Demonstrating Substantial Evidence of Effectiveness for Human Drug and Biological Products

Guidance for Industry

DRAFT GUIDANCE

This guidance document is being distributed for comment purposes only.

Comments and suggestions regarding this draft document should be submitted within 60 days of publication in the *Federal Register* of the notice announcing the availability of the draft guidance. Submit electronic comments to https://www.regulations.gov. Submit written comments to the Dockets Management Staff (HFA-305), Food and Drug Administration, 5630 Fishers Lane, Rm. 1061, Rockville, MD 20852. All comments should be identified with the docket number listed in the notice of availability that publishes in the *Federal Register*.

For questions regarding this draft document, contact (CDER) Ei Thu Lwin, Office of New Drug Policy, 301-796-0728 or (CBER) Office of Communication, Outreach and Development, 800-835-4709 or 240-402-8010, ocod@fda.hhs.gov.

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

December 2019
Clinical/Medical

Demonstrating Substantial Evidence of Effectiveness for Human Drug and Biological Products

Guidance for Industry

Additional copies are available from:
Office of Communications, Division of Drug Information
Center for Drug Evaluation and Research
Food and Drug Administration
10001 New Hampshire Ave., Hillandale Bldg., 4th Floor
Silver Spring, MD 20993-0002
Phone: 855-543-3784 or 301-796-3400; Fax: 301-431-6353
Email: druginfo@fda.hhs.gov
https://www.fda.gov/drugs/guidance-compliance-regulatory-information/guidances-drugs

and/or

Office of Communication, Outreach and Development
Center for Biologics Evaluation and Research
Food and Drug Administration
10903 New Hampshire Ave., Bldg. 71, Room 3128
Silver Spring, MD 20993-0002
Phone: 800-835-4709 or 240-402-8010
Email: ocod@fda.hhs.gov
https://www.fda.gov/vaccines-blood-biologics/guidance-compliance-regulatory-information-biologics

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

December 2019
Clinical/Medical

Contains Nonbinding Recommendations

Draft — Not for Implementation

TABLE OF CONTENTS

Demonstrating Substantial Evidence of Effectiveness for Human Drug and Biological Products Guidance for Industry

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 guidance complements and expands on the 1998 guidance entitled Providing Clinical Evidence of Effectiveness for Human Drug and Biological Products (the 1998 guidance)[1].

The 1998 guidance was issued in response to the Food and Drug Administration Modernization Act of 1997 (FDAMA) (Pub. L. 105–115), which stated that the substantial evidence requirement for effectiveness, which had generally been interpreted as calling for two adequate and well-controlled trials, could also be met by a single trial[2] plus confirmatory evidence. The 1998 guidance, therefore, provided many examples of the types of evidence that could be considered confirmatory evidence, with a specific focus on adequate and well-controlled trials of the test agent in related populations or indications, as well as a number of illustrations of a single adequate and well-controlled trial supported by convincing evidence of the drug's mechanism of action in treating a disease or condition.

FDAMA thus introduced a specific new area of flexibility in the evidence needed to support effectiveness, but there are many other characteristics of the evidence supporting effectiveness that can vary (notably, trial designs, trial endpoints, statistical methodology), and evidence that varies in such ways potentially can provide substantial evidence of effectiveness but because of these characteristics may provide greater or lesser certainty. These characteristics also deserve consideration and were not discussed in the 1998 guidance. FDA's consideration of these various designs, endpoints, and analyses which can differ in the strength of evidence they

[1] FDA updates guidances periodically. To make sure you have the most recent version of a guidance, check the FDA webpage. The guidances mentioned in this document are available on the guidance web page at https://www.fda.gov/drugs/guidance-compliance-regulatory-information/guidances-drugs, and https://www.fda.gov/vaccines-blood-biologics/guidance-compliance-regulatory-information-biologics.

[2] In this guidance, the terms "trial" and "clinical trial" have the same meaning as the term "clinical investigation" as the latter is defined in FDA regulations (see, e.g., 21 CFR 312.3(b)).

provide, reflects the Agency's longstanding flexibility when considering the types of data and evidence that can meet the substantial evidence requirement.

Although FDA's evidentiary standard for effectiveness has not changed since 1998, the evolution of drug development and science has led to changes in the types of drug development programs submitted to the Agency. Specifically, there are more programs studying serious diseases lacking effective treatment, more programs in rare diseases, and more programs for therapies targeted at disease subsets. There is a need for more Agency guidance on the flexibility in the amount and type of evidence needed to meet the substantial evidence standard in these circumstances. The approaches discussed in this guidance can yield evidence that meets the statutory standard for substantial evidence and reflect the evolving landscape of drug development.

The "substantial evidence" of effectiveness standard in the statute (discussed in Section II) refers to both the quality and the quantity of the evidence. It clearly provides that all clinical investigations supporting effectiveness should be of appropriate design and of high quality (i.e., adequate and well-controlled; discussed in Section III). Sponsors often seek advice on what trial design will be considered acceptable in various development programs. This guidance discusses, in part, what clinical trial designs are considered adequate and well-controlled, and under what circumstances it may be appropriate to use a given design (discussed in Section III.A).

The clinical endpoints studied are a critical aspect of evidence quality (discussed in Section III.B). The Agency accepts clinical endpoints that reflect patient benefits (i.e., how patients feel, function, or survive) or validated surrogate endpoints[3] (i.e., those that have been shown to predict a specific clinical benefit) as the basis for traditional approval. In contrast to traditional approval, accelerated approval can be based on a demonstrated effect on a surrogate endpoint that is reasonably likely to predict a clinical benefit but where there are not sufficient data to show that it is a validated surrogate endpoint. Effects on intermediate clinical endpoints[4] can also be a basis for accelerated approval. For drugs granted accelerated approval, FDA requires post-approval trials to verify the predicted clinical benefit.

This guidance also discusses the quantity of evidence needed in a given development program – i.e., two adequate and well-controlled trials, one adequate and well-controlled trial plus confirmatory evidence, or reliance on a previous finding of effectiveness of an approved drug when scientifically justified and legally permissible (i.e., no new effectiveness or pharmacodynamic data would be needed) (discussed in the 1998 guidance and Section IV.A, IV.B, and IV.C, respectively). It also expands upon the discussions included in the 1998 guidance on the types of mechanistic and pharmacologic evidence and non-clinical evidence that can constitute confirmatory evidence.

[3] For more information on validated surrogate endpoints, see the BEST (Biomarkers, EndpointS, and other Tools) Resource available at: https://www.ncbi.nlm.nih.gov/books/NBK453484/.

[4] An intermediate clinical endpoint is "a clinical endpoint that can be measured earlier than irreversible morbidity or mortality, that is reasonably likely to predict an effect on irreversible morbidity or mortality or other clinical benefit." Section 506(c)(1)(A) of the Federal Food, Drug, and Cosmetic Act (FD&C Act).

Although randomized superiority trials with a placebo- or active-control design generally provide the strongest evidence of effectiveness, this guidance discusses the circumstances under which trials not using a placebo control, superiority design, or randomization may be acceptable (discussed in Section V.A and V.B). In addition, this guidance also discusses situations in which human efficacy trials are not ethical or feasible, and the animal rule may be applied (discussed in Section V.C).

The finding of substantial evidence of effectiveness is necessary but not sufficient for FDA approval. The approval decision also requires a determination that the drug is safe for the intended use. As all drugs have adverse effects, evaluating whether a drug is "safe" involves weighing whether the benefits of the drug outweigh its risks under the conditions of use defined in labeling. Uncertainties regarding benefits and risks are considered when making an approval determination; a drug with greater risks may require a greater magnitude and certainty of benefit to support approval. This benefit-risk analysis, as well as other determinations necessary for approval, is outside the scope of this guidance.

In general, FDA's guidance documents do not establish legally enforceable responsibilities. Instead, guidances describe the Agency's current thinking on a topic and should be viewed only as recommendations, unless specific regulatory or statutory requirements are cited. The use of the word *should* in Agency guidances means that something is suggested or recommended, but not required.

II. STANDARD OF EFFECTIVENESS FOR DRUGS AND BIOLOGICS

A. Statutory standard

In 1962, Congress required for the first time that drugs be shown to be effective as well as safe. A drug's effectiveness must be established by "substantial evidence," which is defined 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."[5]

Under section 351 of the Public Health Service Act (PHS Act) (42 U.S.C.§ 262) licenses for biologics have been issued only upon a showing that the products are "safe, pure, and potent." Potency has long been interpreted to include effectiveness (21 CFR 600.3(s)). FDA has also generally considered "substantial evidence" of effectiveness to be necessary to support licensure

[5] The FD&C Act section 505(d) (21 U.S.C. § 355(d)).

of a biological product under section 351 of the PHS Act.[6]

FDA has interpreted the law as generally requiring at least two adequate and well-controlled clinical investigations,[7] each convincing on its own, to establish effectiveness (discussed in Section IV.A.1). Under specific circumstances, however, FDA has considered a large multicenter trial that has certain characteristics to satisfy the legal requirement for substantial evidence of effectiveness (discussed in Section II.C.3 of the 1998 guidance and Section IV.A.2). FDA may also rely on a previous finding of effectiveness of an approved drug when scientifically justified and legally permissible; in this case there is no need for additional adequate and well-controlled clinical efficacy trials (discussed in Section IV.C).

In addition to reliance on a single large multicenter trial or previous finding of effectiveness of an approved drug, there are other circumstances where substantial evidence of effectiveness can be provided outside of the setting of two adequate and well-controlled clinical investigations. Congress specifically provided for these in section 115(a) of FDAMA, which amended the statutory provision on substantial evidence of effectiveness, 21 U.S.C. § 355(d), to add the following:

> "If [FDA] determines, based on relevant science, that data from one adequate and well-controlled clinical investigation and confirmatory evidence (obtained prior to or after such investigation) are sufficient to establish effectiveness, [FDA] may consider such data and evidence to constitute substantial evidence."

This modification explicitly recognized the potential for FDA to find that one adequate and well-controlled clinical investigation with confirmatory evidence, including supportive data outside of a controlled trial, is sufficient to establish effectiveness (discussed in Section IV.B).

B. Scientific basis for the statutory standard

To establish a drug's effectiveness, it is essential to distinguish the effect of the drug "from other influences, such as spontaneous change in the course of the disease, placebo effect, or biased observation."[8] This is the basis for the statutory requirement that approval be based on adequate and well-controlled investigations, as well as the basis for FDA's regulations describing the characteristics of such investigations (i.e., design elements that are generally intended to minimize bias and permit a valid comparison with a control to provide a quantitative assessment of drug effect).

---

[6] 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, with limited exceptions, consist of controlled clinical investigations as defined in the provision for "adequate and well-controlled studies" for new drugs (21 CFR 314.126) (see former 21 CFR 601.25(d)(2) (2015) (revoked as no longer necessary, 81 FR 7445 (Feb. 12, 2016))). We note that, in section 123(f)) of FDAMA, 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 FD&C Act.

[7] See FDA regulation regarding adequate and well-controlled studies at 21 CFR 314.126.

[8] 21 CFR 314.126(a).

A second adequate and well-controlled investigation or confirmatory evidence provides substantiation of experimental results, which is a widely accepted scientific principle. This approach is intended to minimize the possibility that other influences such as bias and chance findings could result in a false conclusion that a drug is effective when in fact it is not (false positive).

III. THE QUALITY OF CLINICAL EVIDENCE TO ESTABLISH EFFECTIVENESS

The quality of clinical evidence to establish effectiveness and the resulting level of certainty about the demonstration of substantial evidence is impacted by the selection of trial design and trial endpoint(s) as well as statistical considerations, as discussed below.

A. Trial designs

Adequate and well-controlled clinical investigations provide the primary basis for determining whether there is substantial evidence to support the claims of effectiveness.[9] FDA regulation at 21 CFR 314.126(b) describes characteristics of an adequate and well-controlled clinical investigation, including choice of control, method of patient assignment to treatment (e.g., randomization), adequate measures to minimize bias (e.g., blinding), well-defined and reliable assessment of individuals' response (i.e., efficacy endpoint), and adequate analysis of the clinical investigation's results to assess the effects of the drug (i.e., statistical methods). Although randomized double-blinded, concurrently controlled superiority trials are usually regarded as the most rigorous design, as discussed further below, five types of controls are described in section 314.126:[10] placebo concurrent control, dose-comparison concurrent control, no treatment concurrent control, active treatment concurrent control, and historical control (a type of external control).[11] Of note, when the first version of the rule was published in 1970, historical controls and active treatment controls were included.[12] Thus, from its earliest description of adequate and well-controlled trials, FDA included trial designs (as discussed below) that may be more difficult to interpret, which reflected FDA's recognition that different trial designs (including choice of control) may be appropriate in different disease settings.

Establishing superiority to a concurrent control group (whether an active agent, including a lower dose of the test drug, or placebo) generally provides strong evidence of effectiveness, because a superiority design does not depend on assumptions regarding the effectiveness of the control.

---

[9] The FD&C Act section 505(d) (21 U.S.C. § 355(d)); 21 CFR 314.126(a).

[10] See 50 FR 7452, 7487 (February 22, 1985).

[11] The regulation uses the term "historical control," which is a subset of "external control." FDA also accepts other types of external controls. An externally controlled trial compares a group of subjects receiving the test treatment with a group of patients external to the trial, rather than to an internal control group consisting of patients from the same trial population assigned to a different treatment. The external control can be a group of patients, treated or untreated, at an earlier time (historical control) or a group, treated or untreated, during the same time period but in another setting. An important subset of externally controlled trials are "baseline controlled trials," where there is not a specific external control group but assurance, based on experience, that no change could occur (e.g., tumors are known not to shrink spontaneously or patients not given general anesthetic remain awake). See International Conference on Harmonisation E10 guidance on Choice of Control Group and Related Issues in Clinical Trials (ICH E10). This guidance uses the term "external control," except when referring to section 314.126.

[12] See 35 FR 7250, 7251-7252 (May 8, 1970).

However, each of the trial designs has distinct considerations; for example, the lack of blinding when using a no treatment control could introduce bias, which may attenuate confidence in the trial's results. The dose-comparison design may support the effectiveness of the highest dose when a positive dose response is seen, but could leave uncertainty about whether lower tested doses were effective.

Although demonstrating that a new drug is superior to an active control provides strong evidence of effectiveness, a common goal of active controlled trials is to show non-inferiority (NI), i.e., that the new drug is not less effective than the active control by a specified amount, that amount being no larger than the effect the active control was expected (the effect is not measured) to have had in the NI trial based on the drug's past performance in trials. Showing such non-inferiority allows a conclusion that the new drug is effective.[13] In general, with regard to establishing effectiveness, NI designs are credible and appropriate only in situations in which the active control has shown a consistent effect (generally compared with placebo) in prior superiority trials conducted in a patient population similar to the population in the clinical investigation being planned. Unless a placebo group (or other treatment group where the intent is to demonstrate superiority of the test drug) is also included, these NI trials depend on the assumption, not confirmed in the trial, that the active control had its anticipated effect (which is the basis for the NI margin) in the trial. As a result, the strength of evidence that may result from an NI trial can vary considerably depending on the specific disease setting and the choice of active control. An NI trial that meets its objective (with respect to the pre-specified statistical testing plan) could mean either that both drugs were effective or, if neither control nor drug has its expected effect, that neither was effective in the trial. Because interpretation of NI trials depends on assumptions not confirmed in the trial, this design is usually chosen when it would be unethical or infeasible to conduct one of the superiority designs discussed above (e.g., when withholding available therapy would not be clinically acceptable and the new drug is being studied as an alternative, rather than as an adjunct, to available therapy).

Externally controlled trials differ in several important ways from the other trial designs identified in 21 CFR 314.126. Most notably, random assignment is not a feature of external control designs. As a result, there may be differences in patient characteristics or concomitant treatments in the trial population compared to the external control population that lead to differences in outcomes that are unrelated to the investigational treatment. In addition, the lack of blinding could introduce bias. For these reasons, external control designs are usually reserved for specific circumstances, such as trials of diseases with high and predictable mortality or progressive morbidity (e.g., certain malignancies or certain rare diseases) and trials in which the effect of the drug is self-evident (e.g., general anesthetics).

Despite the limitations of externally controlled trials compared with concurrently controlled trials, strong support for effectiveness can emerge from externally controlled trials, especially when (1) the natural history of a disease is well defined, (2) the external control population is very similar to that of the treatment group, (3) concomitant treatments that affect the primary endpoint are not substantially different between the external control population and the trial population, and (4) the results provide compelling evidence of a change in the established progression of disease. Such results could include partial or complete response in a disease

[13] FDA guidance on Non-Inferiority Clinical Trials to Establish Effectiveness.

where spontaneous regression is not observed, or stabilization or improvement in function in a disease where progressive functional decline is well documented to occur over the duration of the treatment period in the trial. Another example of where there is strong evidence of drug effectiveness is reversal of clinical signs and symptoms following a toxic exposure or overdose after administration of a drug antidote. In all such circumstances, a detailed understanding of the full range of possible clinical outcomes, with a well-documented natural history of the disease in the absence of treatment, is essential to interpreting trial results and, therefore, drawing a conclusion about the effectiveness of the drug.

It is important to recognize that trial design alone does not determine whether evidence from the trial is sufficient to establish substantial evidence of effectiveness. For example, compelling results may overcome challenges associated with less rigorous trial designs, such as those with an external control. As discussed above, a small externally controlled trial with an outcome markedly superior to the well-established natural history of a disease may provide a compelling case for drug effectiveness. Similarly, a successful active-controlled NI trial of a new antimicrobial drug or of a new anticoagulant to prevent stroke in patients with atrial fibrillation can provide strong evidence of effectiveness when it is well-established that the effect of the control antimicrobial or anticoagulant drug is large.

Poor execution can render a trial of any design to be not adequate or not well-controlled and, therefore, unable to provide substantial evidence of effectiveness. Examples of this include (1) a randomized, double-blind, placebo-controlled trial where there is extensive drop-out of trial patients (with the potential for informative censoring), and (2) a randomized, double-blind, placebo-controlled trial in which unblinding is common due to an effect of the test drug, and where a modest treatment effect is found on a primary endpoint that is subject to bias when drug assignment is known (e.g., a physician global impression). In these cases, the trials might not be considered adequate and well-controlled.

B. Trial endpoints

One of the characteristics of an adequate and well-controlled clinical investigation is that "the methods of assessment of subjects' response are well-defined and reliable."[14] Such a method of assessment can be a clinical endpoint[15] or, where appropriate, a surrogate endpoint.[16]

[14] 21 CFR 314.126(b)(6).

[15] An endpoint is a precisely defined variable intended to reflect an outcome of interest as a measure of drug effect that is prespecified (i.e., chosen before the data are analyzed) and statistically analyzed to address a particular research question. A definition of "clinical endpoint" is provided in FDA guidance on Expedited Programs for Serious Conditions – Drugs and Biologics (FDA guidance on expedited programs). A clinical endpoint can be used to support traditional approval.

[16] A definition of "surrogate endpoint" is provided in FDA guidance on expedited programs. A surrogate endpoint that has been shown to predict a specific clinical benefit can be used to support traditional approval. A surrogate endpoint that is reasonably likely to predict clinical benefit can be used to support accelerated approval. Accelerated approval can also be based on an effect on a clinical endpoint that can be measured earlier than irreversible morbidity or mortality (IMM) and that is reasonably likely to predict an effect on IMM or other clinical benefit. See FDA web page on Table of Surrogate Endpoints That Were the Basis of Drug Approval or Licensure, available at https://www.fda.gov/Drugs/DevelopmentApprovalProcess/DevelopmentResources/ucm613636.htm.

Although the statutory standard for effectiveness does not refer to particular endpoints or state a preference for clinical endpoints over surrogate endpoints, it is well established that the effect shown in the adequate and well-controlled clinical investigations, must be, in FDA's judgment, clinically meaningful.[17]

Many disease specific guidances have been issued by the Agency that can assist sponsors in identifying an appropriate trial endpoint. In addition, discussion with appropriate review divisions early in clinical development can assist sponsors in identifying appropriate trial endpoints for a particular development program.

C. Statistical considerations

The strength of evidence in each trial contributing to meeting the substantial evidence standard should be assessed by appropriate statistical methods. The uncertainty about the findings from each trial should be sufficiently small and the findings should be unlikely to result from chance alone, as demonstrated by a statistically significant result or a high posterior probability of effectiveness.[18] Statistical approaches should be specified in advance, to limit erroneous conclusions resulting from multiplicity.

IV. THE QUANTITY OF CLINICAL EVIDENCE TO ESTABLISH EFFECTIVENESS

A. Meeting the substantial evidence standard based on two adequate and well-controlled clinical investigations

1. Two adequate and well-controlled clinical investigations

In many situations FDA requires two adequate and well-controlled trials to establish effectiveness. This reflects the need for substantiation of experimental results, which 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. Although two positive identically designed and conducted trials can provide substantial evidence of effectiveness, precise replication of a trial is only one of a number of possible means of obtaining substantiation of a clinical finding and, at times, can provide less persuasive evidence of benefit, as it could leave the conclusions of both trials vulnerable to any systematic biases inherent to the particular study design.

Two positive trials with differences in design and conduct may be more persuasive, as unrecognized design flaws or biases in study conduct will be less likely to impact the outcomes of both trials. The consistency of results across two trials also greatly reduces the possibility that a biased, chance, site-specific, or fraudulent result will lead to an erroneous conclusion that a

[17] See preamble to FDA final rule on accelerated approval (57 FR 58942, 58944 (December 11, 1992)).

[18] In a Bayesian framework the strength of evidence is assessed by the probability that the drug is effective given the data rather than by statistical significance.

drug is effective. Such trials also may be more informative: for example, two positive trials using the same endpoint but with distinct study populations within the same proposed indication (e.g., one trial studying a new glucose-lowering drug in patients with type 2 diabetes receiving only diet and exercise therapy, and a second trial in patients with type 2 diabetes already on two or three oral antihyperglycemic agents) may provide evidence that is more generalizable to the population that will take the drug than two identical trials in a narrower population. Similarly, two trials in the same disease using different but related clinical endpoints could support effectiveness and provide broader information about the drug's effect (e.g., one trial showing symptom improvement and a second trial showing improved survival in a more severely ill population).

2. One adequate and well-controlled large multicenter trial that can provide substantial evidence of effectiveness

In general, substantiation of a drug's effectiveness obtained with two trials, especially with complementary design, as discussed above, will provide more convincing evidence of effectiveness than would a single trial. In some circumstances, however, there may not be a meaningful difference between the strength of evidence provided by a single large multicenter adequate and well-controlled trial and that provided by two smaller adequate and well-controlled trials. In such cases, the large multicenter trial can be considered, both scientifically and legally, to be, in effect, multiple trials and can be relied on to provide substantial evidence of effectiveness. Large multicenter trials can include a broad range of subjects and investigation sites and have procedures in place to ensure trial quality (e.g., investigation site selection, monitoring, and auditing). They generally are less vulnerable to certain biases such as selection or measurement bias, are often more generalizable to the intended population, and can often be evaluated for internal consistency across subgroups, centers, and multiple endpoints.

Reliance on a single large multicenter trial to establish effectiveness should generally be limited to situations in which the trial has demonstrated a clinically meaningful and statistically very persuasive effect on mortality, severe or irreversible morbidity, or prevention of a disease with potentially serious outcome, and with other characteristics described below, and confirmation of the result in a second trial would be impracticable or unethical. For example, conducting a second trial after a strongly positive trial had demonstrated a decrease in post-infarction mortality, or prevention of pertussis would generally present significant ethical concerns. Repetition of positive trials showing only symptomatic benefit would generally not present the same ethical concerns.

In addition to the expectation that the single trial is large and multicenter, there should be no single trial site that is the main contributor to the observed effect, either by virtue of having a much bigger effect or many more patients than other sites; these characteristics help address concerns about bias and chance findings associated with a single trial. As noted above it would also be expected that the effect size on the primary endpoint and the statistical analysis results are both persuasive.

Other characteristics, discussed below, also support the persuasiveness of a single trial in supporting the conclusion that there is substantial evidence of effectiveness. Finding consistent,

and clinically meaningful effects on distinct prospectively specified endpoints (e.g., an effect on both myocardial infarction and stroke for a drug being studied for cardiovascular benefit) can provide further evidence that the results are not due to chance. Moreover, an effect on a meaningful, objective endpoint, such as certain imaging endpoints, may complement a more subjective endpoint, such as a clinician- or patient-reported outcome. In these cases, the internal consistency across endpoints not only reduces the possibility of a chance finding but also may further support the clinical utility of the results.

Frequently, large multicenter trials have relatively broad entry criteria and the trial 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 important patient subgroups can address 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.

Furthermore, there may be other characteristics of a large multicenter trial that increase confidence in its results. For example, the multicenter trial may sometimes be appropriately analyzed as "multiple trials" within a single trial. An example is a 4-arm ("2×2 factorial") trial (placebo, drug A, drug B, and drug A + drug B) in which the effectiveness of drug A could be supported by two controlled comparisons if the combination of drug A + drug B is superior to drug B alone *and* drug A is superior to placebo.

Although a large multicenter trial with robust results can be persuasive, even a robust result can arise from bias. For example, although two consistent findings within a single trial usually provide reassurance that a positive treatment effect is not due to chance, they do not protect against bias in trial conduct, biased analyses, or fraud. Thus, close scrutiny of trial conduct, including, for example, completeness of follow-up, methods of analysis, imputation of missing data, evaluation of trial endpoints, is critical to evaluating such trials. Findings from other trials that are not consistent with the findings of the single positive trial would need to be considered collectively, and could weaken the overall strength of evidence.

B. Meeting the substantial evidence standard based on one adequate and well-controlled clinical investigation plus confirmatory evidence

Under certain circumstances and consistent with FDAMA, FDA can conclude that one adequate and well-controlled clinical investigation plus confirmatory evidence is sufficient to establish effectiveness. FDA will consider a number of factors when determining whether reliance on a single adequate and well-controlled clinical investigation plus confirmatory evidence is appropriate. These factors may include the persuasiveness of the single trial; the robustness of the confirmatory evidence; the seriousness of the disease,[19] particularly where there is an unmet medical need; the size of the patient population; and whether it is ethical and practicable to conduct more than one adequate and well-controlled clinical investigation. Sponsors intending to establish substantial evidence of effectiveness using one adequate and well-controlled clinical

[19] While seriousness of the disease is one of the factors that FDA considers, reliance on a single trial plus confirmatory evidence to establish effectiveness is not limited only to drugs for "serious diseases," as the term is defined in 21 CFR 312.300(b)(1).

investigation plus confirmatory evidence should consult FDA in advance to discuss the appropriateness of such an approach for their development program.

Confirmatory evidence could include, for example, adequate and well-controlled clinical investigations in a related disease area, certain types of real world evidence[20] such as extensive data on outcomes that provide further support for the lack of effect seen in the control group in the randomized trial, compelling mechanistic evidence in the setting of well-understood disease pathophysiology (e.g., pharmacodynamic data or compelling data from nonclinical testing), or well-documented natural history of the disease.

Below are examples of when a single adequate and well-controlled clinical investigation, together with confirmatory evidence, can establish effectiveness. The strength of the single trial will affect the extent of confirmatory evidence required – for example, a trial showing compelling efficacy results (but not rising to the level that would be provided by a large multicenter trial, as discussed in Section IV.A.2) may require less confirmatory evidence.

1. One adequate and well-controlled clinical investigation on a new indication for an approved drug, supported by existing adequate and well-controlled clinical investigation(s) that demonstrated the effectiveness of the drug for its other, closely related approved indication(s)

To establish effectiveness for a new indication of a product already approved by FDA – where the new indication is closely related to the other approved indication(s) – substantial evidence of effectiveness can be based on one adequate and well-controlled clinical investigation, generally a randomized concurrently controlled trial, of the new indication, supported by the confirmatory evidence provided by the existing adequate and well-controlled clinical investigation(s) that established effectiveness of the product for the related indication(s). See Section II.C.2 of the 1998 guidance for more details.

2. One adequate and well-controlled clinical investigation supported by data that provide strong mechanistic support

A single adequate and well-controlled clinical investigation, generally a randomized concurrently controlled trial, together with earlier phase clinical results and/or testing that provide compelling mechanistic evidence in the setting of well-understood disease pathophysiology, may be sufficient to provide substantial evidence of effectiveness of a new drug or a new indication. The mechanistic evidence would generally be obtained from clinical testing using a relevant and well understood pharmacodynamic endpoint not accepted by itself as an endpoint to establish evidence of effectiveness. It also could be collected from other sources, such as animal studies (e.g., those using an established, relevant animal model to study the effect of the drug on a pharmacodynamic marker of known relevance to humans), or a combination of

[20] Real world evidence is the clinical evidence regarding the usage, and potential benefits or risks, of a medical product derived from analysis of real world data. Real world data are data relating to patient health status and/or the delivery of health care routinely collected from a variety of sources. See FDA real world evidence web page, available at https://www.fda.gov/science-research/science-and-research-special-topics/real-world-evidence.

the two.[21] An example is enzyme replacement therapy, where a single adequate and well-controlled clinical investigation that demonstrates the therapy's efficacy is supported by evidence that the condition is caused by the enzyme deficiency and by earlier results that show the therapy increases enzyme activity to biologically active levels at the appropriate site and/or reduces disease-specific substrates. Another example could be a trial of a drug which is a mineral or vitamin replacement that showed restoration of accepted normal concentrations, in concert with a prior large body of information showing the clinical consequences of deficiency states.

3. One adequate and well-controlled clinical investigation with compelling results, supported by additional data from the natural history of the disease

In certain circumstances, FDA accepts one adequate and well-controlled clinical investigation that has generated compelling results as the basis to demonstrate effectiveness, when the single trial is supported by additional data from the natural history of the disease that reinforce the very persuasive finding. For example, a single trial showing marked improvement in survival compared to a control group, either external to the trial or concurrent, could be supported by data from separate sources (e.g., a natural history study, case report forms, or registries) that demonstrate a very limited median survival time or other clinically highly important outcome without treatment. In this case, the natural history data would represent confirmatory evidence.

4. One adequate and well-controlled clinical investigation of the new drug, supported by scientific knowledge about the effectiveness of other drugs in the same pharmacological class

In certain circumstances, FDA accepts one adequate and well-controlled clinical investigation as the basis to demonstrate effectiveness, when the single trial is supported by confirmatory evidence of effectiveness from adequate and well-controlled trials of other drugs in the same pharmacological class.[22] For example, the approval of two angiotensin II receptor blockers, losartan and irbesartan, for the treatment of diabetic nephropathy in patients with type 2 diabetes, hypertension, and abnormal kidney function, was based on effectiveness data from a single trial of each drug, supported by similarly favorable results from a single trial of the other drug. In this

[21] FDA supports the principles of the "3Rs," to reduce, refine, and replace animal use in testing when feasible. FDA encourages sponsors to consult with us if they wish to use a non-animal testing method they believe is suitable, adequate, validated, and feasible. FDA will consider if such an alternative method could be assessed for equivalency to an animal test method.

[22] Reliance on data concerning a different drug raises legal issues that will need to be considered in each case. If the applicant owns the data concerning the other drug, or has a right to refer to those data, such as a license, then the legal concerns are satisfied. In the example of losartan and irbesartan cited in the text, the two applicants each agreed to permit the other to rely on their data. If there is not such permission, for an NDA, the question will be raised whether the reliance makes the application a 505(b)(2) application. If so, that may require compliance with patent certification requirements applicable to such applications and may mean that the submission or approval of the application will be affected by statutory exclusivity provisions. For a BLA, in certain circumstances reliance on data not owned by the applicant, that is not in the public domain, and for which the applicant does not have a right of reference would raise additional legal considerations.

case, the two single trials supplied the needed confirmatory evidence for each other, as neither drug would have been approved for this indication based on the single trial alone.[23]

Whether this scenario applies to a particular development program depends on a number of factors, including but not limited to: (1) the strength of the evidence for effectiveness from the single trial; and (2) the relevance of the additional data derived from other drugs in the same class, including the similarity between the new drug and other drugs in the same class, particularly the pharmacologic activity or specificity of mechanism of action.[24]

C. Meeting the substantial evidence standard for a new population or a different dose, regimen, or dosage form, based on reliance of FDA's previous finding of effectiveness of an approved drug when scientifically justified and legally permissible

When scientifically justified and legally permissible, FDA can rely on its previous finding of effectiveness of an approved drug to conclude that the drug "will have the effect it purports or is represented to have,"[25] thus not requiring additional adequate and well-controlled clinical efficacy trials. Ordinarily, this will be because other types of evidence provide a way to apply the known effectiveness to a new population or a different dose, regimen, or dosage form. For example, the effectiveness of a drug for pediatric use can sometimes be based on FDA's previous finding of effectiveness of the drug in adults, together with scientific evidence that justifies such reliance.[26] In this case, the scientific evidence may include, for example, evidence supporting a conclusion of similar disease course and pathophysiologic basis in adult and pediatric populations, and similar pharmacologic activity of the drug in adults and children (e.g., similar concentration-response relationships), as well as similar blood levels of the drug in adults and children. The effectiveness of new dosage forms or dosing regimens may be demonstrated by the effectiveness trial(s) on the original dosage form or regimen, together with evidence that both the dosage forms or regimens have similar pharmacokinetic (PK) profiles. In this case no new effectiveness or pharmacodynamic data would be needed, but sufficient safety data would still be needed. See Section II.C.1. of the 1998 guidance for more details.

[23] See Secondary Review Memo on losartan, May 3, 2002, available at https://www.accessdata.fda.gov/drugsatfda_docs/nda/2002/20386-S028_COZAAR_Medr1.pdf; see also the FDA-approved labels for both products.

[24] A product development program under this scenario may result in a small safety database. Sponsors should consult FDA guidance on Premarketing Risk Assessment, which notes that the appropriate size of a safety database depends on a number of factors specific to the product; two of them are particularly relevant to this scenario, i.e., the product's novelty (i.e., whether it represents a new treatment or is similar to available treatment) and the availability of alternative therapies and the relative safety of those alternatives as compared to the new product. For more details, see FDA guidance on Premarketing Risk Assessment.

[25] See the statutory definition of "substantial evidence" in section 505(d) of the FD&C Act.

[26] Section 505B(a)(2)(B)(i) of the FD&C Act (21 U.S.C. § 355c(a)(2)(B)(i)).

V. EXAMPLES OF CLINICAL CIRCUMSTANCES WHERE ADDITIONAL FLEXIBILITY MAY BE WARRANTED

The statutory standard of "substantial evidence" contains both a statement of what kind of evidence must exist ("adequate and well-controlled investigations") and also an element of expert judgment. Thus the standard requires that the investigations be such that "it could fairly and responsibly be concluded by [qualified] experts that the drug will have the effect it purports or is represented to have,"[27] and permits approval on the basis of one trial and confirmatory evidence only "If [FDA] determines, based on relevant science, that data . . . are sufficient to establish effectiveness." For example, while FDA regulations outline five different types of studies that might be considered adequate and well-controlled,[28] it has always been recognized that some designs (e.g., placebo concurrent control) provide more certainty than others (e.g., external controls). FDA experts may "fairly and responsibly" rely on study designs that produce less certainty in some circumstances when a better design is not feasible or ethical. This may be the case for life-threatening and severely debilitating diseases with an unmet medical need, for certain rare diseases, or potentially even for a more common disease where the availability of existing treatments makes certain design choices infeasible or unethical. FDA would not, however, find it responsible to rely on such design choices in other situations in which, for example, the drug will be used for a less serious disease and greater certainty about benefits and risks is needed, or in cases where designs providing more certainty are possible. In all cases, FDA must reach the conclusion that there is substantial evidence of effectiveness to approve a drug; however, the degree of certainty supporting such a conclusion may differ, depending on clinical circumstances (e.g., severity and rarity of the disease and unmet medical need).

This reflects the longstanding awareness that, in certain settings, a somewhat greater risk (compared to placebo-controlled or other randomized superiority trials) of false positive conclusions – and therefore less certainty about effectiveness – may be acceptable, when balanced against the risk of rejecting or delaying the marketing of an effective therapy, as described below for an unmet medical need. The data supporting effectiveness could, despite the greater risk of error, support a conclusion that there is substantial evidence of effectiveness. Therefore, when selecting a trial design, a sponsor should consider the specific clinical circumstance, including the severity of the disease, unmet medical need (e.g., whether there is available therapy), the rarity of the disease, and whether it is feasible and ethical to conduct a randomized concurrently controlled superiority trial.

A. When the disease is life-threatening or severely debilitating with an unmet medical need

As defined in 21 CFR 312, subpart E (21 CFR 312.81), the term "life-threatening" means diseases or conditions where the likelihood of death is high unless the course of the disease is interrupted, and diseases or conditions with potentially fatal outcomes, where the endpoint of clinical trial analysis is survival; the term "severely debilitating" means diseases or conditions

[27] The law is clear that it is the FDA which "must determine, after giving full consideration to all of the evidence that has been submitted, including expert opinions, if the studies meet the regulatory criteria and show effectiveness." Warner-Lambert Co. v. Heckler, 787 F.2d 147, 154 (3rd Cir. 1986).

[28] 21 CFR 314.126(b)(2).

that cause major irreversible morbidity. An unmet medical need is a condition whose treatment or diagnosis is not addressed adequately by available therapy.[29]

Subpart E regulations promulgated in 1988[30] call for FDA to exercise its broad scientific judgment in applying the evidentiary approval standards to drugs for life-threatening and severely debilitating diseases, especially where there is no satisfactory alternative therapy. In addition, the accelerated approval regulations built upon this recognition by acknowledging that reliance on a surrogate endpoint "almost always introduces some uncertainty into the risk/benefit assessment, because clinical benefit is not measured directly and the quantitative relation of the effect on the surrogate to the clinical effect is rarely known."[31] Together these regulations recognize the importance of facilitating the development of, and access to, safe and effective treatment options for life-threatening and severely debilitating diseases with unmet medical needs. This approach has been reinforced by FDA's interactions with patients and their caregivers who describe their willingness to accept less certainty about effectiveness in return for earlier access to much needed medicines. For example, for a life-threatening disease without any available treatment, FDA might accept the results of adequate and well-controlled investigations with less rigorous designs, such as a historically controlled study. Below are considerations for drugs developed for life-threatening and severely debilitating diseases.

1. Trial design

While a randomized placebo-controlled trial can provide more definitive evidence of a small treatment effect than any other kind of trial of the same size, there are instances when this design and other concurrently controlled superiority designs may not be feasible or ethical. In such settings, other trial designs, such as non-inferiority trials or externally controlled trials can be acceptable if they provide substantial evidence of effectiveness (see discussion of noninferiority design and external control in Section III.A).

2. Trial endpoints

As discussed in Section III.B, endpoint selection is an important consideration in clinical trial design. The most straightforward and readily interpreted endpoints are those that directly measure clinical benefit or are validated surrogate endpoints shown to predict clinical benefit. Surrogate endpoints that are reasonably likely to predict clinical benefit can be relied on to establish effectiveness under the accelerated approval pathway. Effects on intermediate clinical endpoints can also be a basis for accelerated approval. Surrogate and intermediate clinical endpoints often can be assessed sooner than an endpoint that directly measures the clinical benefit or irreversible morbidity or mortality. Note that for accelerated approval the evidentiary standard still applies – that is, there must be substantial evidence that the drug has a meaningful effect on the surrogate or intermediate clinical endpoint.

[29] FDA guidance on expedited programs.

[30] 21 CFR 312.80, subpart E; 21 CFR 314.105(c).

[31] The preamble to the final rule on accelerated approval also notes, when responding to a comment, that "[a]lthough studies using surrogate endpoints may provide less assurance of clinical benefit than studies using clinical endpoints, FDA believes compliance with all of the elements of the accelerated approval program will not result in the marketing of large numbers of clinically ineffective drugs." 57 FR 58942, 58944 (December 11, 1992).

590 *3. Number of trials*
591
592 Although two adequate and well-controlled clinical investigations remain the standard approach
593 to generating substantial evidence of effectiveness in many disease settings, there are scenarios
594 where the conduct of a second trial is not ethical or feasible.
595
596 For example, as discussed in section IV.A.2, when a large multicenter trial has demonstrated a
597 clinically meaningful and statistically very persuasive effect on mortality, irreversible morbidity,
598 or prevention of a disease with potentially serious outcome, a second trial would be
599 impracticable or unethical. In this case the single large multicenter trial would be considered
600 sufficient to establish effectiveness.
601
602 *4. Statistical considerations*
603
604 A typical criterion for concluding that a trial is positive (showed an effect) is a p value of < 0.05
605 (two sided). A lower p value, for example, would often be expected for reliance on a single trial.
606 For a serious disease with no available therapy or a rare disease where sample size might be
607 limited, as discussed further below, a somewhat higher p value – if prespecified and
608 appropriately justified – might be acceptable.
609
610 **B. When the disease is rare**
611
612 By statutory definition, a rare disease – including a genetically defined subset of a disease –
613 affects fewer than 200,000 people in the U.S.;[32] but many rare diseases affect far fewer patients.
614 A large number of rare diseases are pediatric diseases or have childhood onset. In addition,
615 many rare disorders are life-threatening or severely debilitating diseases with no approved
616 treatments, leaving substantial unmet medical needs for patients. Therefore, many of the
617 considerations discussed above also apply to development programs for rare diseases.
618
619 FDA has a history of applying the philosophy underlying subpart E regulations to drugs for rare
620 diseases. FDA recognizes that certain aspects of drug development that are feasible for common
621 diseases may not be feasible for rare diseases and that development challenges are often greater
622 with increasing rarity of the disease. The small population affected by a rare disease presents
623 additional considerations that must be addressed and also calls for appropriate flexibility,
624 discussed below.
625
626 *1. Trial design*
627
628 Because of the small number of patients with a rare disease, the number of patients eligible for
629 enrollment in a trial may be small. In such situations, it is especially important to consider the
630 advantages and disadvantages of various trial designs to achieve the objectives of establishing
631 evidence of effectiveness as well as safety. Randomized, placebo-controlled trials with equal
632 allocation are generally the most efficient designs to assess effectiveness; however, depending on
633 the circumstances, sponsors should consider alternatives such as unequal allocation in a
634 randomized controlled trial (i.e., more patients receive the new drug than the control), which can

[32] Section 526(a)(2) of the FD&C Act (21 U.S.C. 360bb(a)(2)).

provide increased safety experience and reduce the use of placebo, or a dose-comparison design (i.e., randomization to more than one dose, with or without placebo). If the effect of the drug can be discerned relatively quickly after starting or discontinuing the drug, designs such as cross-over trials, randomized withdrawal, or randomized delayed start should also be considered. Sometimes, as noted previously, a single-arm trial with an external control is an appropriate option. The ability of these or other trial designs to generate substantial evidence of effectiveness is dependent on the specifics of each situation.

Sponsors of drugs intended for rare diseases should consider designing their first-in-human trial to be an adequate and well-controlled clinical investigation that has the potential, depending on the trial results, to provide part of the substantial evidence of effectiveness to support a marketing application.[33]

2. Trial endpoints

Understanding of the pathophysiology of the underlying disease is important in planning clinical trials, including selection of endpoints. For many rare diseases, well-characterized clinical efficacy endpoints appropriate for the disease may need to be developed. In cases where utilizing clinical endpoints is not feasible because changes in symptoms and disease status occur too slowly to be measured in a clinical trial of reasonable duration, surrogate endpoints may be considered. It will be particularly important to understand the pathophysiology and natural history of the disease to help identify potential surrogate endpoints.

3. Number of trials

A second trial may be infeasible in certain rare disease settings where the limited patient populations preclude the conduct of a second trial. A similar situation may also arise when a drug is developed to target, for example, a low-frequency, molecularly defined subset of a more common disease and it may not be possible to screen and enroll enough patients within a reasonable period of time to conduct the second trial.[34] In these cases, the substantial evidence of effectiveness would typically be provided by a single trial plus confirmatory evidence.

4. Statistical considerations

As noted above, treatments for rare diseases often are intended to address unmet medical needs, and the considerations of balancing the harmful consequences of false positive and false negative results will often apply. In addition, the amount of evidence that can practically be acquired may be limited by the number of patients who can be recruited for trials. FDA may interpret the substantial evidence standard flexibly considering the harmful consequences of false negative and false positive results and the amount of evidence that can practically be acquired. Statistical approaches to evaluating treatments for rare diseases should consider the feasibility of trial

[33] Draft guidance for industry *Human Gene Therapy for Rare Diseases* (July 2018). When final, this guidance will represent the Agency's thinking on the topic is addresses.

[34] Guidance for industry *Developing Targeted Therapies in Low-Frequency Molecular Subsets of a Disease* (October 2018).

design, sample size, and endpoints, using methods and thresholds for demonstrating substantial evidence that are appropriate to these settings.

C. When conducting a human efficacy trial is not ethical or feasible

When it is not ethical or feasible to conduct clinical trials, FDA can allow the use of appropriate animal models to generate evidence to establish effectiveness for products intended to treat or prevent serious or life-threatening conditions caused by exposure to toxic biological, chemical, radiological, or nuclear substances. FDA's regulation governing these trials is known as the Animal Rule.[35]

[35] The Animal Rule "applies to certain new drug products that have been studied for their safety and efficacy in ameliorating or preventing serious or life-threatening conditions caused by exposure to lethal or permanently disabling toxic biological, chemical, radiological, or nuclear substances." 21 CFR 314.600; see also 21 CFR 601.90 (same restriction with respect to biological products).