively, and Δ is the equivalence or noninferiority margin of interest. The null hypothesis is rejected if the upper bound of the confidence interval for the difference between the treatment being tested and control is smaller than the specified margin. In this case, the new agent is considered effective. A confidence interval that cannot exclude a difference greater than the margin would not permit rejection of the null hypothesis, and noninferiority would not be supported.

Choice of the margin is critical and depends on both knowledge of the effect of the control drug and clinical judgment. The margin chosen must be no larger than the smallest difference between control drug and placebo that could regularly be demonstrated in controlled trials. Exclusion of a difference greater than that margin would therefore mean that at least some part of

the effect of the control agent was preserved for the test drug and that the test drug therefore had at least some effect. If it were thought medically compelling to assure preservation of a specific fraction of the control drug effect, the specified margin would have to be made smaller than the largest possible noninferiority margin. If the control drug has not regularly shown superiority to placebo in trials of adequate size and design, the noninferiority margin must be set at zero: that is, only superiority to the control could be interpreted as evidence of effectiveness of the new drug. The ability to choose a margin representing the “guaranteed” effect of the control agent thus becomes functionally equivalent to saying that the trial has “assay sensitivity,” an ability to distinguish active from inactive treatments.

The use of equivalence margins is shown in the **Appendix Figure**. The hypothetical results of five different active-control studies are presented. The result of each study is shown on the y-axis as the difference between treatments ($E_S - E_T$, the difference between the standard drug control and test drug, with a positive difference favoring the standard drug) and confidence intervals for these differences. Three possible noninferiority margins are shown. M1 is the smallest effect the standard drug can be presumed to have in the study compared to a placebo treatment. M2 is a fraction of M1, chosen because it is considered essential to assure that the new drug retains some substantial fraction of the effect of the standard drug. M0 is the margin that must be used when the standard drug is not regularly superior to placebo:

Appendix Figure. Interpretation of equivalence margins in active-control trials.



Hypothetical results of five studies are shown. The result of each study is shown on the y-axis as the difference between treatments ($E_S - E_T$), where E_S represents the control and E_T represents the test drug. A positive difference favors the standard drug. Error bars show hypothetical 95% CIs for these differences. Three possible noninferiority margins are shown (M0, M1, M2). For more explanation, see the Appendix.

that is, only superiority of the new drug is acceptable evidence of effectiveness.

The results of studies 1 through 5 (**Appendix Figure**) are summarized in the following paragraphs.

Study 1: Where an effectiveness margin M_1 can be defined, effectiveness is shown because the confidence interval of the difference favoring the standard drug excludes inferiority greater than M_1 . Moreover, the study shows that more than 50% of the standard drug effect is preserved. If, however, assay sensitivity cannot be assumed (that is, there is no assurance that the standard drug had any effect in the study so that the noninferiority margin is M_0), the study would not show effectiveness of the new drug.

Study 2: Effectiveness would be shown by noninferiority of the new drug based on the M_1 margin but not if the more stringent M_2 margin were used.

Study 3: Effectiveness is not demonstrated because noninferiority to the standard drug based on the effectiveness margin M_1 is not demonstrated; the test drug may have no effect at all.

Study 4: Effectiveness of the test drug is demonstrated for any choice of margin by a showing of superiority of the test drug to the standard drug.

Study 5: Effectiveness is not demonstrated, despite the favorable point estimate of the effect of the test drug, because the wide confidence interval does not exclude inferiority to the standard drug greater than the M_1 margin.

From the Center for Drug Evaluation and Research and the Center for Biologics Evaluation and Research, U.S. Food and Drug Administration, Rockville, Maryland.

Disclaimer: The views expressed in these papers are those of the authors. They are similar to the conclusions and principles published in a Food and Drug Administration draft guidance document (33) based on a guideline developed by the International Conference on Harmonization, a body representing regulators and industry from the United States, the European Union, and Japan. The inferential problems associated with use of equivalence designs to show effectiveness are reflected in regulations (15) and guidelines (41).

Requests for Single Reprints: Robert Temple, MD, Center for Drug Evaluation and Research, U.S. Food and Drug Administration, 5600 Fishers Lane, HFD-101/WOC II—6014, Rockville, MD 20857.

Current Author Addresses: Dr. Temple: Center for Drug Evaluation and Research, U.S. Food and Drug Administration, 5600 Fishers Lane, HFD-101/WOC II—6014, Rockville, MD 20857.
Dr. Ellenberg: Center for Biologics Evaluation and Research, U.S. Food and Drug Administration, 1401 Rockville Pike, Rockville, Maryland 20852.

References

1. Rothman KJ, Michels KB. The continued unethical use of placebo controls. *N Engl J Med.* 1994;331:394-8.
2. Agostini A. Placebo and EU Guidelines [Letter]. *Lancet.* 1995;310:1710.
3. Aspinall RL, Goodman NW. Denial of effective treatment and poor quality of clinical information in placebo controlled trials of ondansetron for postoperative nausea and vomiting: a review of published trials. *BMJ.* 1995; 311:844-6.
4. Henry D, Hill S. Comparing treatments [Editorial]. *BMJ.* 1995;310: 1279.
5. Freedman B, Weijer C, Glass KC. Placebo orthodoxy in clinical research. I: Empirical and methodological myths. *J Law Med Ethics.* 1996;24: 243-51.
6. Freedman B, Weijer C, Glass KC. Placebo orthodoxy in clinical research. II: Ethical, legal, and regulatory myths. *J Law Med Ethics.* 1996;24: 252-9.
7. Angell M. The ethics of clinical research in the Third World [Editorial]. *N Engl J Med.* 1997;337:847-9.
8. World Medical Association Declaration of Helsinki. Recommendations guiding physicians in biomedical research involving human subjects. *JAMA.* 1997;277:925-6.
9. World Medical Association. Summary History of the World Medical Association Declaration of Helsinki. Ferney-Voltaire, France: World Medical Association; 1985.
10. American Medical Association Council on Ethical and Judicial Affairs. The use of placebo controls in clinical trials. Report 2-A-1996. Chicago: American Medical Assoc; 1996.
11. Grof P, Akhter MI, Campbell M, Gottfries CG, Khan I, Lapierre YD, et al. Clinical Evaluation of Psychotropic Drugs for Psychiatric Disorders: Principles and Proposed Guidelines. WHO Expert Series on Biological Psychiatry. vol 2. Seattle: Hogrefe & Huber; 1993:28-9.
12. Council for International Organizations of Medical Sciences. International Ethical Guidelines for Biomedical Research Involving Human Subjects. Geneva: Council for International Organizations of Medical Sciences; 1993.
13. Ellenberg SS, Temple R. Placebo-controlled trials and active control trials in the evaluation of new treatments. Part 2: Practical issues and specific cases. *Ann Intern Med.* 2000;133:464-70.
14. Levine RJ. Ethics and Regulation of Clinical Research. 2nd ed. New Haven, CT: Yale Univ Pr; 1986.
15. Adequate and well-controlled studies. Code of Federal Regulations, 21 Part 314.126. Revised as of 1 April 2000. Washington, DC: U.S. Government Printing Office; 2000.
16. Temple R. Government viewpoint of clinical trials. *Drug Information Journal.* 1982;16:10-7.
17. Temple R. Difficulties in evaluating positive control trials. *Proceedings of the American Statistical Association.* 1983;1-7.
18. Leber P. Hazards of inference: the active control investigation. *Epilepsia.* 1989;30(Suppl 1):S57-63.
19. Lasagna L. Placebos and controlled trials under attack [Editorial]. *Eur J Clin Pharmacol.* 1979;15:373-4.
20. Makuch R, Johnson M. Issues in planning and interpreting active control equivalence studies. *J Clin Epidemiol.* 1989;42:503-11.
21. Modell W, Houde RW. Factors influencing clinical evaluation of drugs

- with special reference to the double-blind technique. *JAMA*. 1958;167:2190-9.
22. **Fleming TR**. Treatment evaluation in active control studies. *Cancer Treat Rep*. 1987;71:1061-5.
 23. **Fleming TR**. Evaluation of active control trials in AIDS. *J Acquired Immune Defic Syndr*. 1990;3(Suppl 2):S82-7.
 24. **Jones B, Jarvis P, Lewis JA, Ebbutt AF**. Trials to assess equivalence: the importance of rigorous methods. *BMJ*. 1996;313:36-9.
 25. **Temple R**. Problems in interpreting active control equivalence trials. *Accountability in Research*. 1996;4:267-75.
 26. **Dollery CT**. A bleak outlook for placebos (and for science) [Editorial]. *Eur J Clin Pharmacol*. 1979;15:219-21.
 27. **Smith JL**. Placebos in clinical trials of peptic ulcer. ACG Committee on FDA-Related Matters. *Am J Gastroenterol*. 1989;84:469-74.
 28. **Freston JW**. Dose-ranging in clinical trials: rationale and proposed use with placebo or positive controls. *Am J Gastroenterology*. 1986;81:307-11.
 29. **Prien RF**. Methods and models for placebo use in pharmacotherapeutic trials. *Psychopharmacol Bull*. 1988;24:4-8.
 30. **Sachar DB**. Placebo-controlled trials clinical trials in gastroenterology. A position paper of the American College of Gastroenterology. *Am J Gastroenterology*. 1984;79:913-7.
 31. **Laska EM, Klein DF, Lavori PW, Levine J, Robinson DS**. Design issues for the clinical evaluation of psychotropic drugs. In: Prien RF, Robinson DS, eds. *Clinical Evaluation of Psychotropic Drugs: Principles and Guidelines*. New York: Raven Pr; 1994:29-35.
 32. **Pledger G, Hall DB**. Active control equivalence studies: do they address the efficacy issue? In: Peace KE, ed. *Statistical Issues in Drug Research and Development*. New York: Marcel Dekker; 1990:226-38.
 33. International Conference on Harmonization: choice of control group in clinical trials. *Federal Register*. 1999;64:51767-80.
 34. **Sramek JJ, Cutler NR**. The use of placebo controls [Letter]. *N Engl J Med*. 1995;332:62.
 35. **Max MB**. Divergent traditions in analgesic clinical trials. *Clin Pharmacol Ther*. 1994;56:237-41.
 36. **Yusuf S, Peto R, Lewis J, Collins R, Sleight P**. Beta blockade during and after myocardial infarction: an overview of the randomized trials. *Prog Cardiovasc Dis*. 1985;27:335-71.
 37. **Tramer MR, Reynolds DJ, Moore RA, McQuay HJ**. When placebo controlled trials are essential and equivalence trials are inadequate. *BMJ*. 1998;317:875-80.
 38. **Temple R**. Special study designs: early escape, enrichment, studies in non-responders. *Communications in Statistics*. 1994;23:499-531.
 39. **Amery W, Dony J**. A clinical trial design avoiding undue placebo treatment. *J Clin Pharmacol*. 1975;15:674-9.
 40. **Blackwelder WC**. "Proving the null hypothesis" in clinical trials. *Control Clin Trials*. 1982;3:345-53.
 41. Guideline for the format and content of the clinical and statistical sections of new drug applications. Rockville, MD: U.S. Department of Health and Human Services, Public Health Service, Food and Drug Administration; 1988.

Doubts and uncertainties suck my blood and poison my strength, and there are times when they implant despair—cold and terrible as death—in the soul. . . . Can you possibly restore my pure and powerful faith, unblemished by any shadow of doubt and despair!

Michael Bar-Zohar
Ben-Gurion: A Biography
 New York: Lambda Publishers; 1986

Submitted by:
 Jorge Goldberg, MD
 Mexico City, Mexico

Submissions from readers are welcomed. If the quotation is published, the sender's name will be acknowledged. Please include a complete citation (along with page number on which the quotation was found), as done for any reference.—*The Editor*