35

Guidance for Industry

Providing Clinical Evidence of Effectiveness for Human Drug and Biological Products

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 1998
Clinical 6

Guidance for Industry

Providing Clinical Evidence of Effectiveness for Human Drugs and Biological Products

Additional copies are available from:
the Drug Information Branch (HFD-210),
Center for Drug Evaluation and Research (CDER),
5600 Fishers Lane, Rockville, MD 20857 (Tel) 301-827-4573
Internet at http://www.fda.gov/cder/guidance/index.htm
or
Office of Communication,
Training, and Manufacturers Assistance (HFM-40)
Center for Biologics Evaluation and Research (CBER)
1401 Rockville Pike, Rockville, MD 20852-1448
http://www.fda.gov/cber/guidelines.htm
(Fax) 888-CBERFAX or 301-827-3844
(Voice Information) 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 1998
Clinical 6

TABLE OF CONTENTS

I. INTRODUCTION 1

II. QUANTITY OF EVIDENCE NECESSARY TO SUPPORT EFFECTIVENESS 2

A. Legal Standards for Drug and Biological Products 2

B. Scientific Basis for the Legal Standard 4

C. The Quantity of Evidence to Support Effectiveness 6

III. DOCUMENTATION OF THE QUALITY OF EVIDENCE SUPPORTING AN EFFECTIVENESS CLAIM 16

A. Reliance on Less Than Usual Access to Clinical Data or Detailed Study Reports 17

B. Reliance on Studies with Alternative, Less Intensive Quality Control/On-Site Monitoring 20

GUIDANCE FOR INDUSTRY[1]

Providing Clinical Evidence of Effectiveness[2] for Human Drug and Biological Products

I. INTRODUCTION

This document is intended to provide guidance to applicants planning to file new drug applications (NDAs), biologics license applications (BLAs), or applications for supplemental indications on the evidence to be provided to demonstrate effectiveness.

This document is also intended to meet the requirements of subsections 403(b)(1) and (2) of the Food and Drug Administration Modernization Act (the Modernization Act) of 1997 for human drug and biological products (P.L. 105-115).[3] Subsection 403(b)(1) directs FDA to provide guidance on the circumstances in which published matter may be the basis for approval of a supplemental application for a new indication. Section III of this guidance satisfies this requirement by describing circumstances in which published matter may partially or entirely support approval of a supplemental application. Subsection 403(b)(2) directs FDA to provide guidance on data requirements that will avoid duplication of previously submitted data by recognizing the availability of data previously submitted in support of an original application to support approval of a supplemental application. Section II of this guidance satisfies this requirement by describing a range of circumstances in which related existing data, whether from an original application or other sources, may be used to support approval of a supplemental application.

In 1962, Congress amended the Federal Food, Drug, and Cosmetic Act to add a requirement that, to obtain marketing approval, manufacturers demonstrate the effectiveness of their products through the conduct of adequate and well-controlled studies. Since then, the issue of what constitutes sufficient evidence of effectiveness has been debated by the Agency, the scientific community, industry, and others. Sound evidence of effectiveness is a crucial component of the Agency's benefit-risk assessment of a new product or use. At the same time, the demonstration of effectiveness represents a major component of drug development time and cost; the amount

---

[1] This guidance document represents the agency's current thinking on providing clinical evidence of effectiveness for human drug and biological products. 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 statute, regulations, or both.

[2] As used in this guidance, the term efficacy refers to the findings in an adequate and well-controlled clinical trial or the intent of conducting such a trial and the term effectiveness refers to the regulatory determination that is made on the basis of clinical efficacy and other data.

[3] The Modernization Act requirements in Section 403 also apply to animal drugs and medical devices. These products will be addressed in separate guidances.

and nature of the evidence needed can therefore be an important determinant of when and whether new therapies become available to the public. The public health is best served by the development of sound evidence of effectiveness in an efficient manner.

The science and practice of drug development and clinical evaluation have evolved significantly since the effectiveness requirement for drugs was established, and this evolution has implications for the amount and type of data needed to support effectiveness in certain cases. As a result of medical advances in the understanding of pathogenesis and disease staging, it is increasingly likely that clinical studies of drugs will be more narrowly defined to focus, for example, on a more specific disease stage or clinically distinct subpopulation. As a consequence, product indications are often narrower, the universe of possible indications is larger, and data may be available from a number of studies of a drug in closely related indications that bear on a determination of its effectiveness for a new use. Similarly, there may be studies of a drug in different populations, studies of a drug alone or in combination, and studies of different doses and dosage forms, all of which may support a particular new use of a drug. At the same time, progress in clinical evaluation and clinical pharmacology have resulted in more rigorously designed and conducted clinical efficacy trials, which are ordinarily conducted at more than one clinical site. This added rigor and scope has implications for a study's reliability, generalizability, and capacity to substantiate effectiveness.

Given this evolution, the Agency has determined that it would be appropriate to articulate its current thinking concerning the quantitative and qualitative standards for demonstrating effectiveness of drugs and biologics. FDA hopes that this guidance will enable sponsors to plan drug development programs that are sufficient to establish effectiveness without being excessive in scope. The guidance should also bring greater consistency and predictability to FDA's assessment of the clinical trial data needed to support drug effectiveness.

Another major goal of this guidance is to encourage the submission of supplemental applications to add new uses to the labeling of approved drugs. By articulating how it currently views the quantity and quality of evidence necessary to support approval of a new use of a drug, FDA hopes to illustrate that the submission of supplements for new uses need not be unduly burdensome.

II. QUANTITY OF EVIDENCE NECESSARY TO SUPPORT EFFECTIVENESS

A. Legal Standards for Drug and Biological Products

Drugs: The effectiveness requirement for drug approval was added to the Federal Food, Drug, and Cosmetic Act (the Act or the FDC Act) in 1962. Between passage of the Act in 1938 and the 1962 amendments, drug manufacturers were required to show only that their drugs were safe. The original impetus for the effectiveness requirement was Congress's growing concern about the misleading and unsupported claims being made by pharmaceutical companies about their drug products coupled with high drug prices. After two years of hearings on these issues, Congress adopted the 1962 Drug Amendments,

which included a provision requiring manufacturers of drug products to establish a drug's effectiveness by "substantial evidence." *Substantial evidence* was defined in section 505(d) of the Act as "evidence consisting of adequate and well-controlled investigations, including clinical investigations, by experts qualified by scientific training and experience to evaluate the effectiveness of the drug involved, on the basis of which it could fairly and responsibly be concluded by such experts that the drug will have the effect it purports or is represented to have under the conditions of use prescribed, recommended, or suggested in the labeling or proposed labeling thereof."

Since the 1962 Amendments added this provision to the statute, discussions have ensued regarding the quantity and quality of the evidence needed to establish effectiveness. With regard to quantity, it has been FDA's position that Congress generally intended to require at least two adequate and well-controlled studies, each convincing on its own, to establish effectiveness. (See e.g., Final Decision on Benylin, 44 FR 51512, 518 (August 31, 1979); *Warner-Lambert Co. V. Heckler,* 787 F. 2d 147 (3d Cir. 1986)). FDA's position is based on the language in the statute[4] and the legislative history of the 1962 amendments. Language in a Senate report suggested that the phrase "adequate and well-controlled investigations" was designed not only to describe the quality of the required data but the "quantum" of required evidence. (S. Rep. No. 1744, Part 2, 87th Cong. 2d Sess. 6 (1962))

Nevertheless, FDA has been flexible within the limits imposed by the congressional scheme, broadly interpreting the statutory requirements to the extent possible where the data on a particular drug were convincing. In some cases, FDA has relied on pertinent information from other adequate and well-controlled studies of a drug, such as studies of other doses and regimens, of other dosage forms, in other stages of disease, in other populations, and of different endpoints, to support a single adequate and well-controlled study demonstrating effectiveness of a new use. In these cases, although there is only one study of the exact new use, there are, in fact, multiple studies supporting the new use, and expert judgment could conclude that the studies together represent substantial evidence of effectiveness. In other cases, FDA has relied on only a single adequate and well-controlled efficacy study to support approval — generally only in cases in which a single multicenter study of excellent design provided highly reliable and statistically strong evidence of an important clinical benefit, such as an effect on survival, and a confirmatory study would have been difficult to conduct on ethical grounds.

In section 115(a) of the Modernization Act, Congress amended section 505(d) of the Act to make it clear that the Agency may consider "data from one adequate and well-controlled clinical investigation and confirmatory evidence" to constitute substantial

---

[4] Section 505(d) of the Act uses the plural form in defining "substantial evidence" as "adequate and well-controlled investigations, including clinical investigations." See also use of "investigations" in section 505(b) of the Act, which lists the contents of a new drug application.

evidence if FDA determines that such data and evidence are sufficient to establish effectiveness. In making this clarification, Congress confirmed FDA's interpretation of the statutory requirements for approval and acknowledged the Agency's position that there has been substantial progress in the science of drug development resulting in higher quality clinical trial data.

Biologics. Biological products are approved under authority of section 351 of the Public Health Service Act (PHS Act) (42 U.S.C.§ 262). Under section 351, as in effect since 1944, licenses for biologics have been issued only upon a showing that the products meet standards designed to ensure the "continued safety, purity, and potency" of the products. *Potency* has long been interpreted to include effectiveness (21 CFR 600.3(s)). In 1972, FDA initiated a review of the safety and effectiveness of all previously licensed biologics. The Agency stated then that proof of effectiveness would consist of controlled clinical investigations as defined in the provision for "adequate and well-controlled studies" for new drugs (21 CFR 314.126), unless waived as not applicable to the biological product or essential to the validity of the study when an alternative method is adequate to substantiate effectiveness (21 CFR 601.25 (d) (2)). One such adequate alternative was identified to be serological response data where a previously accepted correlation with clinical effectiveness exists. As with nonbiological drug products, FDA has approved biological products based on single, multicenter studies with strong results.

Although section 123(a) of the Modernization Act amended section 351 of the PHS Act to make it clear that separate licenses are not required for biological products and the establishments at which the products are made, the evidentiary standard for a biological product was not changed: the product must be shown to be "safe, pure, and potent" (section 351 (a)(2) of the PHS Act as amended). In the Modernization Act (section 123(f)) Congress also directed the agency to take measures to "minimize differences in the review and approval" of products required to have approved BLAs under section 351 of the PHS Act and products required to have approved NDAs under section 505(b)(1) of the FDC Act.

B. Scientific Basis for the Legal Standard

The usual requirement for more than one adequate and well-controlled investigation reflects the need for *independent substantiation* of experimental results. A single clinical experimental finding of efficacy, unsupported by other independent evidence, has not usually been considered adequate scientific support for a conclusion of effectiveness. The reasons for this include the following.

- Any clinical trial may be subject to unanticipated, undetected, systematic biases. These biases may operate despite the best intentions of sponsors and investigators, and may lead to flawed conclusions. In addition, some investigators may bring conscious biases to evaluations.

- The inherent variability in biological systems may produce a positive trial result by chance alone. This possibility is acknowledged, and quantified to some extent, in the statistical evaluation of the result of a single efficacy trial. It should be noted, however, that hundreds of randomized clinical efficacy trials are conducted each year with the intent of submitting favorable results to FDA. Even if all drugs tested in such trials were ineffective, one would expect one in forty of those trials to "demonstrate" efficacy by chance alone at conventional levels of statistical significance.[5] It is probable, therefore, that false positive findings (i.e., the chance appearance of efficacy with an ineffective drug) will occur and be submitted to FDA as evidence of effectiveness. Independent substantiation of a favorable result protects against the possibility that a chance occurrence in a single study will lead to an erroneous conclusion that a treatment is effective.

- Results obtained in a single center may be dependent on site or investigator specific factors (e.g., disease definition, concomitant treatment, diet). In such cases, the results, although correct, may not be generalizable to the intended population. This possibility is the primary basis for emphasizing the need for independence in substantiating studies.

- Rarely, favorable efficacy results are the product of scientific fraud.

Although there are statistical, methodologic, and other safeguards to address the identified problems, they are often inadequate to address these problems in a single trial. Independent substantiation of experimental results addresses such problems by providing consistency across more than one study, thus greatly reducing the possibility that a biased, chance, site-specific, or fraudulent result will lead to an erroneous conclusion that a drug is effective.

The need for independent substantiation has often been referred to as the need for replication of the finding. Replication may not be the best term, however, as it may imply that precise repetition of the same experiment in other patients by other investigators is the only means to substantiate a conclusion. Precise replication of a trial is only one of a number of possible means of obtaining independent substantiation of a clinical finding and, at times, can be less than optimal as it could leave the conclusions vulnerable to any systematic biases inherent to the particular study design. Results that are obtained from studies that are of different design and independent in execution, perhaps evaluating different populations, endpoints, or dosage forms, may provide support for a conclusion of effectiveness that is as convincing as, or more convincing than, a repetition of the same study.

---

[5] p-value = 0.05, two-tailed, which implies an error rate in the efficacy (false positive) tail of 0.025 or one in forty.

C. The Quantity of Evidence to Support Effectiveness

The following three sections provide guidance on the quantity of evidence needed in particular circumstances to establish substantial evidence of effectiveness. Section 1 addresses situations in which effectiveness of a new use may be extrapolated entirely from existing efficacy studies. Section 2 addresses situations in which a single adequate and well-controlled study of a specific new use can be supported by information from other related adequate and well-controlled studies, such as studies in other phases of a disease, in closely related diseases, of other conditions of use (different dose, duration of use, regimen), of different dosage forms, or of different endpoints. Section 3 addresses situations in which a single multicenter study, without supporting information from other adequate and well-controlled studies, may provide evidence that a use is effective.

In each of these situations, it is assumed that any studies relied on to support effectiveness meet the requirements for adequate and well-controlled studies in 21 CFR 314.126. It should also be appreciated that reliance on a single study of a given use, whether alone or with substantiation from related trial data, leaves little room for study imperfections or contradictory (nonsupportive) information. In all cases, it is presumed that the single study has been appropriately designed, that the possibility of bias due to baseline imbalance, unblinding, post-hoc changes in analysis, or other factors is judged to be minimal, and that the results reflect a clear prior hypothesis documented in the protocol. Moreover, a single favorable study among several similar attempts that failed to support a finding of effectiveness would not constitute persuasive support for a product use unless there were a strong argument for discounting the outcomes in the studies that failed to show effectiveness (e.g., study obviously inadequately powered or lack of assay sensitivity as demonstrated in a three-arm study by failure of the study to show efficacy of a known active agent).

Whether to rely on a single study to support an effectiveness determination is not often an issue in contemporary drug development. In most drug development situations, the need to find an appropriate dose, to study patients of greater and lesser complexity or severity of disease, to compare the drug to other therapy, to study an adequate number of patients for safety purposes, and to otherwise know what needs to be known about a drug before it is marketed will result in more than one adequate and well-controlled study upon which to base an effectiveness determination.

This guidance is not intended to provide a complete listing of the circumstances in which existing efficacy data may provide independent substantiation of related claims; rather, it provides examples of the reasoning that may be employed. The examples are applicable whether the claim arises in the original filing of an NDA or BLA, or in a supplemental application.

1. Extrapolation from Existing Studies

In certain cases, effectiveness of an approved drug product for a new indication, or effectiveness of a new product, may be adequately demonstrated without additional adequate and well-controlled clinical efficacy trials. Ordinarily, this will be because other types of data provide a way to apply the known effectiveness to a new population or a different dose, regimen or dosage form. The following are examples of situations in which effectiveness might be extrapolated from efficacy data for another claim or product.

a. Pediatric uses

The rule revising the Pediatric Use section of product labeling (21 CFR 201.57(f)(9)(iv)) makes allowance for inclusion of pediatric use information in labeling without controlled clinical trials of the use in children. In such cases, a sponsor must provide other information to support pediatric use, and the Agency must conclude that the course of the disease and the effects of the drug are sufficiently similar in the pediatric and adult populations to permit extrapolation from adult efficacy data to pediatric patients. Evidence that could support a conclusion of similar disease course and similar drug effect in adult and pediatric populations includes evidence of common pathophysiology and natural history of the disease in the adult and pediatric populations, evidence of common drug metabolism and similar concentration-response relationships in each population, and experience with the drug, or other drugs in its therapeutic class, in the disease or condition or related diseases or conditions. Examples in which pediatric use labeling information has been extrapolated from adult efficacy data include ibuprofen for pain and loratidine for seasonal allergic rhinitis.

b. Bioequivalence

The effectiveness of alternative formulations and new dosage strengths may be assessed on the basis of evidence of bioequivalence.

c. Modified-release dosage forms

In some cases, modified release dosage forms may be approved on the basis of pharmacokinetic data linking the new dosage form to a previously studied immediate-release dosage form. Because the pharmacokinetic patterns of modified-release and immediate-release dosage forms are not identical, it is generally important to have some understanding of the relationship of blood concentration to response, including an understanding of the time course of that relationship, to extrapolate the immediate-release

data to the modified-release dosage form.

d. Different doses, regimens, or dosage forms

Dose-response relationships are generally continuous such that information about the effectiveness of one dose, dosage regimen, or dosage form is relevant to the effectiveness of other doses, regimens, or dosage forms. Where blood levels and exposure are not very different, it may be possible to conclude that a new dose, regimen, or dosage form is effective on the basis of pharmacokinetic data alone. Even if blood levels are quite different, if there is a well-understood relationship between blood concentration and response, including an understanding of the time course of that relationship, it may be possible to conclude that a new dose, regimen, or dosage form is effective on the basis of pharmacokinetic data without an additional clinical efficacy trial. In this situation, pharmacokinetic data, together with the well-defined pharmacokinetic/pharmacodynamic (PK/PD) relationship, are used to translate the controlled trial results from one dose, regimen, or dosage form to a new dose, regimen, or dosage form (See also section II.C.2.a).

2. Demonstration of Effectiveness by a Single Study of a New Use, with Independent Substantiation From Related Study Data

The discussion that follows describes specific examples in which a single study of a new use, with independent substantiation from study data in related uses, could provide evidence of effectiveness. In these cases, the study in the new use and the related studies support the conclusion that the drug has the effect it is purported to have. Whether related studies are capable of substantiating a single study of a new use is a matter of judgment and depends on the quality and outcomes of the studies and the degree of relatedness to the new use.

a. Different doses, regimens, or dosage forms

As discussed in Sections II.C.1.d, it may be possible to conclude that a new dose, regimen, or dosage form is effective on the basis of pharmacokinetic data without an additional clinical efficacy trial where blood levels and exposure are not very different or, even if quite different, there is a well-understood relationship between blood concentration and response. Where the relationship between blood concentration and response is not so well understood and the pharmacokinetics of the new dose, regimen, or dosage form differ from the previous one, clinical efficacy data will likely be necessary to support effectiveness of a new regimen. In this case, a single additional efficacy study should ordinarily be sufficient. For example, a single controlled trial was needed to support the recent approval of a once

daily dose of risperidone because the once daily and twice daily regimens had different pharmacokinetics and risperidone's PK/PD relationship was not well understood.

b. Studies in other phases of the disease

In many cases, therapies that are effective in one phase of a disease are effective in other disease phases, although the magnitude of the benefit and benefit-to-risk relationship may differ in these other phases. For example, if a drug is known to be effective in patients with a refractory stage of a particular cancer, a single adequate and well-controlled study of the drug in an earlier stage of the same tumor will generally be sufficient evidence of effectiveness to support the new use.

c. Studies in other populations

Often, responses in subsets of a particular patient population are qualitatively similar to those in the whole population. In most cases, separate studies of effectiveness in demographic subsets are not needed (see also discussion of the pediatric population in section II.C.1.a) However, where further studies are needed, a single study would ordinarily suffice to support effectiveness in age, race, gender, concomitant disease, or other subsets for a drug already shown to be generally effective in a condition or to be effective in one population. For example, a single study was sufficient to support tamoxifen use in breast cancer in males.

d. Studies in combination or as monotherapy

For a drug known to be effective as monotherapy, a single adequate and well-controlled study is usually sufficient to support effectiveness of the drug when combined with other therapy (as part of a multidrug regimen or in a fixed-dose combination). Similarly, known effectiveness of a drug as part of a combination (i.e., its contribution to the effect of the combination is known) would usually permit reliance on a single study of appropriate design to support its use as monotherapy, or as part of a different combination, for the same use. For example, a single study of a new combination vaccine designed to demonstrate adequate immune response will ordinarily provide sufficient evidence of effectiveness if the new combination contains products or antigens already proven to be effective alone or in other combinations. These situations are common for oncologic and antihypertensive drugs, but occur elsewhere as well.

e. Studies in a closely related disease

Studies in etiologically or pathophysiologically related conditions, or studies of a symptom common to several diseases (e.g., pain) can support each other, allowing initial approval of several uses or allowing additional claims based on a single adequate and well-controlled study. For example, certain anti-coagulant or anti-platelet therapies could be approved for use in two different settings based on individual studies in unstable angina/acute coronary syndrome and in the postangioplasty state. Because the endpoints studied and the theoretical basis for use of an anti-coagulant or anti-platelet drug are similar, each study supports the other for each claim. Similarly, single analgesic studies in several painful conditions would ordinarily be sufficient to support either a general analgesic indication or multiple specific indications. The recent approval of lamotrigine for treatment of Lennox-Gastaut Syndrome (a rare, largely pediatric, generalized seizure disorder) was based on a single adequate and well-controlled trial, due in part to related data showing efficacy of the drug in partial-onset seizures in adults.

f. Studies in less closely related diseases, but where the general purpose of therapy is similar

Certain classes of drug therapy, such as antimicrobials and antineoplastics, are appropriate interventions across a range of different diseases. For therapies of this type, evidence of effectiveness in one disease could provide independent substantiation of effectiveness in a quite different disease. For example, it is possible to argue that evidence of effectiveness of an antimicrobial in one infectious disease setting may support reliance on a single study showing effectiveness in other settings where the causative pathogens, characteristics of the site of infection that affect the disease process (e.g., structure and immunology) and patient population are similar.[6] Similarly, for an oncologic drug, evidence of effectiveness in one or more tumor types may support reliance on a single study showing effectiveness against a different kind of tumor, especially if the tumor types have a common biological origin.

g. Studies of different clinical endpoints

Demonstration of a beneficial effect in different studies on two different clinically meaningful endpoints could cross-substantiate a claim for

[6] See Division of Anti-Infective Drug Products: Points to Consider in the Clinical Development and Labeling of Anti-Infective Drug Products, October 1992.

effectiveness for each outcome. For example, the initial claim for effectiveness of enalapril for heart failure was supported by one study showing symptom improvement over several months and a second study showing improved survival in a more severely ill population. The two different findings, each from an adequate and well-controlled study, led to the conclusion that enalapril was effective in both treating symptoms and improving survival.

h. Pharmacologic/pathophysiologic endpoints

When the pathophysiology of a disease and the mechanism of action of a therapy are very well understood, it may be possible to link specific pharmacologic effects to a strong likelihood of clinical effectiveness. A pharmacologic effect that is accepted as a validated surrogate endpoint can support ordinary approval (e.g., blood pressure effects, cholesterol-lowering effects) and a pharmacologic effect that is considered reasonably likely to predict clinical benefit can support accelerated approval under the conditions described in 21 CFR 314 Subpart H and 21 CFR 601 Subpart E (e.g., CD4 count and viral load effects to support effectiveness of anti-viral drugs for HIV infection). When the pharmacologic effect is not considered an acceptable effectiveness endpoint, but the linkage between it and the clinical outcome is strong, not merely on theoretical grounds but based on prior therapeutic experience or well-understood pathophysiology, a single adequate and well-controlled study showing clinical efficacy can sometimes be substantiated by persuasive data from a well-controlled study or studies showing the related pharmacologic effect.

For example, a single clearly positive trial can be sufficient to support approval of a replacement therapy such as a coagulation factor, when it is combined with clear evidence that the condition being treated is caused by a deficiency of that factor. Demonstration of physical replacement of the deficient factor or restoration of the missing physiologic activity provides strong substantiation of the clinical effect. The corrective treatment of an inborn error of metabolism could be viewed similarly. In the case of preventive vaccines, one adequate and well-controlled clinical trial may be supported by compelling animal challenge/protection models, human serological data, passive antibody data, or pathogenesis information. The more evidence there is linking effects on the pharmacologic endpoint to improvement or prevention of the disease, the more persuasive the argument for reliance on a single clinical efficacy study.

Note, however, that plausible beneficial pharmacologic effects have often not correlated with clinical benefit, and, therefore, caution must be observed in relying on a pharmacologic effect as contributing to evidence

of effectiveness. For example, pharmacologic effects such as arrhythmia suppression by Type 1 antiarrhythmics and increased cardiac output by phosphodiesterase inhibitors or beta adrenergic inotropes resulted in increased mortality, rather than, as was expected, decreased sudden death and improved outcome in heart failure. The reasons for the absence of an expected correlation between pharmacologic and clinical effects are diverse and can include an incompletely understood relationship between the pharmacologic effect and the clinical benefit and the presence of other pharmacologic effects attributable to a drug in addition to the effect being measured and thought to be beneficial. Generally, the utility of pharmacologic outcomes in providing independent substantiation will be greatest where there is prior experience with the pharmacologic class. Even in this case, however, it is difficult to be certain that a pharmacologic effect that correlates with a clinical benefit accounts for all the clinical benefit or that other effects are not present and relevant.

3. Evidence of Effectiveness from a Single Study

When the effectiveness requirement was originally implemented in 1962, the prevailing efficacy study model was a single institution, single investigator, relatively small trial with relatively loose blinding procedures, and little attention to prospective study design and identification of outcomes and analyses. At present, major clinical efficacy studies are typically multicentered, with clear, prospectively determined clinical and statistical analytic criteria. These studies are less vulnerable to certain biases, are often more generalizable, may achieve very convincing statistical results, and can often be evaluated for internal consistency across subgroups, centers, and multiple endpoints.

The added rigor and size of contemporary clinical trials have made it possible to rely, in certain circumstances, on a single adequate and well-controlled study, without independent substantiation from another controlled trial, as a sufficient scientific and legal basis for approval. For example, the approval of timolol for reduction of post-infarction mortality was based on a single, particularly persuasive (low p-value), internally consistent, multicenter study that demonstrated a major effect on mortality and reinfarction rate. For ethical reasons, the study was considered unrepeatable. The Center for Biologics Evaluation and Research has also approved a number of products based upon a single persuasive study. The Agency provided a general statement in 1995 describing when a single, multicenter study may suffice (60 FR 39181; August 1, 1995), but the Agency has not comprehensively described the situations in which a single adequate and well-controlled study might be considered adequate support for an effectiveness claim, or the characteristics of a single study that could make it adequate support for an effectiveness claim.

Whether to rely on a single adequate and well-controlled study is inevitably a matter of judgment. A conclusion based on two persuasive studies will always be more secure than a conclusion based on a single, comparably persuasive study. For this reason, reliance on only a single study will generally be limited to situations in which a trial has demonstrated a clinically meaningful effect on mortality, irreversible morbidity, or prevention of a disease with potentially serious outcome and confirmation of the result in a second trial would be practically or ethically impossible. For example, sequential repetition of strongly positive trials that demonstrated a decrease in post-infarction mortality, prevention of osteoporotic fractures, or prevention of pertussis would present significant ethical concerns. Repetition of positive trials showing only symptomatic benefit would generally not present the same ethical concerns.

The discussion that follows identifies the characteristics of a single adequate and well-controlled study that could make the study adequate support for an effectiveness claim. Although no one of these characteristics is necessarily determinative, the presence of one or more in a study can contribute to a conclusion that the study would be adequate to support an effectiveness claim.

a. Large multicenter study

In a large multicenter study in which (1) no single study site provided an unusually large fraction of the patients and (2) no single investigator or site was disproportionately responsible for the favorable effect seen, the study's internal consistency lessens concerns about lack of generalizability of the finding or an inexplicable result attributable only to the practice of a single investigator. If analysis shows that a single site is largely responsible for the effect, the credibility of a multicenter study is diminished.

b. Consistency across study subsets

Frequently, large trials have relatively broad entry criteria and the study populations may be diverse with regard to important covariates such as concomitant or prior therapy, disease stage, age, gender or race. Analysis of the results of such trials for consistency across key patient subsets addresses concerns about generalizability of findings to various populations in a manner that may not be possible with smaller trials or trials with more narrow entry criteria. For example, the timolol postinfarction study randomized patients separately within three severity strata. The study showed positive effects on survival in each stratum supporting a conclusion that the drug's utility was not limited to a particular disease stage (e.g., relatively low or high severity).

c. Multiple *studies* in a single study

Properly designed factorial studies may be analyzed as a series of pairwise comparisons, representing, within a single study, separate demonstrations of activity of a drug as monotherapy and in combination with another drug. This model was successfully used in ISIS II, which showed that for patients with a myocardial infarction both aspirin and streptokinase had favorable effects on survival when used alone and when combined (aspirin alone and streptokinase alone were each superior to placebo; aspirin and streptokinase in combination were superior to aspirin alone and to streptokinase alone). This represented two separate (but not completely independent) demonstrations of the effectiveness of aspirin and streptokinase.

d. Multiple endpoints involving different events

In some cases, a single study will include several important, prospectively identified primary or secondary endpoints, each of which represents a beneficial, but different, effect. Where a study shows statistically persuasive evidence of an effect on more than one of such endpoints, the internal weight of evidence of the study is enhanced. For example, the approval of beta-interferon (Betaseron) for prevention of exacerbations in multiple sclerosis was based on a single multicenter study, at least partly because there were both a decreased rate of exacerbations and a decrease in MRI-demonstrated disease activity — two entirely different, but logically related, endpoints.

Similarly, favorable effects on both death and nonfatal myocardial infarctions in a lipid-lowering, postangioplasty, or postinfarction study would, in effect, represent different, but consistent, demonstrations of effectiveness, greatly reducing the possibility that a finding of reduced mortality was a chance occurrence. For example, approval of abciximab as adjunctive treatment for patients undergoing complicated angioplasty or atherectomy was supported by a single study with a strong overall result on the combined endpoint (decreased the combined total of deaths, new infarctions, and need for urgent interventions) and statistically significant effects in separate evaluations of two components of the combined endpoint (decreased new infarctions and decreased need for urgent interventions). In contrast, a beneficial effect on multiple endpoints that evaluate essentially the same phenomenon and correlate strongly, such as mood change on two different depression scales or SGOT and CPK levels postinfarction, does not significantly enhance the internal weight of the evidence from a single trial.

Although two consistent findings within a single study usually provide reassurance that a positive treatment effect is not due to chance, they do not protect against bias in study conduct or biased analyses. For example, a treatment assignment not well balanced for important prognostic variables could lead to an apparent effect on both endpoints. Thus, close scrutiny of study design and conduct are critical to evaluating this type of study.

e. Statistically very persuasive finding

In a multicenter study, a very low p-value indicates that the result is highly inconsistent with the null hypothesis of no treatment effect. In some studies it is possible to detect nominally statistically significant results in data from several centers, but, even where that is not possible, an overall extreme result and significance level means that most study centers had similar findings. For example, the thrombolysis trials of streptokinase (ISIS II, GISSI) had very sizable treatment effects and very low p-values, greatly adding to their persuasiveness. Preventive vaccines for infectious disease indications with a high efficacy rate (e.g., point estimate of efficacy of 80% or higher and a reasonably narrow 95% confidence interval) have been approved based on a single adequate and well-controlled trial.

4. Reliance on a Single, Multicenter Study — Caveats

While acknowledging the persuasiveness of a single, internally consistent, strong multicenter study, it must be appreciated that even a strong result can represent an isolated or biased result, especially if that study is the only study suggesting efficacy among similar studies. Recently, the apparent highly favorable effect of vesnarinone, an inotropic agent, in heart failure (60% reduction of mortality in what appeared to be a well-designed, placebo-controlled, multicenter trial with an extreme p-value) has proven to be unrepeatable. In an attempt to substantiate the finding, the same dose of the drug that seemed lifesaving in the earlier study significantly increased mortality (by 26%), and a lower dose also appeared to have a detrimental effect on survival. Although the population in the second study was, on the whole, a sicker population than in the first, the outcomes in similarly sick patients in each study were inconsistent so this factor does not explain the contradictory results.

When considering whether to rely on a single multicenter trial, it is critical that the possibility of an incorrect outcome be considered and that all the available data be examined for their potential to either support or undercut reliance on a single multicenter trial. In the case of vesnarinone, there were other data that were not consistent with the dramatically favorable outcome in the multicenter study. These data seemed to show an inverse dose-response relationship, showed no suggestion

of symptomatic benefit, and showed no effect on hemodynamic endpoints. These inconsistencies led the Agency, with the advice of its Cardio-Renal Advisory Committee, to refuse approval — a decision borne out by the results of the subsequent study.

This example illustrates how inadequacies and inconsistencies in the data, such as lack of pharmacologic rationale and lack of expected other effects accompanying a critical outcome, can weaken the persuasiveness of a single trial. Although an unexplained failure to substantiate the results of a favorable study in a second controlled trial is not proof that the favorable study was in error — studies of effective agents can fail to show efficacy for a variety of reasons — it is often reason not to rely on the single favorable study.

III. DOCUMENTATION OF THE QUALITY OF EVIDENCE SUPPORTING AN EFFECTIVENESS CLAIM

When submitting the requisite quantity of data to support approval of a new product or new use of an approved product, sponsors must also document that the studies were adequately designed and conducted. Essential characteristics of adequate and well-controlled trials are described in 21 CFR 314.126. To demonstrate that a trial supporting an effectiveness claim is adequate and well-controlled, extensive documentation of trial planning, protocols, conduct, and data handling is usually submitted to the Agency, and detailed patient records are made available at the clinical sites.

From a scientific standpoint, however, it is recognized that the extent of documentation necessary depends on the particular study, the types of data involved, and the other evidence available to support the claim. Therefore, the Agency is able to accept different levels of documentation of data quality, as long as the adequacy of the scientific evidence can be assured. This section discusses the factors that influence the extent of documentation needed, with particular emphasis on studies evaluating new uses of approved drugs.

For the purposes of this section, the phrase *documentation of the quality of evidence* refers to (1) the completeness of the documentation and (2) the ability to access the primary study data and the original study-related records (e.g., subjects' medical records, drug accountability records) for the purposes of verifying the data submitted as evidence. These interrelated elements bear on a determination of whether a study is adequate and well-controlled.

In practice, to achieve a high level of documentation, studies supporting claims are ordinarily conducted in accordance with good clinical practices (GCPs). Sponsors routinely monitor all clinical sites, and FDA routinely has access to the original clinical protocols, primary data, clinical site source documents for on-site audits, and complete study reports.

However, situations often arise in which studies that evaluate the efficacy of a drug product lack the full documentation described above (for example, full patient records may not be available) or in which the study was conducted with less monitoring than is ordinarily seen in commercially sponsored trials. Such situations are more common for supplemental indications because postapproval studies are more likely to be conducted by parties other than the drug sponsor and those parties may employ less extensive monitoring and data-gathering procedures than a sponsor. Under certain circumstances, it is possible for sponsors to rely on such studies to support effectiveness claims, despite less than usual documentation or monitoring. Some of those circumstances are described below.

A. Reliance on Less Than Usual Access to Clinical Data or Detailed Study Reports

FDA's access to primary data has proven to be important in many regulatory decisions. There are also reasons to be skeptical of the conclusions of published reports of studies. Experience has shown that such study reports do not always contain a complete, or entirely accurate, representation of study plans, conduct and outcomes. Outright fraud (i.e., deliberate deception) is unusual. However, incompleteness, lack of clarity, unmentioned deviation from prospectively planned analyses, or an inadequate description of how critical endpoint judgments or assessments were made are common flaws. Typically, journal article peer reviewers only have access to a limited data set and analyses, do not see the original protocol and amendments, may not know what happened to study subjects that investigators determined to be non-evaluable, and thus may lack sufficient information to detect critical omissions and problems. The utility of peer review can also be affected by variability in the relevant experience and expertise of peer reviewers. FDA's experiences with the Anturane Reinfarction Trial, as well as literature reports of the efficacy of tacrine and the anti-sepsis HA-1A antibody, illustrate its concerns with reliance on the published medical literature.

Notwithstanding these concerns, the presence of some of the factors discussed below can make it possible for FDA to rely on studies for which it has less than usual access to data or detailed study reports to partially or entirely (the so-called *paper* filing) support an effectiveness claim. FDA's reliance on a literature report to support an effectiveness claim is more likely if FDA can obtain additional critical study details. Section 1 below describes additional information that, if available, would increase the likelihood that a study could be relied on to support an effectiveness claim. Section 2 describes factors that may make efficacy findings sufficiently persuasive to permit reliance on the published literature alone. Note that the factors outlined in Section 2 are relevant to an assessment of the reliability of literature reports generally, whether alone, or accompanied by other important information as discussed in Section 1.

1. Submission of Published Literature or Other Reports in Conjunction with Other Important Information that Enhances the Reliability of the Data

If a sponsor wishes to rely on a study conducted by another party and cannot obtain the primary data from the study, for most well-conducted studies it is possible to obtain other important information, such as a protocol documenting the prospective plans for the trial, records of trial conduct and procedures, patient data listings for important variables, and documentation of the statistical analysis. FDA has considerable experience evaluating large multicenter outcome studies sponsored by U.S. and European government agencies (NIH, British Medical Research Council) and private organizations (the ISIS studies, the SAVE study) for which there was limited access to primary study data, but for which other critical information was available. Providing as many as possible of the following important pieces of information about a study, in conjunction with the published report, can increase the likelihood that the study can be relied on to support an effectiveness claim:

a. The protocol used for the study, as well as any important protocol amendments that were implemented during the study and their relation to study accrual or randomization.

b. The prospective statistical analysis plan and any changes from the original plan that occurred during or after the study, with particular note of which analyses were performed pre- and post-unblinding.

c. Randomization codes and documented study entry dates for the subjects.

d. Full accounting of all study subjects, including identification of any subjects with on-treatment data who have been omitted from analysis and the reasons for omissions, and an analysis of results using all subjects with on-study data.

e. Electronic or paper record of each subject's data for critical variables and pertinent baseline characteristics. Where individual subject responses are a critical variable (e.g., objective responses in cancer patients, clinical cures and microbial eradications in infectious disease patients, death from a particular cause), detailed bases for the assessment, such as the case report, hospital records, and narratives, should be provided when possible.

f. Where safety is a major issue, complete information for all deaths and drop-outs due to toxicity. For postapproval supplemental uses, however, there is generally less need for the results of lab tests or for details of adverse event reports and, consequently, much more limited documentation may be sufficient (e.g., only for unexpected deaths and previously undescribed serious adverse effects). Exceptions to this

approach would include situations in which the population for the supplemental use is so different that existing safety information has limited application (e.g., thrombolysis in stroke patients versus myocardial infarction patients) or where the new population presents serious safety concerns (e.g., extension of a preventive vaccine indication from young children to infants).

2. Submission of Published Literature Reports Alone

The following factors increase the possibility of reliance on published reports alone to support approval of a new product or new use:

a. Multiple studies conducted by different investigators where each of the studies clearly has an adequate design and where the findings across studies are consistent.

b. A high level of detail in the published reports, including clear and adequate descriptions of statistical plans, analytic methods (prospectively determined), and study endpoints, and a full accounting of all enrolled patients.

c. Clearly appropriate endpoints that can be objectively assessed and are not dependent on investigator judgment (e.g., overall mortality, blood pressure, or microbial eradication). Such endpoints are more readily interpreted than more subjective endpoints such as cause-specific mortality or relief of symptoms.

d. Robust results achieved by protocol-specified analyses that yield a consistent conclusion of efficacy and do not require selected post hoc analyses such as covariate adjustment, subsetting, or reduced data sets (e.g., analysis of only responders or compliant patients, or of an "eligible" or "evaluable" subset).

e. Conduct of studies by groups with properly documented operating procedures and a history of implementing such procedures effectively.

There have been approvals based primarily or exclusively on published reports. Examples include the initial approval of secretin for evaluation of pancreatic function and recent approvals of bleomycin and talc for malignant pleural effusion and doxycycline for malaria.

B. Reliance on Studies with Alternative, Less Intensive Quality Control/On-Site Monitoring

Industry-sponsored studies typically use extensive on-site and central monitoring and auditing procedures to assure data quality. Studies supported by other sponsors may employ less stringent procedures and may use no on-site monitoring at all. An International Conference on Harmonisation guideline on good clinical practices,[7] recently accepted internationally, emphasizes that the extent of monitoring in a trial should be based on trial-specific factors (e.g., design, complexity, size, and type of study outcome measures) and that different degrees of on-site monitoring can be appropriate. In recent years, many credible and valuable studies conducted by government or independent study groups, often with important mortality outcomes, had very little on-site monitoring. These studies have addressed quality control in other ways, such as by close control and review of documentation and extensive guidance and planning efforts with investigators. There is a long history of reliance on such studies for initial approval of drugs as well as for additional indications. Factors that influence whether studies with limited or no monitoring may be relied on include the following:

1. The existence of a prospective plan to assure data quality.

2. Studies that have features that make them inherently less susceptible to bias, such as those with relatively simple procedures, noncritical entry criteria, and readily assessed outcomes.

3. The ability to sample critical data and make comparisons to supporting records (e.g., hospital records).

4. Conduct of the study by a group with established operating procedures and a history of implementing such procedures effectively.

---

[7] International Conference on Harmonisation Guidance for Industry E6, *Good Clinical Practice: Consolidated Guideline*, April 1996.

B. Reliance on Studies with Alternative, Less Intensive Quality Control/On-Site Monitoring

Industry-sponsored studies typically use extensive on-site and central monitoring and auditing procedures to assure data quality. Studies supported by other sponsors may employ less stringent procedures and may use no on-site monitoring at all. An International Conference on Harmonisation guideline on good clinical practices,[7] recently accepted internationally, emphasizes that the extent of monitoring in a trial should be based on trial-specific factors (e.g., design, complexity, size, and type of study outcome measures) and that different degrees of on-site monitoring can be appropriate. In recent years, many credible and valuable studies conducted by government or independent study groups, often with important mortality outcomes, had very little on-site monitoring. These studies have addressed quality control in other ways, such as by close control and review of documentation and extensive guidance and planning efforts with investigators. There is a long history of reliance on such studies for initial approval of drugs as well as for additional indications. Factors that influence whether studies with limited or no monitoring may be relied on include the following:

1. The existence of a prospective plan to assure data quality.

2. Studies that have features that make them inherently less susceptible to bias, such as those with relatively simple procedures, noncritical entry criteria, and readily assessed outcomes.

3. The ability to sample critical data and make comparisons to supporting records (e.g., hospital records).

4. Conduct of the study by a group with established operating procedures and a history of implementing such procedures effectively.

---

[7] International Conference on Harmonisation Guidance for Industry E6, *Good Clinical Practice: Consolidated Guideline*, April 1996.