

I. INTRODUCTION

This document covers a range of topics on the appropriate use of meta-analysis of randomized controlled clinical trials (RCTs) to evaluate the safety of pharmaceutical products in the context of regulatory decision-making. Comments received from the public on this document will be considered in developing FDA guidance for applicants submitting new drug applications (NDAs), biologics licensing applications (BLAs), or supplemental applications on the appropriate use of meta-analyses of RCTs for safety evaluation. The planned guidance is also intended for FDA reviewers and for third-party entities that prepare or evaluate meta-analyses assessing the safety of regulated products, as there is currently no FDA guidance on how to evaluate the quality and persuasiveness of meta-analysis for regulatory decision-making. Specifically, this guidance will describe our view of various aspects of the evidentiary criteria considered important by FDA when evaluating the strength and quality of evidence provided by a meta-analysis.

FDA uses several approaches to summarize and combine safety data from multiple clinical studies, including formal meta-analysis, to assess the safety of drugs, either before approval or after marketing. Examples of the use of these strategies by FDA include:

1. Summary of observed safety events at the time of approval. The results are often displayed in tabular form in a product's labeling to describe common adverse events seen in the clinical trials conducted for approval. Regulations at 21 C.F.R. 314.50(d)(5)(iv) specifically call for such integrated analyses of safety.
2. Assessment of safety signals that arise from other analyses conducted by FDA. An example is the meta-analysis of the combined controlled studies' data conducted by FDA to assess a possible relationship between the use of "statins" and amyotrophic lateral sclerosis, a relationship suggested by individual reports of adverse events.
3. Assessment of safety signals raised by other investigators that require further investigation. An example is the meta-analysis conducted by FDA to investigate a possible link between rosiglitazone and cardiovascular morbidity/mortality.

These examples illustrate a critical feature of FDA's use of strategies to combine safety data, namely, the important consequences of these analyses when used to support a regulatory decision. This document focuses specifically on the use of formal meta-analysis of RCTs with safety endpoints. While many groups conduct meta-analyses, FDA's use of meta-analyses and other safety evaluation tools has the potential to result in consequential regulatory actions, including market withdrawal or a conclusion that a safety concern is not supported by data. FDA must therefore adopt a rigorous approach to these analyses and be transparent regarding its evidentiary standards and how it weighs the evidence of a meta-analysis in arriving at a decision or regulatory action. The planned guidance will set out the best practices for the conduct and assessment of meta-analyses in this setting.

II. INTRODUCTORY CONCEPTS AND PRINCIPLES

A. Scope and purpose of document

Meta-analysis is an important tool for safety assessment in the regulation of pharmaceutical products. The term *meta-analysis*, as used in this document, refers to the combining of evidence from relevant studies using appropriate statistical methods to allow inferences to be made to the population of interest. The most common reason for performing a meta-analysis is to provide an estimate of a treatment effect or measure of risk associated with a therapeutic intervention and to quantify the uncertainty about the estimated effect or risk when data from a single study are insufficient for this purpose, and the conduct of a new, large study would be impractical or take too long.

This document focuses specifically on meta-analyses conducted for purposes of safety evaluation using data from randomized controlled clinical trials (RCTs). Meta-analyses conducted to evaluate a product's effectiveness, either overall or within specific subgroups, are occasionally of interest to FDA, but the primary use of meta-analyses in the regulatory setting is for assessment of product risk. While meta-analyses of non-randomized studies may be informative for the assessment of certain safety endpoints, the issues related to such meta-analyses will not be addressed in this document.

Although methods for conducting and reporting meta-analyses are described in the medical and statistical literature, best practices for the design and conduct of a meta-analysis conducted for regulatory purposes, and interpretation and communication of their results, have not been widely discussed. This document raises issues that are critical to FDA's evaluation of the evidence provided by a meta-analysis and about which FDA is seeking input. It is apparent that, although we will focus on FDA's use of meta-analysis, development of best practices for the design and conduct of meta-analyses should be pertinent to all users.

FDA continually weighs the benefits and risks of drugs and therapeutic biologics to meet its mission to protect consumers and enhance public health. When safety concerns of significant public health impact are identified with the use of a regulated product, FDA must assess the evidence and decide the regulatory impact of such evidence. Such assessments can include meta-analyses that have been conducted in a variety of contexts, either by FDA or sponsors:

- (1) Meta-analyses may be conducted by sponsors and submitted to FDA as part of an IND, NDA, BLA or supplemental submission.
- (2) FDA may recommend that a sponsor or group of sponsors conduct a prospective meta-analysis, as it has in two recent guidance documents, FDA draft guidance for industry *Diabetes Mellitus: Developing Drugs and Therapeutic Biologics for Treatment and Prevention* (2008) and FDA guidance for industry and investigators *Safety Reporting Requirements for INDs and BA/BE Studies* (2010).
- (3) FDA may initiate its own meta-analysis in response to safety signals that FDA is aware of, or because FDA has access to study data that are otherwise unavailable to sponsors and researchers. Recent analyses include meta-analyses of suicidal thinking and behavior in randomized trials of anti-depressants and anti-epileptic drugs.

In addition to considering formal submissions of meta-analyses by drug or therapeutic biologic manufacturers, FDA also evaluates meta-analyses conducted by academic or other researchers based on published or otherwise available study results. In many such cases, the quality of the studies included (particularly if they have not been reviewed by FDA as part of a regulatory submission) and the quality of the meta-analysis itself may be unknown or undocumented. When the results raise concerns for the public health, FDA recognizes the importance of the agency's review and evaluation of the evidence provided by the published meta-analysis. Such a review poses particular challenges for FDA, in that it is usually not possible from the publication alone to determine the value of the evidence provided. Unlike randomized controlled clinical trials, for which objectives, hypotheses, and methods are almost always pre-specified prior to the conduct of the study (and the need for this pre-specification is well-established), meta-analyses are often 'data-driven,' based on studies already completed and for which results may be widely known. The extent to which prior knowledge of the studies' results can bias the meta-analysis findings is difficult to determine. Similarly, randomized controlled clinical trials are carefully designed to control the probability of making a Type I error (i.e., concluding a significant risk exists, when in fact, none does) by properly accounting for multiplicity arising from analyses of multiple endpoints or multiple subgroups of interest, or conducting multiple analyses for other reasons. Because it is often not apparent from a published meta-analysis how many different analyses were planned and conducted prior to the publication, the extent to which multiplicity can impair interpretation of results is also difficult to determine. Potential sources of bias and multiplicity are two critical issues that must be addressed before a regulatory decision can be made based on meta-analysis results, and we are seeking input on how best to address them.

To address these concerns, the planned guidance will provide a consistent framework for how meta-analyses should be designed, analyzed, reported, and interpreted in the context of product safety regulation for each of the above scenarios, namely, when FDA requires industry sponsors or other non-FDA entities to conduct a meta-analysis for submission and review, when FDA conducts its own meta-analysis, and when FDA evaluates an unrequested meta-analysis that is submitted by sponsors or published in the peer review literature. In each setting, there are preferred designs and reporting practices that FDA can apply to its own and to industry-requested meta-analyses, but they may pose problems for meta-analyses conducted by a third party. FDA is seeking input on whether the framework outlined in this document is sufficiently comprehensive, transparent, and clear to achieve the goals of a new guidance. Where more is needed, we ask for your advice and suggestions for improvements.

In the remaining subsections of this section, the reasons for the regulatory use of meta-analyses are described. Section III then summarizes the expectations of the FDA for meta-analyses intended for regulatory decision making and delineates the evidentiary criteria a meta-analysis must meet to merit consideration. A number of issues that may arise in planning a meta-analysis are described in this section for comment. Section IV poses several questions for which we are seeking input concerning best statistical approaches for conducting a safety meta-analysis.

B. Why meta-analysis of randomized controlled clinical trials is appropriate for safety assessments

Meta-analysis enables the estimation and quantification of risk for a single drug or a class of drugs in situations where single trials are not sufficient to assess risk. Unless a randomized controlled clinical trial is prospectively designed with a particular safety outcome as its primary endpoint, the trial may not have a large enough sample size to reliably evaluate whether there is increased risk of such events. This is because most serious drug-induced adverse events are rare or, if common, occur at only slightly increased rates compared to no treatment, and are not obviously drug-related (e.g., cardiovascular events, cancers). When more than one randomized controlled clinical trial is conducted, and the safety outcome results for the trials are available, meta-analyses can improve the ability to detect and characterize risks of serious adverse events that occur at low rates. It should be appreciated that a concern for both individual trials and meta-analysis is how well the safety event of interest was assessed, e.g., how suicidal thinking and behavior were assessed in trials of anti-depressants and anti-epileptic drugs.

Many safety concerns are not anticipated at the time studies are designed but rather are discovered later, and the question arises as to whether the occurrence is a chance finding or a true effect of drug exposure. Meta-analysis can help estimate with more precision than is available in a single trial the risk of a low-incidence safety outcome and may allow the examination of risk factors associated with the occurrence of the adverse event. However, many meta-analyses are conducted as exploratory exercises (without stated hypotheses) to determine whether a drug might be causing any untoward events. Such analyses are discovery-oriented and should usually be considered hypothesis generating, as no pre-specified safety concern motivated the conduct of the meta-analysis. FDA is interested in both types of meta-analyses. They are conducted for different but equally valid objectives; however, the weight given to each type may differ.

C. Other views of the use of meta-analysis as evidence

The planned guidance will provide scientific regulatory advice on how FDA uses meta-analysis to make regulatory decisions about the safety of new drugs and to describe the evidentiary principles that FDA considers or follows when making regulatory safety decisions that rely upon meta-analyses of randomized trials. As stated earlier, the guidance will be directed to industry, researchers and FDA staff, so that all are aware of the expectations and criteria against which meta-analyses will be judged and to encourage the use of best practices when planning, conducting and reporting meta-analyses to FDA.

There are other professional organizations, government agencies, and research consortia that have provided guidance on appropriate standards for conducting meta-analyses when used for their purposes. These include:

- Institute of Medicine (IOM) – ‘Finding What Works in Health Care: Standards for Systematic Reviews’ was issued in 2011, directed by Congress as part of the *Medicare Improvement for Patients and Providers Act of 2008*. The IOM report focuses on comparative effectiveness of medical and surgical interventions.

- Agency for Health Care Research and Quality (AHRQ) – ‘Methods Guide for Medical Test Reviews’ was developed as a practical guide for those preparing and using systematic reviews of medical tests and as a resource for the Evidence-based Practice Center Program.
- Cochrane Collaboration – ‘Handbook for Systematic Reviews of Interventions’ includes the PRISMA (Preferred Reporting Items for Systematic Reviews and Meta-Analyses) Statement and serves as a reporting guideline to assist authors of systematic reviews and/or meta-analyses of randomized trials and other study designs. The PRISMA Statement is an update and expansion of the QUOROM (Quality of Reporting of Meta-analyses) Statement, published ten years ago.
- Grading of Recommendations, Assessment, Development, and Evaluation (GRADE) working group – a series of articles published in the British Medical Journal in 2008 and in the Journal of Clinical Epidemiology in 2011 summarized the GRADE approach to rating quality of evidence and grading strength of recommendations in health care.

The above documents give guidelines for planning, conducting, and reporting results from meta-analyses or systematic reviews of research studies. FDA recognizes the benefits that these and other approaches lend to the general consensus for good practices in meta-analysis. The intent of FDA’s planned guidance is to emphasize methods and techniques that provide the rigor required in the evaluation of drug safety for regulatory decision-making.

III. REGULATORY CRITERIA FOR EVALUATING EVIDENCE FROM A META-ANALYSIS

In the following sub-sections, we review issues commonly associated with meta-analyses and describe the evidentiary standards used to evaluate meta-analyses conducted for regulatory decision-making.

A. Bias and multiplicity issues

Meta-analysis is typically a form of retrospective research in that most meta-analyses are conducted based on clinical trials that have been published or that, even if not published, have already been completed and whose results are known. As is well recognized, there is a potential for bias to occur from a variety of sources, including:

- Advance knowledge of individual study results when selecting the studies to be included in the meta-analysis
- Lack of pre-specification of the meta-analysis hypothesis
- Inclusion of the hypothesis-generating study in a meta-analysis designed to confirm the hypothesis
- Other biases that may exist but cannot be identified

In addition, the individual studies comprising the meta-analysis should be designed and conducted to ensure unbiased comparisons of the outcomes of interest. Because the studies may be completed well in advance of designing the meta-analysis, however, this may not be possible. The level of evidence that can be supported by a meta-analysis depends on the number and types

of biases present, both within the individual studies and as a result of the way the study-specific information is combined.

Another problem frequently encountered when evaluating the evidence provided by a meta-analysis is the potential for spurious findings because of multiple hypotheses being tested, multiple endpoints being evaluated, multiple or iterative analyses being conducted, and multiple patient subgroups being investigated. The result is often substantial inflation of the Type I error associated with the tests of hypotheses, making the meta-analysis conclusions difficult to interpret. When each of these sources of multiplicity is not well-described in advance, it is impossible to apply a statistical method of adjustment for multiplicity, because the full range of factors that were evaluated is impossible to determine post-hoc. When the analysis plan does contain a clear description of the sequence of tests to be conducted (across hypotheses, endpoints, subgroups, etc.), there may be too little power available for each of the tests to detect an important safety signal.

We are seeking recommendations on appropriate methods to guard against the bias and multiplicity issues described above as well as methods to evaluate their impact on the meta-analysis results. One possibility is to create an expectation for pre-specification of the meta-analysis protocol and statistical analysis plan prior to conducting the meta-analysis, such that people conducting these analyses would create such a protocol and store it in an accessible place. Input on an appropriate infrastructure that could be used for this purpose and that would be acceptable to researchers and industry is being sought.

B. The hierarchy of evidence from regulatory meta-analyses

The highest level of evidence derives from a meta-analysis that is prospectively planned and conducted, with the meta-analysis protocol prepared prior to the conduct of the trials to be included. Analogous to an adequate and well-controlled clinical trial, all aspects of the meta-analysis design, including the endpoints and analysis plan, are specified in advance without knowledge of the individual study results, and the availability of patient-level data from all studies is expected. In this scenario, a cumulative meta-analysis may be planned in which results are sequentially updated as each study concludes; however, such cumulative analyses require prospective statistical analysis plans that describe the methods to be used for repeated assessments of safety.

The second highest level of evidence is from a meta-analysis that is prospectively planned, with a protocol in place prior to the selection of studies or conduct of analyses, but where the studies are completed, and their results may be published or otherwise known to FDA and other interested parties. Although in some sense this is a retrospective meta-analysis, the prospectively developed protocol for the conduct of the meta-analysis can protect against some sources of bias associated with a fully retrospective one. If the protocol is developed without examining study results, then one of the major sources of bias, namely, knowing the results of some or all of the component studies in advance, is avoided. In addition, it should be possible to standardize the outcomes and the study selection criteria, thereby better controlling for biases in study selection and choice of endpoint. By pre-specifying the study endpoints and statistical analysis plan, there is little opportunity for manipulating the meta-analysis conclusions based on judgments made

along the way by the analyst, often not well documented in published meta-analyses. Note, however, that even when the meta-analysis is carried out as pre-specified in the protocol and analysis plan, thereby minimizing the potential for bias, the limitations of the actual data collected and the study results already reported may lessen the effectiveness of the pre-specification. These limitations will be taken into account in evaluating the weight of evidence from the meta-analysis.

From a regulatory review perspective, an important aspect of either meta-analysis described above that contributes substantially to the quality and strength of evidence is the availability of patient-level data for each study in the meta-analysis. Such data availability allows FDA to evaluate each study's quality and eligibility for inclusion in the meta-analysis and also allows for confirmation of study outcomes, particularly time-to-event outcomes. The consistency of outcome ascertainment across studies can also be assessed. Analysis strategies can control for and evaluate the impact of patient-level covariates in addition to study-level covariates on risk and risk factors. Lack of patient-level data forces the meta-analysis to be based on study-level summary data, thereby limiting the ability to evaluate the robustness of its conclusions. The selection criteria for studies to be included in the meta-analysis could be related to the availability of patient-level data, but such a determination should be prospective.

The level of evidence from a meta-analysis that is based solely on study-level summary data, either prospective or retrospective, is generally considered to be lower. In this scenario, study-specific treatment effects (i.e., risk differences, relative risks, odds ratios, or hazard ratios) are the input into the meta-analyses, and no patient-level data are available for quality control or analysis. While it may be the case that a study-level meta-analysis and a patient-level meta-analysis yield similar results, it is difficult to identify such cases without the availability of the patient-level data. A hybrid would be a combination of studies for which patient-level data are available for a subset; the mix would determine where in the hierarchy such a meta-analysis is placed.

The prospective nature of the meta-analysis and availability of patient-level data are but two dimensions in the hierarchy of evidence. In determining where in the hierarchy a particular meta-analysis lies, consideration should also be given to the knowledge and documentation of eligible studies, both published and unpublished; the quality and quantity of the studies selected as well as the process and timing of selection; and the validity of the statistical analysis that supports the inferential conclusions, evaluated against all sources of potential or real bias.

To supplement this hierarchy, there are specific criteria that are used to evaluate the level of evidence represented by a meta-analysis. These criteria are not unique to FDA's perspective on evidence derived from a meta-analysis. As described in Section II.C, criteria for evaluating the weight of evidence of meta-analyses are available from other published sources.

FDA considers the following when evaluating a meta-analysis; this list should not be viewed as exhaustive:

- Whether the meta-analysis (purpose, methods, endpoints, inclusion criteria) was prospectively planned

- The level, quality, and availability of study data from the clinical trials included in the meta-analysis
- The quality of the trials relative to the purpose of the meta-analysis
- The quality of the meta-analysis, including its protocol, statistical analysis methods, including accounting for multiplicity, potential for bias, and other limitations
- The strength of evidence of the meta-analysis, including the magnitude of the risk, the probability that the meta-analytic finding is due to chance, and a rigorous assessment of other plausible explanations for the results
- The adequacy of the documentation of the meta-analysis, allowing for independent confirmation of the results
- The biological and clinical plausibility of the finding, including evidence of a dose versus risk relationship
- The homogeneity or heterogeneity of risks among the studies included in the meta-analysis and the generalizability of the findings
- The robustness of results to other selection criteria and/or judgments bearing upon the inclusion or exclusion of eligible studies in the meta-analysis

FDA anticipates addressing all of these factors in the planned guidance. Some of them are discussed in more detail in the following Sections.

C. The prospective plan for the meta-analysis

As discussed above, for any given finding, a prospectively planned meta-analysis will always be viewed as providing stronger evidence than one conducted with no such plan in place. First and foremost in the plan should be the stated purpose. Of relevance to the setting considered here are purposes relating to evaluation of safety, including but not limited to the following:

- To estimate a specific risk more precisely than can be accomplished with data from a single trial, e.g., a general concern for a drug class or a risk suggested by the results of a single trial, an epidemiologic observation, or spontaneous reports of adverse events
- To address publication and/or selection bias in estimating risk by synthesizing findings from all relevant trials
- To attempt to resolve apparent conflicting results about risk among several trials
- To evaluate risk in one or more subgroups of patients, when sufficient sample size for this purpose is not available in a single trial
- To examine dose-risk relationships using data from more than one trial
- To assess accumulating evidence on product safety, as ongoing studies are completed
- To examine whether risk estimated from trials conducted at different times changes over time
- To identify factors (covariates) that are associated with risk using data from more than one trial
- To assess risk across studies of a class of products
- To provide evidence for a benefit-risk assessment when updating risk communication and risk management for a product or class of products

The enhanced persuasiveness of a meta-analysis with a specified purpose and analytic plan cannot be overstated. A clearly stated purpose and plan in the protocol removes concerns about inflation of type one error due to multiplicity or data-driven analyses that modify or change one or more of the initially specified hypotheses.

In addition to the stated purpose, other design choices should be described as completely as possible in the meta-analysis protocol. Any meta-analysis involves many choices, e.g., of endpoints, study inclusion criteria, comparator, measure of risk, etc., and it is often not possible to know when and how such choices were made. Without a prospective plan, it is difficult to determine which analyses were planned and which were exploratory or suggested during the conduct of the meta-analysis. As a result, the ability to draw a conclusion or assess the statistical uncertainty of the findings is difficult.

D. The quality of the individual studies

FDA will evaluate whether the quality of the individual study data used in a meta-analysis is sufficient to support the meta-analysis conclusions. As mentioned earlier, the availability and use of patient-level data versus study-level data increases the value of evidence of a meta-analysis in most cases. In addition, trials designed with the intent of assessing a specific safety hypothesis will have the advantage that outcome specification, collection, monitoring, and adjudication, if appropriate, were planned in advance, and both quality and consistency of the safety data being analyzed should therefore be enhanced when compared to safety data from trials designed to support efficacy hypotheses. Studies planned with a primary goal of providing evidence of efficacy may also include data on safety outcomes that are of sufficient quality and consistency to support a safety meta-analysis, but verification will be retrospective and possibly more difficult. In these cases, study conduct issues may affect the quality and reliability of safety outcomes that are attributed to patients in the study. For example, the duration of patient exposure, the timing of withdrawal from assigned study drug, and the reasons for such early termination are all critical data elements that govern the rules for assigning a serious adverse event to the appropriate treatment arm, particularly if the event occurred post exposure. Some studies are able to collect such data, and others are not; the ability to correctly make this attribution depends wholly on the quality of the data collected.

The procedures used for randomization and concealment of treatment assignments, the study discontinuation rate, and the amount of missing data are important factors in determining the quality of an individual study. Open-label studies, even when randomized, will generally add less value to a meta-analysis than their double-blind counterparts, and studies with unexpectedly high rates of early termination or large amounts of missing data will provide less value than well-executed studies with better follow-up rates. It should be clear, however, that decisions about the standard for including or omitting the studies should not be based on the results of the studies.

E. Statistical persuasiveness of evidence from a meta-analysis

There are several statistical approaches to combining and analyzing the results of individual randomized controlled clinical trials for purposes of assessing risk. All methods provide an estimate of the overall risk, a test of hypothesis about the risk, and the associated p-value

indicating the strength of evidence against the null hypothesis that there is no increased risk. Several authors have proposed that because of the potential for many inherent biases in meta-analyses, the commonly used test size or alpha level for hypothesis testing, 0.05, should be replaced by a lower value, e.g., 0.01 or 0.001, for the results to be considered convincing [Pogue and Yusuf 1997] [Flather, Farkouh, Pogue, Yusuf, 1997]. The retrospective nature of most meta-analyses and the multiple comparisons that are carried out, often after observing the statistical significance of the individual studies' findings, support the more stringent control of Type I error that a lower alpha level provides. If the meta-analysis will be the sole basis for decision-making, especially when there is no possibility of a future trial to confirm the findings, such a low value for alpha would be warranted, even when other reasons do not apply. Multiplicity and bias concerns can be mitigated, however, by careful planning and pre-specification, as discussed in a previous section.

Many meta-analyses are conducted in a sequential manner. Study results are combined and analyzed as each study is completed. Such practice essentially presents the same issues for statistical Type 1 error control as repeated statistical hypothesis testing from multiple interim analyses of an ongoing trial. A sequential meta-analysis usually requires a lower alpha level for testing, particularly when there is no pre-specified limit to the number of studies that can or will be combined.

Arguments for more stringent control of Type I error, although supportable by the post-facto nature of many meta-analysis procedures, must also be weighed against the importance of the safety finding. The potential for harm may be so serious that even marginally significant findings may support a regulatory action.

IV. BEST PRACTICES

The quality of the meta-analysis is in large part dependent on the appropriateness of the statistical methods used to conduct the analysis. Although there is a large literature and existing guidelines on meta-analysis, as stated earlier, there is a lack of consensus on the best statistical practice in a number of areas. Recognizing that there may be no single method that performs best in all situations, the planned guidance will underscore the need for the statistical methods used in a meta-analysis to be appropriate for the hypotheses under investigation and consistent with the individual studies' designs and distributions. In preparation of the guidance, FDA is exploring several issues related to meta-analysis for the evaluation of risk in the regulatory framework, including:

- Evaluating the results of a meta-analysis, when one or a few large studies dominate the findings
- The ascertainment of outcome and exposure in the individual studies included in the meta-analysis
- The performance of meta-analytic methods with rare events
- The use of fixed effects versus random effects models in evaluating a meta-analytic hypothesis, especially with regard to power and generalizability
- The use of frequentist versus Bayesian methods for meta-analysis

- The use of sensitivity analyses to evaluate the potential sources of bias outlined in Section III.A and how to address conflicting findings that may result from these analyses
- Reporting meta-analyses results

These issues are discussed further in the remainder of this section.

Large studies that dominate meta-analysis findings: A meta-analysis may be greatly influenced by the inclusion of one or a few large trials, and there is no currently accepted practice on how best to evaluate the findings when this is the case. One possibility is to design the meta-analysis to exclude trials that are extremely large relative to the other studies. With this approach, the meta-analysis would summarize the results from the smaller trials, and the large trials could then be used to contrast with or confirm the meta-analysis results. An alternative approach is to include the large trials in the primary meta-analysis and then conduct sensitivity analyses to explore the impact their inclusion had on the meta-analysis findings. FDA is seeking input on these and other approaches to evaluating results in the presence of an overly influential study.

Ascertainment of outcomes and exposure: The choice of the safety endpoint should be carefully and prospectively considered when designing a meta-analysis. Differences may exist in endpoint ascertainment from study to study that can induce bias in the meta-analysis results. Even in the presence of randomization, differences in outcome ascertainment may exist among treatment arms of a single study that can induce bias. For example, if a side effect of an investigational treatment is unrelated to the outcome of interest but results in additional health care visits, there will be more opportunity to detect the side effect for patients in the investigational arm during those visits. The use of “hard outcomes” such as mortality may offer some protection against this phenomenon. For example, vital status may be able to be obtained for the entire meta-analysis population from a national death index search, eliminating any bias due to differential outcome ascertainment. Similarly, stroke and myocardial infarction may be ascertained more consistently and with less bias than other cardiovascular events, such as arrhythmia and transient ischemic attacks.

Exposure duration may also differ across studies and among comparator arms within individual studies, creating another potential for bias in meta-analysis results. FDA is considering design and analysis methods that would address both sources of bias and is seeking input on the choice of appropriate methods. At a minimum, differential outcome ascertainment and differential exposure duration should be evaluated for the studies included in the meta-analysis to determine if the potential for bias exists, preferably before the meta-analysis is conducted.

Rare events: We note that there are special statistical considerations for meta-analysis methods for the evaluation of safety. This is primarily because safety outcomes may be infrequent, even if the outcomes would have strong bearing on the risk-benefit consideration of a drug. Individual trials may have few or no outcome events. We refer to this situation as sparse data. With sparse data, statistical estimates may be unstable in that they have large variance, and the variance may be difficult to measure. Traditional meta-analysis methods developed for more frequently occurring outcomes may not be appropriate for meta-analysis for the evaluation of safety. FDA is exploring methods appropriate for sparse data and seeking input on the choice of available methods.

Fixed versus random effects models: An important consideration in the choice of meta-analysis model is the use of fixed effects versus random effects models. It is generally stated that a fixed effects meta-analysis model assumes all studies are estimating a common treatment effect or risk. That is, there is no between-study heterogeneity in the true effect or risk, and any variation in the observed study-specific estimates is due to chance differences arising from patient-to-patient variability. In contrast, a random effects meta-analysis model assumes the observed estimates of effects or risks can vary across studies because of real differences in the study-specific effects as well as chance variation. In practice, it is often clear that there are differences in trial characteristics such that differences in the trial-specific effects can be expected.

It is also noted that fixed effects and random effects meta-analyses can be thought of as answering two different questions. The fixed effects model provides an estimate of the average treatment effect or risk among the trials in the meta-analysis, whereas the random effects model assumes the trials in the meta-analysis are a random sample from a population of trials that are of interest to the research question and uses the sample to estimate the population-average effect. The choice of fixed effects versus random effects meta-analysis models warrants further consideration, and FDA is seeking input prior to making a recommendation.

Frequentist versus Bayesian methods of meta-analysis: Bayesian methods provide a broad and flexible approach to meta-analysis, and their use has been advocated for evaluating safety issues of marketed pharmaceutical products in a recent IOM Report (*Ethical and Scientific Issues in Studying the Safety of Approved Drugs, released May 1, 2012*). Their advantages include the ability to incorporate a random effects structure or additional sources of variation into the meta-analysis model and to incorporate information on trials with no events, even when risk is assessed via relative measures such as the odds ratio. However, at this point, their performance and robustness are not fully evaluated in the context of safety meta-analyses to support regulatory decisions. FDA seeks input on the use of Bayesian models for safety meta-analysis and their robustness to different model specifications and prior distributions.

Sensitivity analyses: Sensitivity analyses are vital to understanding and interpreting meta-analysis findings. The impact of the various design choices and analysis strategies selected in designing the meta-analysis can be explored through sensitivity analyses, as can the impact of potential sources of bias, particularly biases that are unavoidable. In fact, the primary purpose of most sensitivity analyses conducted in relation to a meta-analysis is to assess the impact of potential biases. FDA is interested in identifying sensitivity analyses that are best able to explore the potential biases that may be in play for a particular meta-analysis as well as to support the findings of the meta-analysis and to gauge the strength of those findings, and is seeking input on this topic.

Reporting: Completeness and transparency are key objectives in the reporting of a meta-analysis for the evaluation of safety in the regulatory framework. A protocol describing the design of the meta-analysis is expected, as is a report describing the conduct of the analysis and any sensitivity analyses in addition to presenting the findings. The protocol and study report should describe what information was available prior to designing the meta-analysis and what specific information, including individual study results, motivated the research objectives of the meta-analysis. Potential problems anticipated in designing the meta-analysis and methods

547 incorporated to manage those problems as well as sensitivity analyses designed to evaluate their
548 impact, if they are not manageable, should also be documented. The protocol should be
549 completed and finalized prior to the conduct of the meta-analysis. FDA is interested in a suitable
550 framework for making meta-analysis protocols and reports publicly available to the extent
551 permissible under applicable laws and regulations. Such a framework would be particularly
552 useful if its use provided some assurances that the protocol was complete prior to conducting the
553 meta-analysis.