Design Considerations for Pivotal Clinical Investigations for Medical Devices

---

Guidance for Industry, Clinical Investigators, Institutional Review Boards and Food and Drug Administration Staff

Document issued on: November 7, 2013

The draft of this document was issued on August 15, 2011.

For questions regarding this document that relate to devices regulated by CDRH, contact Gregory Campbell, PhD at (301) 796-5750 or by email at greg.campbell@fda.hhs.gov, if desired.

For questions regarding this document that relate to devices regulated by CBER, contact Stephen Ripley at 301-827-6210.



U.S. Department of Health and Human Services
Food and Drug Administration
Center for Devices and Radiological Health
Center for Biologic Evaluation and Research

GRZAN

Preface

Public Comment

Written comments and suggestions may be submitted at any time for Agency consideration to the Division of Dockets Management, Food and Drug Administration, 5630 Fishers Lane, Room 1061, (HFA-305), Rockville, MD, 20852.

Additional Copies

Additional copies are available from the Internet. You may also send an e-mail request to dsmica@fda.hhs.gov to receive an electronic copy of the guidance or send a fax request to 301-847-8419 to receive a hard copy. Please use the document number 1776 to identify the guidance you are requesting.
Additional copies of this guidance document are also available from:

Center for Biologics Evaluation and Research (CBER),
Office of Communication, Outreach and Development (HFM-40),
1401 Rockville Pike, Suite 200N, Rockville, MD 20852-1448,
or by calling 1-800-835-4709 or 301-827-1800, or email ocod@fda.hhs.gov, or from the Internet at
http://www.fda.gov/BiologicsBloodVaccines/GuidanceComplianceRegulatoryInformation/Guidances/default.htm.

Table of Contents

1 INTRODUCTION 4
2 SCOPE 5
2.1 Types of Studies Addressed in this Guidance 5
2.2 Types of Studies Not Addressed in this Guidance 6
3 REGULATORY FRAMEWORK FOR LEVEL OF EVIDENCE AND STUDY DESIGN 7
3.1 The Statutory Standard for Approval of a PMA: Reasonable Assurance of Safety and Effectiveness 8
3.2 Valid Scientific Evidence 9
3.3 Benefit-Risk Assessment 10
3.4 Clinical Study Level of Evidence and Regulation 11
3.5 The Least Burdensome Concept and Principles of Study Design 12
4 TYPES OF MEDICAL DEVICES 13
4.1 Types of Devices Based on Intended Use 14
4.2 Special Considerations for Clinical Studies of Devices 15
5 THE IMPORTANCE OF EXPLORATORY STUDIES IN PIVOTAL STUDY DESIGN 16
6 SOME PRINCIPLES FOR THE CHOICE OF CLINICAL STUDY DESIGN 17
6.1 Types of Studies 18
6.2 General Considerations: Bias and Variability in Device Performance 19
6.3 Study Objectives 19
6.4 Subject Selection 20
6.6 Site Selection 22
6.7 Comparative Study Designs 23
7 CLINICAL OUTCOME STUDIES 24
7.1 Endpoints in Clinical Studies 24
7.2 Intervention Assignment (Randomization) for Clinical Outcome Studies 26
7.3 Blinding (Masking) 28
7.4 Controls in Comparative Clinical Outcome Studies 29
7.5 Placebo Effect and Other Phenomena 31
7.6 Non-Comparative Clinical Outcome Studies 32
7.7 Diagnostic Clinical Outcome Studies 34
7.8 Advantages and Disadvantages of Some Clinical Outcome Studies 35
7.9 Some Regulatory Considerations 38
8 DIAGNOSTIC CLINICAL PERFORMANCE STUDIES 38
8.1 Consideration of Intended Use 39
8.2 Clinical Reference Standard for the Target Condition 40
8.3 Study Population for Evaluation of Diagnostic Performance 41
8.4 Study Planning, Subject Selection and Specimen Collection 42
8.5 Diagnostic Clinical Performance Comparison Studies 43
8.6 Blinding (Masking) in Diagnostic Performance Studies 44
8.7 Skill and Behavior of Persons Interacting with the Device (Total Test Concept) 44
8.8 Common Types of Bias in Diagnostic Clinical Performance Studies 45
9 SUSTAINING THE QUALITY OF CLINICAL STUDIES 46
9.1 Handling Clinical Data 46
9.2 Study Conduct 47
9.3 Study Analysis 49
9.4 Anticipating Changes to the Pivotal Study 50
10 THE PROTOCOL 50
11 GLOSSARY 51

Design Considerations for Pivotal Clinical Investigations for Medical Devices

Guidance for Industry, Clinical Investigators, and Food and Drug Administration Staff

This guidance represents the Food and Drug Administration's (FDA's) current thinking on this topic. It does not create or confer any rights for or on any person and does not operate to bind FDA or the public. You can use an alternative approach if the approach satisfies the requirements of the applicable statutes and regulations. If you want to discuss an alternative approach, contact the FDA staff responsible for implementing this guidance. If you cannot identify the appropriate FDA staff, call the appropriate number listed on the title page of this guidance.

1 Introduction

This document is intended to provide guidance to those involved in designing clinical studies intended to support pre-market submissions for medical devices and FDA staff who review those submissions. Although the Agency has articulated policies related to design of studies intended to support specific device types, and a general policy of tailoring the evidentiary burden to the regulatory requirement, the Agency has not attempted to describe the different clinical study designs that may be appropriate to support a device pre-market submission, or to define how a sponsor should decide which pivotal clinical study design should be used to support a submission for a particular device. This guidance document describes different study design principles relevant to the development of medical device clinical studies that can be used to fulfill pre-market clinical data requirements. This guidance is not intended to provide a comprehensive tutorial on the best clinical and statistical practices for investigational medical device studies.

Medical devices can undergo three general stages of clinical development. These stages may be extremely dependent on each other and doing a thorough evaluation in one stage can make the next stage much more straightforward. To begin, medical devices may undergo an exploratory clinical stage. In this stage, the limitations and advantages of the medical device are evaluated. This stage includes first-in-human studies and feasibility studies. The next stage, the pivotal stage, is used to develop the information necessary to evaluate the safety and effectiveness of the device for the identified intended use. It usually consists of one or more pivotal studies. Finally, devices undergo a post-market stage which can include an additional study or studies for better understanding of device safety, such as rare adverse events and long-term effectiveness. This guidance provides information on design issues related to pivotal clinical investigations and does not address the other stages in any detail.

A medical device pivotal study is a definitive study in which evidence is gathered to support the safety and effectiveness evaluation of the medical device for its intended use. Evidence from one or more pivotal clinical studies generally serves as the primary basis for the determination of reasonable assurance of safety and effectiveness of the medical device of a pre-market approval application (PMA) and FDA's overall benefit-risk determination. In some cases, a PMA may include multiple studies designed to answer different scientific questions.

The focus of this guidance is providing recommendations to sponsors on how to design clinical investigations to support a PMA. However, sponsors who conduct clinical studies to support pre-market notification (510(k)) and de novo submissions may also rely on the principles in this guidance document.

FDA's guidance documents, including this guidance, 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.

2 Scope

This guidance describes principles for the design of pre-market clinical studies[1] that are pivotal in establishing the safety and effectiveness of a medical device. Practical issues and pitfalls in pivotal clinical study design are discussed, along with their effects on the conclusions that can be drawn from the studies concerning safety and effectiveness. The principles discussed in this guidance are intended to assist sponsors of marketing applications and investigators in designing studies adequate to provide a reasonable assurance of safety and effectiveness concerning a device.

2.1 Types of Studies Addressed in this Guidance

Due to the range of intended uses and risks associated with medical devices and constraints in executing clinical studies, this guidance treats pivotal clinical studies in a general manner. It frames FDA's recommendations in terms of two broad categories of medical devices:

- Therapeutic and aesthetic devices
- Diagnostic devices

From this guidance, device developers can gain insight about important pivotal study design issues for devices in each of these categories. At the same time, communication with FDA review staff (e.g., through a pre-submission interaction) is often valuable in arriving at pivotal clinical study designs that are both practical and adequate.

[1] For purposes of this guidance the term "studies" is equivalent to the term "investigations."

This guidance also includes principles that are applicable to the device-specific issues for combination products defined under 21 CFR Part 3 (e.g., device-drug products; device-biologic products). However, drug-specific or biologic-specific issues that may also be relevant for a combination product are not described in this guidance.

This guidance is intended to complement other existing guidance, and is not intended to replace the policies described in other guidance documents. In cases where questions arise, consult the appropriate FDA review division directly or the Center for Devices and Radiological Health (CDRH) Division of Small Manufacturers, International and Consumer Assistance and Consumer Assistance or the Center for Biologics Evaluation and Research (CBER) Office of Communication, Outreach and Development (OCOD) depending on which Center is responsible for review of the device.

2.2 Types of Studies Not Addressed in this Guidance

Although this guidance does not address the following kinds of studies, some principles discussed herein are applicable to many of them:

- Non-clinical studies (e.g., bench, animal or measurement studies and, for *in vitro* diagnostic devices, analytical validation studies);
- Studies intended to support Humanitarian Device Exemption (HDE) applications;[2]
- Pre-market feasibility clinical studies, or other pre-market clinical studies that are not part of the pivotal stage;
- Studies to establish the clinical validity of companion diagnostic devices (i.e., *in vitro* diagnostic tests that provide essential information for the safe and effective use of a corresponding therapeutic product). Clinical development programs for companion diagnostic devices are typically part of the clinical development programs of the corresponding therapeutic products;
- Post-market clinical studies. Though the need for post-market clinical studies might arise from interpretation of pre-market clinical results, post-market studies do not drive the initial determination of safety and effectiveness, and their design is not addressed in this guidance. However, the principles discussed in this guidance may be useful in designing such studies;
- Studies of products regulated by CBER that require an Investigational New Drug application and Biologics License Application, such as donor screening tests, are not included in the scope of the guidance.

[2] See FDA's Guidance for HDE Holders, Institutional Review Boards (IRBs), Clinical Investigators, and FDA Staff - Humanitarian Device Exemption (HDE) Regulation: Questions and Answers, at http://www.fda.gov/MedicalDevices/DeviceRegulationandGuidance/GuidanceDocuments/ucm110194.htm, for detailed information.

Although this guidance is developed primarily for clinical studies used to support PMAs, the recommendations of this guidance may also be used in designing clinical studies used to support some 510(k) and de novo submissions with clinical data when applicable.

3 Regulatory Framework for Level of Evidence and Study Design

Clinical studies of medical devices must conform to certain legal requirements. This section describes the:

- Regulatory framework applicable to the design of clinical studies that support pre-market submissions for medical devices;
- Statutory standard for approval of a PMA;
- Regulatory requirements that apply to clinical and other data used to meet the statutory standard for approval of a PMA;
- How FDA evaluates the data to assess the risks and benefits of a device;
- Basic information about Investigational Device Exemption (IDE) applications; and
- FDA's current thinking on good regulatory practice as identified in the least burdensome concept.

This guidance reflects the Agency's consideration of standards for designing, conducting, recording, and reporting studies that involve the participation of human subjects. Related international documents include the "Declaration of Helsinki" and are further explained in the *International Standards Organization (ISO) 14155:2011, Clinical investigation of medical devices for human subjects - Good clinical practice* and through the International Conference on Harmonisation (ICH) of Technical Requirements for Registration of Pharmaceuticals for Human Use (ICH) *E6 Good Clinical Practice: Consolidated Guidance*. FDA regulations under 21 CFR Parts 50, 54, 56, and 812 articulate good clinical practice (GCP) requirements applicable to clinical investigations of medical devices. In addition, FDA guidance documents describe FDA's current thinking on GCP and the conduct of clinical studies.[3] Compliance with GCP, as applicable to medical devices, protects the rights, safety, and well-being of human subjects, ensures appropriate scientific conduct of the clinical investigation and the credibility of the results, defines the responsibilities of the sponsor and the clinical investigator, and assists sponsors, investigators, institutional review boards (IRBs), other ethics committees, regulatory authorities, and other bodies involved in the development and review of medical devices.

If a clinical investigation of a device is conducted in the United States it must comply with the applicable regulations found in 21 CFR Part 56 (IRBs), Part 50 (informed consent), and Part 812 (investigation device exemption (IDE)). If the clinical investigation of the device is conducted

[3] A list of FDA's GCP guidance documents is available at http://www.fda.gov/RegulatoryInformation/Guidances/ucm122046.htm.

outside of the United States, submitted in support of a PMA, and conducted under an IDE, the study shall comply with 21 CFR Part 812. 21 CFR 814.15(a). If a clinical investigation of a device is conducted completely outside the United States and not conducted under an IDE, FDA will accept studies submitted in support of a PMA, if the data are valid and the investigator has conducted the studies in conformance with the "Declaration of Helsinki" or the laws and regulations of the country in which the research is conducted, whichever accords greater protection to the human subjects, in accordance with 21 CFR 814.15. A PMA based solely on foreign clinical data may be approved if the foreign data are applicable to the United States population and medical practice; the studies have been performed by clinical investigators of recognized competence; if an inspection is needed, FDA can validate the data through an on-site inspection or other appropriate means; and the application otherwise meets the criteria for approval. 21 CFR 814.15(d). The sponsor must be able to show that the foreign clinical data are adequate to support approval of the device in the United States, under applicable standards. Section 569B of the Federal Food, Drug, and Cosmetic Act (FD&C Act or the Act). We encourage you to meet with FDA in a Pre-Submission meeting if you intend to seek approval based on foreign data, or based on a combination of foreign and U.S. data.

3.1 The Statutory Standard for Approval of a PMA: Reasonable Assurance of Safety and Effectiveness

As indicated by Sections 513(a)(1)(C) of the FD&C Act, a PMA must provide reasonable assurance of safety and effectiveness of the device. FD&C Act Section 513(a)(2) states:

> *[T]he safety and effectiveness of a device are to be determined---*
> *(A) with respect to the persons for whose use the device is represented or intended,*
> *(B) with respect to the conditions of use prescribed, recommended, or suggested in the labeling of the device, and*
> *(C) weighing any probable benefit to health from the use of the device against any probable risk of injury or illness from such use.*

In addition, FDA has, through regulation, interpreted the statutory standard for approval of a PMA as follows:

> *21 CFR 860.7(d)(1). There is reasonable assurance that a device is safe when it can be determined, based upon valid scientific evidence, that the probable benefits to health from use of the device for its intended uses and conditions of use, when accompanied by adequate directions and warnings against unsafe use, outweigh any probable risks. The valid scientific evidence used to determine the safety of a device shall adequately demonstrate the absence of unreasonable risk of illness or injury associated with the use of the device for its intended uses and conditions of use.*

> *21 CFR 860.7(e)(1). There is reasonable assurance that a device is effective when it can be determined, based upon valid scientific evidence, that in a significant portion of the target population, the use of the device for its intended uses and conditions of use, when accompanied by adequate directions for use and warnings against unsafe use, will provide clinically significant results.*

These statutory and regulatory provisions specify that a finding of reasonable assurance of safety and effectiveness must be supported by data relevant to the target population, and evaluated in light of the device labeling. Further, a determination of whether the standard of approval for a PMA has been met is based on balancing probable benefit to health with probable risk.

3.2 Valid Scientific Evidence

The regulations state that the safety and effectiveness of a device will be determined on the basis of valid scientific evidence. 21 CFR 860.7(c)(1). Valid scientific evidence is defined through regulation as follows:

> *21 CFR 860.7(c)(2) Valid scientific evidence is evidence from well-controlled investigations, partially controlled studies, studies and objective trials without matched controls, well-documented case histories conducted by qualified experts, and reports of significant human experience with a marketed device, from which it can fairly and responsibly be concluded by qualified experts that there is reasonable assurance of the safety and effectiveness of a device under its conditions of use. The evidence required may vary according to the characteristics of the device, its conditions of use, the existence and adequacy of warnings and other restrictions, and the extent of experience with its use. Isolated case reports, random experience, reports lacking sufficient details to permit scientific evaluation, and unsubstantiated opinions are not regarded as valid scientific evidence to show safety or effectiveness. Such information may be considered, however, in identifying a device the safety and effectiveness of which is questionable.*

FDA regulations also consider which types of evidence support reasonable assurance of safety and effectiveness:

> *21 CFR 860.7(d)(2) Among the types of evidence that may be required, when appropriate, to determine that there is reasonable assurance that a device is safe are investigations using laboratory animals, investigations involving human subjects, and nonclinical investigations including in vitro studies.*
>
> *21 CFR 860.7(e)(2) The valid scientific evidence used to determine the effectiveness of a device shall consist principally of well-controlled investigations, as defined in [21 CFR 860.7](f), unless [FDA] authorizes reliance upon other valid scientific evidence which [FDA] has determined is sufficient evidence from which to determine the effectiveness of a device, even in the absence of well-controlled investigations. [FDA] may make such a determination where the requirement of well-controlled investigations in [21 CFR 860.7](f) is not reasonably applicable to the device.*

Thus, one key principle evident in 21 CFR 860.7 is that evidence of effectiveness of a medical device must generally be obtained from well-controlled studies (as described in 21 CFR 860.7(f)). However, the regulations provide FDA with some flexibility regarding its determination of the type of evidence that may be considered valid scientific evidence to demonstrate the safety of a medical device.

FDA believes that in most cases, clinical data will be necessary to demonstrate effectiveness for a device being reviewed in an original PMA. Sections 6, 7, and 8 of this guidance provide some principles to help sponsors determine an appropriate study design. The results of the study must provide sufficient evidence for FDA to make a determination of reasonable assurance of safety and effectiveness, as defined in the regulations above. Based on pre-submission discussions with the sponsor, FDA may determine that alternative study designs can yield appropriate data on which the Agency can make a determination of safety and effectiveness.

Even with a well-planned design, the study may not yield the results expected or necessary to demonstrate safety and effectiveness. The sponsor may need to reassess its goals for the medical device and conduct additional studies to obtain evidence necessary to demonstrate safety and effectiveness.

3.3 Benefit-Risk Assessment

FDA considers a number of factors when making benefit-risk determinations during the pre-market review process. FDA considers the factors within the intended use of the device, including the target population. [4]

Under Section 513(a) of the FD&C Act FDA determines whether PMA applications provide a "reasonable assurance of safety and effectiveness" by "weighing any probable benefit to health from the use of the device against any probable risk of injury or illness from such use," among other relevant factors. 21 CFR 860.7(b)(3) states that, in determining the safety and effectiveness of a device, FDA must weigh "the probable benefit to health from the use of the device against any probable injury or illness from such use."

FDA relies on valid scientific evidence to determine whether there is reasonable assurance that the device is safe and effective. 21 CFR 860.7(c)(1). There is reasonable assurance that the device is safe when it can be determined, based on valid scientific evidence, that the probable benefits outweigh any probably risks. 21 CFR 860.7(d)(1). There is a reasonable assurance of effectiveness when it can be determined, based on valid scientific evidence, that the user of the device for its intended use will provide clinically significant results. 21 CFR 860.7(e)(1).

There are a number of factors that FDA considers when weighing the probable benefits and risks of medical devices. The FDA has issued guidance explaining the factors FDA considers when making benefit-risk determinations during the premarket approval process in "Guidance for Industry and Food and Drug Administration Staff; Factors to Consider When Making Benefit-Risk Determinations in Medical Device Premarket Approval and *De Novo* Classifications" (2012).[5] After a thorough review of the data, FDA's evaluation of a device's probable benefits

[4] See Guidance for Industry and Food and Drug Administration Staff; Factors to Consider When Making Benefit-Risk Determinations in Medical Device Pre-market Approval and De Novo Classifications, at http://www.fda.gov/downloads/MedicalDevices/DeviceRegulationandGuidance/GuidanceDocuments/UCM296379.pdf.

[5] See Guidance for Industry and Food and Drug Administration Staff; Factors to Consider When Making Benefit-Risk Determinations in Medical Device Pre-market Approval and De Novo Classifications, at http://www.fda.gov/downloads/MedicalDevices/DeviceRegulationandGuidance/GuidanceDocuments/UCM296379.pdf.

and risks, along with other factors, is intended to ensure that there are reasonable assurances of safety and effectiveness. Factors that FDA takes into account when considering the extent of probable benefit(s) include:

- the type of benefit(s),
- the magnitude of the benefit(s),
- the probability of the patient experiencing one or more benefits, and
- the duration of effect(s).

Factors that FDA takes into account when considering the extent of probable risk(s) or harm(s) include:

- the severity, types, number, and rates of harmful events associated with the use of the device,
- the probability of a harmful event,
- the duration of harmful events, and
- the risk from false-positive or false-negative results for diagnostic devices.

These factors should be considered during clinical trial design and conduct to ensure that the appropriate information is collected in the study to make these benefit-risk determinations at the time of the marketing application. Manufacturers are encouraged to frame their clinical study protocol in a benefit-risk framework.

3.4 Clinical Study Level of Evidence and Regulation

The regulations under 21 CFR Part 812 describe when approval of an IDE application is required prior to the initiation of the clinical study. A sponsor or sponsor's IRB (with or without consultation to FDA) must first determine if the proposed investigation is of a device that is a significant risk device or a non-significant risk device, although ultimately FDA is the final arbiter in this determination. *See* 21 CFR 812.2(b). If the study is with a significant risk device, the sponsor must submit an IDE to FDA for approval prior to commencing the study. *Id.* If the study is with a non-significant risk device, the study is considered to have an approved IDE when the sponsor obtains IRB approval of the investigation after presenting the reviewing IRB with a brief explanation of why the device is not a significant risk device, and maintains such approval and compliance with the abbreviated IDE regulations under 812.2(b). *Id.* In these situations an IDE application to FDA is not required unless FDA has notified a sponsor under 812.20(a) that approval of an application is required. *Id.*

In any case, the scientific rigor necessary for a clinical study and the robustness of evidence that need to be collected are not dependent on whether an IDE is required in order to initiate the study. Further, the rigor and robustness should not be influenced by the categorization of the clinical study as a study of a significant risk device, non-significant risk device, or device exempt from the IDE regulation. FDA encourages sponsors seeking guidance from FDA on the appropriate study design to support a potential pre-market submission to interact with FDA on key elements of a protocol through the Pre-Submission process in advance of finalizing the protocol. Even for studies that do not require IDE approval, FDA encourages submission of the draft protocol in a Pre-Submission to obtain FDA feedback prior to initiation of the study.

3.5 The Least Burdensome Concept and Principles of Study Design

In considering appropriate clinical study designs, FDA is also guided by the principle that the evidentiary burden must be commensurate with the appropriate regulatory and scientific requirements. This principle is reflected in two statutory provisions that apply to clinical and non-clinical data requirements for PMAs and 510(k) submissions. The following two provisions are referred to as the 'least burdensome provisions.'

Section *513(a)(3)(D)* of the Act provides, in relevant part, that:

> *(ii) Any clinical data, including one or more well-controlled investigations, specified in writing by the Secretary for demonstrating a reasonable assurance of device effectiveness shall be specified as a result of a determination by the Secretary that such data are necessary to establish device effectiveness. The Secretary shall consider, in consultation with the applicant, the least burdensome appropriate means of evaluating device effectiveness that would have a reasonable likelihood of resulting in approval.*
>
> *(iii) For purposes of clause (ii), the term "necessary" means the minimum required information that would support a determination by the Secretary that an application provides reasonable assurance of the effectiveness of a device.*
>
> *(iv) Nothing in this subparagraph shall alter the criteria for evaluating an application for premarket approval of a device.*

Similarly, Section *513(i)(1)(D)* the Act, provides*:*

> *(i) Whenever the Secretary requests information to demonstrate that devices with differing technological characteristics are substantially equivalent, the Secretary shall only request information that is necessary to making substantial equivalence determinations. In making such a request, the Secretary shall consider the least burdensome means of demonstrating substantial equivalence and request information accordingly.*
>
> *(ii) For purposes of clause (i), the term "necessary" means the minimum required information that would support a determination of substantial equivalence between a new device and a predicate device.*
>
> *(iii) Nothing in this subparagraph shall alter the standard for determining substantial equivalence.*

The FDA has issued guidance explaining how it intends to apply the least burdensome provisions in "The Least Burdensome Provisions of the FDA Modernization Act of 1997: Concept and Principles; Final Guidance for FDA and Industry" (2002) (The Least Burdensome Guidance). The Least Burdensome Guidance interprets "least burdensome" to mean a successful means of addressing a pre-market issue that involves the most appropriate investment of time, effort, and resources on the part of industry and the FDA. The guidance specifies that the least burdensome provisions do not affect the statutory pre-market review standards for devices, a principle affirmed by amendments added to the least burdensome provisions by the Food and

Drug Administration Safety and Innovation Act of 2012.[6] The guidance specifies further that for purposes of clinical study design, the FDA and industry should consider alternatives to randomized, clinical studies when potential bias associated with alternative controls can be minimized. The principles of study design discussed in this guidance are consistent with the principles discussed in the Least Burdensome Guidance, and expand upon them by discussing the considerations that may affect the level of evidence necessary to meet the standard for pre-market approval or clearance.

3.6 Approval of an Investigational Device Exemption

The principles discussed in this guidance are intended to assist sponsors of marketing applications and investigators in designing studies adequate to provide data to demonstrate a reasonable assurance of safety and effectiveness concerning a device. Following these principles and tailoring the design of a pivotal clinical study to the premarket review standard for devices may streamline the approval process by avoiding requests from FDA reviewers for additional information, re-analyses of data, or entirely new pivotal studies. However, FDA will not disapprove an IDE because the investigation is inadequate to support a marketing application. Section 520(g)(4)(C) of the FD&C Act, as amended by section 601 of FDASIA provides that FDA shall not disapprove an [IDE] because:

(i) the investigation may not support a substantial equivalence or de novo classification determination or approval of a device;
(ii) the investigation may not meet a requirement, including a data requirement, relating to the approval or clearance of a device; or
(iii) an additional or different investigation may be necessary to support clearance or approval of the device.

Thus, approval of an IDE does not necessarily indicate that FDA believes the study is adequate to support approval of a marketing application.

The focus of this guidance is on the set of principles of clinical study design that will help provide a reasonable assurance of safety and effectiveness for a device, rather than on requirements for approval of an IDE.

4 Types of Medical Devices

This document applies to two broad categories of medical devices based on intended use: (1) therapeutic and aesthetic devices and (2) diagnostic devices. Whether a device is intended for use as a therapeutic, aesthetic or diagnostic device depends on the indications for use statement in a device's label. These types of devices are described in this section, along with

[6] See Section 602 of the Food and Drug Administration Safety and Innovation Act of 2012 (FDASIA, Public Law 112-144).

characteristics unique to each type of device that should be considered when designing a pivotal clinical study. This guidance does not cover every type of device in every setting.

4.1 Types of Devices Based on Intended Use

Therapeutic and Aesthetic Devices

Therapeutic devices are generally intended for use in the treatment of a specific condition or disease. Aesthetic devices are intended to provide a desired change in visual appearance through physical modification of bodily structure.

Diagnostic Devices

In this guidance, diagnostic devices are described broadly as devices that provide results that are used alone or with other information to help assess a subject's health condition of interest, or target condition. A target condition can be a past, present, or future state of health, a particular disease, disease stage, or any other identifiable condition in a subject, or a health condition that could prompt clinical action such as the initiation, modification or termination of treatment. For the purposes of this guidance, diagnostic devices include devices intended for use in the collection, preparation and examination of specimens taken from the human body [*in vitro* diagnostic (IVD) devices], diagnostic imaging systems (e.g., digital mammography), *in vivo* diagnostic devices (non-imaging)[7], devices that provide an anatomical measurement (e.g., bone density, brain volume, retinal thickness), devices that provide a measurement of subject function (e.g., cardiac ejection fraction, subject reaction time), and algorithms that combine subject data to yield a subject specific output (e.g., a classification, score, or index). Note that while some understand the term diagnostic to refer to assessing the presence or absence of a disease, diagnostic tests include devices that, for example, are intended to detect pregnancy, assess immunity to a specific disease, provide information regarding the progress of disease, provide information about the disease-causing agent, or provide genotyping information that will be used to match a therapeutic product to a patient, along with devices that can assist the reader by automating various functions such as staining functions for a pathology system.

Devices with More than One Intended Use

While many devices can simply be categorized as therapeutic, aesthetic or diagnostic, there are devices that may fall into more than one of these categories, e.g., a device that both diagnoses a condition and then provides therapy for that condition when determined by the device to be present. There are also devices that may fall into one of these categories, but have more than one intended use in that category, e.g., a therapeutic device used to treat two very different conditions in two very different patient populations, or a diagnostic device that makes an initial diagnosis and also monitors progression of the same condition. Either case may result in a need to have more than one clinical study and possibly more supporting studies, e.g., bench studies or analytical ones.

[7] Diagnostic products that do not meet the "device" definition, for example because they achieve their primary intended purposes through chemical action on or in the body, are not covered by this guidance.

4.2 Special Considerations for Clinical Studies of Devices

Certain considerations unique to medical devices should be taken into account in designing a clinical study of a device. These considerations apply to all devices, be they therapeutic, aesthetic, or diagnostic devices, although the device type may influence study design decisions. The following characteristics and features unique to medical devices can influence how the device is evaluated by FDA and should be addressed in the clinical study design:

- How the device works: In all devices, an understanding of the scientific principles underlying device function and mechanism of action may be relevant in assessing performance and the adequacy of the proposed study design. This information is especially important as part of a Pre-Submission in which a sponsor requests FDA's advice in developing their clinical studies.
- User skill level and training: Some devices require considerable training and skill to use in a safe and effective manner. This clearly would apply to implantable devices requiring the user to be a highly trained surgical specialist, particularly when the procedure involved is complex. Sometimes multiple personnel and skill sets are needed for appropriate use of the device. For example, for an IVD, one person with a certain skill set may collect the specimen, another person with a different skill set may process the specimen, and still another person with a third skill set may interpret the test results. When designing a device study, one should consider the skills necessary for the safe and effective use of the device. The skill sets of study investigators and personnel should reflect the range of skills of personnel likely to use the device in the intended use setting after marketing approval. The training provided to study investigators and personnel in the appropriate use of the device should guide the training that will be provided to users when the device is marketed. If no training will be provided for a marketed device, study personnel should not be specifically trained in the use of the device in order to ensure that the study reflects intended use conditions.
- Learning curve: Some devices are so novel that there is a learning curve associated with use of the device. With novel technologies, it may take time to master the steps prior to using the device in the clinical study. For an implant this would usually include any surgical techniques specific to the implantation procedure. For some devices, determination of a learning curve can be addressed during the exploratory stage, including pilot studies. If hands-on training of device operators is provided by a sponsor in the pre-market pivotal study, then one would expect such training to be provided in the post-market setting. Devices with steep learning curves may not be suitable for some settings (e.g., home use) because they may not be safe and effective in that setting. When a learning curve is evident during the pivotal study, it is important to consider how information gathered during this learning curve period are considered in the protocol (e.g., by clearly defining which subjects in the study are part of the learning curve period) and how results will be reported in the Statistical Analysis Plan (SAP). If the learning curve is steep, this may have ramifications in labeling and in training requirements for users.

- Human factors considerations: Human factors can play a crucial role in the development of a medical device.[8] At any point in the device developmental process, the study of human factors associated with the use of the device may necessitate changes to the design of the device or instructions for use to make it safer or more effective or easier to use for subjects or medical professionals. Devices that incorporate software or provide a user interface should be designed to minimize user error. Devices that require more manual intervention or subjective judgment on the part of the user may require more user skill. Clear documentation of the device-user interface under real-world scenarios should be part of a human factors assessment.

5 The Importance of Exploratory Studies in Pivotal Study Design

Medical devices often undergo design improvement during development, with refinement during lifecycles beginning with early research, extending through investigational use and initial marketing of the approved or cleared product, and continuing on to subsequently approved or cleared commercial device versions.

For new medical devices, as well as for significant changes to marketed devices, clinical development in many cases is marked by the following three stages: the exploratory stage (first-in-human, feasibility), the pivotal stage (determines the safety and effectiveness of the device), and the post-market stage (design improvement, better understanding of device safety and effectiveness and development of new intended uses). While these stages can be distinguished, it is important to point out that device development can be an ongoing, iterative process, requiring additional exploratory and pivotal studies as new information is gained and new intended uses or indications are developed. Insights obtained late in development (e.g., from a pivotal study) can raise the need for additional studies, including clinical or non-clinical.

This section focuses on the importance of the exploratory work (in non-clinical and clinical studies) to development of a pivotal study design plan. Non-clinical testing (e.g., bench, cadaver, modeling, or animal) can often lead to a better understanding of the mechanism of action and can provide basic safety information for those devices that may pose a risk to subjects. The exploratory stage of clinical device development, consisting of first-in-human and feasibility studies, is intended to allow for any iterative improvement of the design of the device, to advance the understanding of how the device works and its safety, and to set the stage for the pivotal study.

Thorough and complete evaluation of the device during the exploratory stage results in a better understanding of the device and how it is expected to perform. This understanding can help to confirm that the intended use of the device will be aligned with sponsor expectations. It also can help with the selection of an appropriate pivotal study design. A robust exploratory stage should also bring the device as close as possible to the form that will be used both in the pivotal trial

[8] See Medical Device Use-Safety: Incorporating Human Factors Engineering into Risk Management (July 18, 2000), http://www.fda.gov/MedicalDevices/DeviceRegulationandGuidance/GuidanceDocuments/ucm094460.htm.

and in the commercial market.[9] This reduces the likelihood that the pivotal study will need to be altered due to unexpected results. This is an important consideration, since altering an ongoing pivotal study can increase cost, time, and patient resources. This might also invalidate the study or lead to its abandonment. In general, in order to make scientifically valid confirmatory inferences, feasibility study data should not be combined with pivotal study data without advanced planning at the exploratory stage.

For diagnostic devices, analytical validation of the device to establish performance characteristics such as analytical specificity, precision (repeatability/reproducibility), and limit of detection are often part of the exploratory stage. In addition, for such devices, the exploratory stage may be used to develop an algorithm, determine the threshold(s) for clinical decisions, or develop the version of the device to be used in the pivotal clinical study. For both *in vivo* and *in vitro* diagnostic devices, results from early clinical studies may prompt device modifications and thus necessitate additional small studies in humans or with specimens from humans. FDA should be consulted prior to initiating these studies.

Exploratory studies may continue even as the pivotal stage of clinical device development gets underway. For example, FDA may require continued animal testing of implanted devices at 6 months, 2 years and 3 years after implant. While the pivotal study might be allowed to begin after the six-month data are available, additional data may also need to be collected. As another example, additional animal testing might be required if pediatric use is intended. For *in vitro* diagnostic devices, it is not uncommon for stability testing of the device (e.g., for shelf life) to continue while (or even after) conducting the pivotal study.

While the pivotal stage is generally the definitive stage during which valid scientific evidence is gathered to support the primary safety and effectiveness evaluation of the medical device for its intended use, the exploratory stage should be used to finalize the device design and the appropriate endpoints for the pivotal stage. This is to ensure that the investigational device is standardized as described in 21 CFR 860.7(f)(2), which states:

> *To insure the reliability of the results of an investigation, a well-controlled investigation shall involve the use of a test device that is standardized in its composition or design and performance.*

6 Some Principles for the Choice of Clinical Study Design

FDA reviews medical device pivotal clinical studies submitted as part of marketing applications to determine whether they provide reasonable assurance of device safety and effectiveness. FDA recognizes that there may be several types of studies that can fulfill this expectation. FDA therefore encourages applicants to meet with the appropriate FDA review division to discuss

[9] In some cases for *in vitro* diagnostic devices that are used as companion diagnostic devices for therapeutic products, a non-final version of the device is used in the clinical trial of the therapeutic product. When this occurs, careful advance planning and execution of "bridging" studies are needed to establish clinical validity of the commercial *in vitro* diagnostic device.

study design choices for demonstrating reasonable assurance of device safety and effectiveness prior to study commencement.

In this document two broad types of clinical studies will be distinguished: clinical outcome studies and diagnostic clinical performance studies. The following discussion is predicated on the choice of appropriate questions to be answered by the study using clinically meaningful and statistically appropriate study endpoints.

This section addresses some of the considerations applicable to all pivotal clinical studies of medical devices. Various factors are important when designing any medical device clinical study, including general considerations of bias, variability, and validity, as well as specific considerations related to study objectives, subject selection, stratification, site selection, and comparative study designs. Each of these is defined and discussed below.

6.1 Types of Studies

Clinical Outcome Studies

In a clinical outcome study, subjects are assigned to an intervention and then studied at planned intervals using validated assessment tools to assess clinical outcome parameters (or their validated surrogates) to determine the safety and effectiveness of the intervention. These studies are described in greater detail in Section 7. It may be the case that device clinical performance (i.e., whether the device has the intended effect in the clinical setting) is also studied but the primary focus of the investigation is one or more clinical outcomes. For purposes of this document, the term "intervention" refers to either the use of an investigational device or a control. The investigational device could be therapeutic or aesthetic. For diagnostic devices, the term "intervention" relates to a strategy for subject management based on the result produced by the diagnostic device. A clinical outcome study is used to evaluate a diagnostic device when the goal is to evaluate the impact of how the device result changes a subject's subsequent course of treatment or management by the health care provider.

Diagnostic Clinical Performance Studies

For the majority of diagnostic devices, the pivotal clinical evaluation is not a clinical outcome study but a diagnostic clinical performance study. It could be that a performance study also may have clinical outcomes but these outcomes are not the primary focus of the study. These studies are described in greater detail in Section 8. In a diagnostic clinical performance study, diagnostic test results are obtained from subjects, but are not used for subject management. Instead, the diagnostic clinical performance of a test is characterized by performance measures that quantify for each subject how well the diagnostic device output agrees with a clinical reference standard that is used to assess subjects for the target condition, as described in greater detail in Section 8.

General Considerations

Devices with both diagnostic and therapeutic functions, e.g., to detect a condition and then administer the treatment, may be assessed using a diagnostic clinical performance study and/or a clinical outcome study with diagnostic performance elements.

Clinical outcome studies and diagnostic clinical performance studies are discussed separately in this document. For more information on clinical outcome studies, please refer to Section 7, and for more information on clinical performance studies for diagnostic devices, please refer to Section 8. For products with both diagnostic and either therapeutic or aesthetic components, please read both Sections 7 and 8. Section 9, which provides information on plans and techniques that sustain the level of evidence of clinical studies, applies to both clinical outcome studies and diagnostic clinical performance studies.

6.2 General Considerations: Bias and Variability in Device Performance

Designing studies to collect the *right* data is more important than designing studies simply to collect *more* data. The study design should consider both bias and variability. When evaluating a study design to determine whether the design will support approval of a PMA, an important consideration is the statistical concept of bias. Bias is the introduction of systematic errors from the truth. Bias can be introduced in subject selection, study design, study conduct, and data analysis procedures. In a clinical study, bias may lead to an incorrect determination of safety and effectiveness. Study designs that introduce little or no bias are preferable to designs that do not control for bias, which can be introduced into clinical studies due to a number of reasons. Some of these are reviewed below with strategies that can help to eliminate or minimize bias in the design phase (see also Sections 7 through 9).

Bias can distort the interpretation of study outcomes. When the performance of the device is good, the presence of moderate bias may not distort the study's ability to support conclusions about overall effectiveness; when the performance is known (or thought) to be marginal, the performance may be overwhelmed by the bias in some study designs. Particularly when there has been insufficient study in the exploratory stage and the device effect may not be well understood, it may be difficult to choose an appropriate study design under which the device effect would not be overwhelmed by bias. Consideration of the potential for study bias is a critical factor in designing a study to reduce the risk that bias may invalidate the final study results.

A second general consideration when evaluating a study design for level of evidence is the sampling variability, which is controlled by the sample size of the study. On the one hand, a larger sample size provides more data so that estimates of performance have less sampling variability and hence become more precise. On the other hand, larger sample size can also result in an analysis for a clinically insignificant outcome that demonstrates it is statistically significant. Studies should be designed to show both clinical and statistical significance. It is also important to note that increased sample size will not necessarily address issues of bias, inappropriate outcomes assessment, a marginal improvement in outcomes that fails to show clinical relevance, or other study design problems.

6.3 Study Objectives

The study objectives provide the scientific rationale for why the study is being performed. The objectives should provide support for the intended use of the device, including any desired labeling claims.

Claims can be supported statistically by formal hypothesis testing or by point estimates with corresponding confidence intervals. For pivotal studies designed to test a scientific hypothesis, the study objectives should include a statement of the null and alternative hypotheses that correspond to any desired claim. For studies with estimation goals (e.g., some diagnostic performance studies), rather than hypothesis testing, claims can be supported with point estimates and confidence intervals describing device performance.

6.4 Subject Selection

21 CFR 860.7(f)(1)(ii) states that the plan or protocol for a study must include:

> *A method of selection of the subjects that:*
>
> *(a) Provides adequate assurance that the subjects are suitable for the purposes of the study, provides diagnostic criteria of the condition to be treated or diagnosed, provides confirmatory laboratory tests where appropriate and, in the case of a device to prevent a disease or condition, provides evidence of susceptibility and exposure to the condition against which prophylaxis is desired;*

Subjects selected for any clinical study should adequately reflect the target population for the device (i.e., the population for whom the device is intended) based on specific enrollment criteria and confirmatory laboratory or other testing. If the study enrolls subjects who do not represent the target population then the study results have the potential for subject selection bias.

To ensure that the subjects in the clinical study reflect the desired target population, the protocol should specifically define eligibility criteria that match the key characteristics of the intended target population. In conducting the trial, the sponsor should ensure that only those individuals meeting these criteria are included. These are referred to as the inclusion/exclusion criteria for subject entry into the study.

In considering the target population, FDA encourages sponsors to enroll subjects who would reflect the demographics of the affected population with regard to age, sex, race and ethnicity.[10,11] Inadequate participation from some segments of the population can lead to insufficient information pertaining to device safety and effectiveness for important subpopulations. We recommend including a background discussion of prevalence, diagnosis and treatment patterns for the type of disease for which the device is intended, if appropriate. This discussion should include: sex- and race-specific prevalence; identification of proportions of women and minorities included in past trials for the target indication; and a discussion of plans to address any factors identified or suggested, which may explain the potential for under-

---

[10] Guidance on the Collection of Race and Ethnicity Data in Clinical Trials, Sept. 2005, at http://www.fda.gov/RegulatoryInformation/Guidances/ucm126340.htm.

[11] To provide information on evaluation of clinical trial outcomes by sex, FDA has published the draft guidance, "Draft Guidance for Industry and Food and Drug Administration Staff – Evaluation of Sex Differences in Medical Device Clinical Studies," at http://www.fda.gov/downloads/MedicalDevices/DeviceRegulationandGuidance/GuidanceDocuments/UCM283707.pdf. When finalized, this guidance will represent the Agency's current thinking on this topic.

representation of women and minorities, if applicable. We recommend including a summary of this information in the protocol and investigator training materials. Consideration should be given to enrollment of investigational sites where recruitment of needed populations for the study can be more easily facilitated. In the description of the patient population [21 CFR 812.25(c)] and use of foreign data [21 CFR 814.15(d)(1)], consideration of how each is applicable to the U.S. population and U.S. medical practice should be included in the study design.

When a clinical study involves vulnerable populations, such as children, prisoners, pregnant women, physically handicapped or mentally disabled persons, or economically or educationally disadvantaged persons, the sponsor should be prepared to discuss potential issues with FDA in advance of the study so that they comply with 21 CFR 56.111(b) and 21 CFR Part 50.

There may be information known in advance of a study that can improve the conduct of the study and enhance its chances for success. In planning a study it is important to consider factors that may be related to outcomes, such as skill of the user/surgeon, disease severity, and sex or age of the subjects. Some caution should be exercised with respect to adequately representing all important subgroups, e.g., sex, age, ethnicity, and groups that are particularly important to the current study. When the condition of interest is rare, subject selection for diagnostic studies can be challenging and alternative approaches may be considered.

The protocol may include one of several possible subject selection methods: random selection, consecutive selection, systematic selection, and/or convenience selection.

- In **random selection**, all subjects have a known (usually the same) chance of being selected for the clinical study. When implementable, a random selection of subjects for the clinical study has the potential to provide unbiased estimates for the population from which they are selected. Random selection is often not practical due to logistical difficulties.
- Clinical studies commonly use **consecutive selection** (i.e., selecting every subject in the order they present at the site) or systematic selection (e.g., selecting every tenth subject) among those who meet the inclusion/exclusion criteria. These selection methods will likely provide unbiased estimates so long as the study period is not confounded with other variables associated with the subject that might affect outcome. For instance, if the study lasts one morning, study subjects may not be representative of the target population since subjects who visit the clinic in the morning may not be representative of all subjects who visit the clinic.
- **Convenience selection** is a method where subjects are selected because of their convenient accessibility to the researcher. This method may provide results that cannot be generalized to the target population.

6.5 Stratification for Subject Selection

When studies enroll subjects at multiple sites, it is necessary to select subjects from sites that adequately represent the target population. Sometimes, this cannot be achieved by simply selecting representative sites. Performance of the device needs to be adequately characterized in important subgroups where differences in performance or utilization are expected. For example,

a device that is indicated for use by both men and women should not enroll mostly men; one that is indicated for all adults should adequately represent all adult age groups.

There are two broad types of techniques for selection of subjects: stratified selection and selection just based on inclusion/exclusion criteria.

Stratification involves dividing the target population into pre-specified non-overlapping subject subgroups or strata. Stratified selection of subjects means that subjects are selected separately from each subgroup (stratum). For example, one may decide to stratify a subject population by sex (male, female) and by age group (below or above a given age), resulting in four strata, each defined by a unique combination of sex and age. These characteristics are recorded as subjects enter a study, and are not the result of the intervention. Stratified subject selection not only ensures adequate representation of important subgroups but may also provide estimates of device performance that are statistically more precise. When there is reason to believe the device performs differently in different subgroups, it may be beneficial to consult with FDA to determine an acceptable design.

Often, subjects are recruited without regard to specific baseline characteristics or strata but just according to pre-specified inclusion/exclusion criteria. This type of selection may be adequate when the device is expected to perform similarly in all subject subgroups.

In such clinical outcome studies, when a decision is made to study important subgroups or strata such as the multiple centers at which the study is being conducted or covariates that are thought to be highly predictive of subject outcomes such as the presence or absence of co-morbidities (e.g., diabetes), it is often wise to also consider stratified randomization in which randomization occurs separately in each of the pre-specified strata.

6.6 Site Selection

Sponsors should select subject enrollment sites (centers) that are appropriate for the intended use of the device. For diagnostic devices, testing sites are usually different than the subject enrollment sites.

Single-center investigator studies may be a useful starting point in evaluating the initial feasibility of a new device since they are logistically easier to coordinate, less resource-intensive, and typically focused on a more homogeneous subject population with fewer confounding variables. They also aid in planning for larger, multicenter studies. A single-center study is rarely adequate for a pivotal clinical study

Rather, evaluation of the safety and effectiveness of an investigational device is typically dependent on demonstrating generally consistent results across a number of study sites in a larger multicenter study. An advantage of multicenter studies is that it is easier to recruit subjects and the required sample size is typically reached faster.

A multicenter study may assure a more representative sample of the target population and make it easier to generalize the findings of the study. Differences in outcomes among centers are very important in the evaluation of medical device study outcomes because they may reflect

differences in subject selection, surgical technique, and clinician skills, as well as any learning curve, all of which could bias interpretation of study results. Similarly, in diagnostic clinical performance studies, the subjects or specimens may be referred from other centers and the skill of the person performing the test, as well as the person interpreting the result, can vary. Since study results may vary considerably from center to center in both clinical outcome and diagnostic clinical performance studies, special statistical techniques may be required to combine study results from several centers (e.g., analysis to support pooling).

Where applicable, special care should be taken to ensure that the study sites will include subjects who reflect the epidemiological distribution of the disease being treated with regard to variables such as sex, age, race, ethnicity, socio-economic status, and coexisting conditions. In addition, the inclusion of subjects who may have a spectrum of disease different from that of the intended use population (e.g., inclusion of additional rare disease subjects for a device used to screen subjects) will likely result in a biased estimate of device performance (referral bias). For some diagnostic studies, different sites may reflect subjects with characteristics such as high risk or average risk of a disease, and these results may need to be considered separately. (See Section 6.5)

Similarly, depending on the device, it may be important to consider diversity of sites in terms of investigator or operator experience. For example, for a clinical outcome study, surgeons at a tertiary care facility may have more specialized experience than those at a community hospital. Therefore, only selecting tertiary care facilities for a clinical study could lead to a biased assessment of device performance.

A study to support a pre-market submission in the United States should be relevant to understanding the safety and effectiveness of the device when used in patients in the United States with regard to subject demographics, standard of care, and practice of medicine. This is important for studies conducted both in and outside of the United States. Studies that fail to meet these criteria may be inadequate to support approval of a device.

The sponsor is responsible for selecting investigators that have the training and experience necessary to investigate the device. *See* 21 CFR 812.43(a). Note that for certain novel devices, additional training may be needed as part of the planned clinical study to facilitate safe use of the device. Sponsors should ensure that users of the investigational device have any training and experience that are necessary for such use. Investigators are also expected to know the applicable regulations and guidances that guide the conduct of clinical research.

6.7 Comparative Study Designs

Studies that compare two or more treatments or the performance of two or more diagnostic tests are called comparative study designs. There are several different types of comparative designs.

- **Parallel group design**: Each subject or sample is assigned only one of the possible treatments or tests being compared. Because a different group of subjects (or samples) is assigned to each treatment (or each diagnostic test), comparisons are made between subject groups. With this type of study, randomization of subjects to the intervention

groups is recommended to ensure a fair comparison, so that groups are comparable at baseline prior to the treatment or test.

- **Paired design**: Each subject or sample receives all of the treatments or tests at the same time. Therefore, treatments or tests can be compared on the same subjects. An example would be a split-face design in which each side of the face was treated with a different device. In general, comparisons are often more precise with a paired design than a parallel group design because with a parallel group design, the comparisons made across subject groups include variability between subjects. The paired design is less common for some therapeutic or aesthetic treatments, but is very common in diagnostic studies where different diagnostic devices may be used on the same subject or sample. One disadvantage of this design is that it may be difficult to distinguish whether a non-local adverse event is associated with one or both of the two interventions.

- **Cross-over design**: Each subject or sample receives two or more treatments (or diagnostic tests) at different times, but in a predetermined sequence. Multiple sequences of interventions are often studied, with each subject receiving the treatments (or diagnostic tests) in a specified sequential order. A cross-over design may be appropriate when all of the treatments (or diagnostic tests) cannot be assigned to a subject at the same time (i.e., a paired design is not feasible) and when the effects of one treatment (or diagnostic test) do not carry over to the next. With this type of design, the order should be randomized unless otherwise justified. Cross-over designs are possible in therapeutic, aesthetic and diagnostic studies.

7 Clinical Outcome Studies

Various important factors need to be considered in designing a clinical outcome study. This section discusses these factors, including:

- Specific considerations of subject endpoint(s), intervention assignment (randomization), blinding, placebo effect, controls, and non-comparative studies.
- Sources of bias and general considerations for bias minimization.

7.1 Endpoints in Clinical Studies

In any clinical study, key study variables are chosen that will demonstrate device performance and clinical outcomes. For clinical outcome studies, these variables are the primary and secondary clinical endpoints. It is important that all primary and secondary endpoints are pre-specified at the design stage of the pivotal clinical study.

Device performance and clinical outcomes should be objectively measured with minimal bias. Some considerations include:

- The endpoints, outcomes, or measurements should provide sufficient evidence to characterize the clinical effect of the device (for both safety and effectiveness) for the desired intended use.

- The protocol should specify what endpoints or outcomes are being measured, how and when they are being measured, and how they will be analyzed statistically. A more detailed description of the statistical analysis may be contained in a separate Statistical Analysis Plan.
- The endpoints, outcomes, or measurements should be clinically meaningful and relevant to the stated study objectives and desired intended use. The pivotal study should be designed to demonstrate clinical benefit to the specified subject population rather than to simply demonstrate how the device functions.
- Whenever possible, the endpoint should be objective, be internally and externally valid, and determined with minimal bias. For example, when the therapeutic endpoint involves a diagnostic assessment such as presence of stroke or myocardial infarction, a diagnostic clinical reference standard would be preferable to a subjective assessment. Relying on the subjective clinical assessments to determine an endpoint in a clinical outcome study is typically inadequate when more objective assessment methods exist. In such cases the use of a committee of clinical experts to adjudicate the endpoints may address some of the concerns.
- The protocol should specify who will evaluate the endpoints, outcomes, or measurements in relation to the subjects and/or study investigators, e.g., an evaluator blinded (masked) to the intervention assignment versus the investigator.
- For some studies, an independent adjudication committee may be warranted to adjudicate an endpoint, for example, when objective assessments do not exist and a subjective assessment is used, such as in the case of an interpretation of a radiograph. The rules by which endpoints, outcomes, or measurements are adjudicated should be defined in advance in the pivotal study protocol.
- A subject-reported outcome instrument can be used when the outcome of interest and desired intended use are best measured from the subject's perspective, (e.g., pain reduction). In such cases, it is important to select a scoring assessment that is validated for the subject population and condition being treated, and consistent with the desired intended use. For this reason, early discussion with FDA during the study design phase is important. These more subjective measures are often used in conjunction with more objective assessments as part of a composite endpoint. When using subjective rating scales in multinational trials, sponsors should make sure that the scales are interpretable and valid across cultures and languages. For more information on the use of subject-reported outcomes and their validation, refer to FDA Guidance.[12]
- A composite endpoint is an endpoint that is a pre-specified combination of more than one endpoint. Use of a composite endpoint can be challenging and requires careful and early discussion with FDA to formulate the appropriate endpoint and analysis plan. When a composite endpoint is used, in addition to analysis of the effect of the device on the overall composite endpoint, FDA will also evaluate the effect of the device on each of the

[12] Guidance for Industry: Patient-Reported Outcome Measures: Use in Medical Product Development to Support Labeling Claims. For more information on this guidance, see http://www.fda.gov/downloads/Drugs/GuidanceComplianceRegulatoryInformation/Guidances/UCM193282.pdf.

component endpoints so that domination of the composite by any of its components or lack of consistency in individual component results can be assessed.

- For multiple primary effectiveness (or safety) endpoints, providing an explanation for the role and relative importance of each endpoint is encouraged. A well-written protocol should carefully define study success/failure criteria with respect to each of multiple primary endpoints, and pre-specify appropriate statistical approaches to handling multiplicity issues and controlling for overall Type I error rates. When multiple secondary endpoints are selected with potential additional intended uses or claims in mind, the protocol should pre-specify appropriate statistical methods to analyze data and interpret results.
- Use of surrogate endpoints may be appropriate when they are validated and directly correlated to clinical benefit. Early discussion with FDA during the study design phase is critical to determine whether the surrogate has been established as valid, or, if not, to determine an acceptable means of validating the surrogate.

The following issues should be considered when choosing a primary endpoint for a clinical outcome study:

- The endpoints, outcomes or measurements should be carefully selected to avoid a situation where they are undefined or may be unobtainable for a substantial proportion of subjects.

- Sponsors should give careful consideration when designing a study as to the total study duration (including the time to complete enrollment and the total length of follow-up), the time-point at which safety and effectiveness endpoints will be evaluated, and for how long subjects will be consented for follow-up via the informed consent process. Among the aspects worth considering are the earliest time-point(s) at which safety and effectiveness should be evaluated for purposes of performing a risk/benefit analysis, as well as the possibility that additional years of follow-up may be required post-approval (i.e., as a condition of PMA approval). The selection of the time point(s) for a primary endpoint should take into account the time course for activity of the product, considering evidence from prior studies. Sponsors should be mindful that if they have committed to a certain length of follow-up in the approved protocol and through the informed consent process, generally the sponsor is expected to comply with the protocol and follow enrolled subjects for the entire period.
- When the understanding of science or medicine changes during the course of a particular device study, the relevance of particular endpoints, outcomes or measurements may change. Changing study endpoints during the trial may make the trial difficult to interpret, for both the FDA review team and the sponsor, especially for a pivotal trial. For a well-controlled trial design, all of the major elements should be pre-specified, and any changes in key elements, including study endpoints, may seriously impact trial interpretation and data analysis. In such cases, sponsors are advised to contact the appropriate FDA review division to discuss the best possible course of action.

7.2 Intervention Assignment (Randomization) for Clinical Outcome Studies

21 CFR 860.7(f)(1) states that the plan or protocol for a study must include:

(ii) A method of selection of the subjects that:

(a) Provides adequate assurance that the subjects are suitable for the purposes of the study, provides diagnostic criteria of the condition to be treated or diagnosed, provides confirmatory laboratory tests where appropriate and, in the case of a device to prevent a disease or condition, provides evidence of susceptibility and exposure to the condition against which prophylaxis is desired;

(b) Assigns the subjects to test groups, if used, in such a way as to minimize any possible bias;

(c) Assures comparability between test groups and any control groups of pertinent variables such as sex, severity or duration of the disease, and use of therapy other than the test device;

Randomization of subjects to intervention groups is generally recommended to assure an appropriate comparison, so that groups are comparable at baseline prior to the intervention or test. Randomization tends to assure balance between intervention groups in terms of pertinent variables such as sex and other demographic variables, severity or duration of the disease, prior therapies, professional user biases and/or preferences, and use of interventions other than the investigational device. Also very importantly, randomization similarly acts to balance unmeasured or unknown covariates. In some instances it may be possible to consider performing randomization not at the individual level but at a different level, for example, at the clinical site level, in which case clinical sites are randomized to the interventions and then all subjects at a particular clinical site are assigned to the same group; this is called cluster randomization. Sponsors are encouraged to consult with FDA prior to assigning interventions by cluster randomization.

In a parallel group clinical outcome design, randomization is typically used to assign each subject to an intervention in an unbiased manner. In the paired clinical outcome design, in which each subject serves as his or her own control, reliance on randomization to assign the order of two interventions or locations (e.g., right vs. left sides of the face, left versus right knee) in which each intervention is applied for each particular subject helps minimize bias. In a cross-over design, the order of interventions to each subject is generally randomly determined. Failure to randomize in a parallel study, a paired study or a cross-over design study risks study failure by allowing bias to distort the results.

When the design of a device or the intended subject population makes it impossible to randomize the intervention assignment, the study may be subject to bias of unknown size and direction, and such bias can adversely impact the level of evidence provided by the study and the ability to rely on the data as valid.

The Agency acknowledges that there are situations in device studies where randomization is impossible, difficult, or potentially inappropriate. For example, investigators may face an ethical dilemma in recommending a randomized study to subjects when they believe that the different interventions in the study are not equally safe and effective (i.e., they lack clinical equipoise). In such cases, sponsors are advised to contact FDA prior to submitting their pre-market approval application or notification to discuss their concerns with randomization and determine an appropriate study design that will provide an adequate level of evidence in such a situation.

7.3 Blinding (Masking)

Limiting knowledge of intervention assignment, without jeopardizing subject care or study objectives, is referred to as blinding, also called masking. (The term masking is more appropriate for ophthalmic products.) In the context of a clinical outcome study, knowledge of the intervention assignment can influence the behavior and decisions of the subject, clinician, investigator, care-givers and third-party evaluators, whether consciously or unconsciously.

If the subject's assignment to an intervention is not blinded (masked), the behavior of the subject may be affected by knowledge of the intervention and consequently a bias can be introduced, particularly if a clinical measurement or endpoint is subjective.

If the investigator or a third-party evaluator is not blinded (masked) to the intervention assignment, then investigator or evaluator bias can adversely affect the study by influencing the interpretation of clinical outcomes, the performance of surgical implantation of a device, and subsequent clinical decision-making.

Even in cases where blinding the subject and/or investigator is not possible, it may still be possible and is strongly recommended that independent, third-party evaluators of clinical measurements and/or endpoints be blinded to the intervention assignment. It is preferable to use evaluators who do not know the study objectives but rather are asked to perform evaluations based on objective criteria (e.g., clinical, radiographic). Alternatively, independent core labs and reading centers, and/or clinical events committees that employ prospectively defined key definitions and Standard Operating Procedures, can be used to minimize the bias that could occur if evaluations were affected by knowledge of the intervention assignment.

In some clinical outcome device studies, particularly those that are highly invasive or in which device treatment is compared to medical therapy or surgical intervention, it may be impossible to blind the subject or the investigator to the intervention assignment. However, even if it is inconvenient or difficult, FDA recommends that blinding be considered and attempted if at all possible. When a study is blinded, it is often very informative for the study design to include an evaluation of the integrity and effectiveness of the blinding by asking the participants (i.e., subjects and/or investigators) blinded to the intervention assignment at the end of the study to indicate which intervention group they thought they were in and why. This could be helpful to better understanding the study results, especially if knowledge of the assigned arm can affect the overall outcome.

In cases where blinding of intervention assignment of study participants throughout the entire study is not possible, the following are considered potential means to minimize bias as much as possible:

- Subjects and study staff should be blinded regarding impending intervention assignment until after a potential subject has been screened and has completed enrollment.

- It is strongly suggested that subjects be blinded until after the procedure to avoid issues with differential dropout that may be related to knowledge of the intervention assignment.

- More objective endpoints are usually preferable to subject reported outcomes if blinding is not performed. In cases in which subject-reported outcomes are employed, consideration should be given to measure them without any clinical staff associated with the investigation present.

- Drafting a script for clinical staff to use to standardize the follow-up questions asked of study participants.

7.4 Controls in Comparative Clinical Outcome Studies

21 CFR 860.7(f)(1)(iv) identifies four types of controls. It states that the plan or protocol for a study must include:

> *A comparison of the results of treatment or diagnosis with a control in such a fashion as to permit quantitative evaluation. The precise nature of the control must be specified and an explanation provided of the methods employed to minimize any possible bias of the observers and analysts of the data. Level and methods of "blinding," if appropriate and used, are to be documented. Generally, four types of comparisons are recognized;*
>
> *(a) No treatments. Where objective measurements of effectiveness are available and placebo effect is negligible, comparison of the objective results in comparable groups of treated and untreated patients;*
>
> *(b) Placebo control. Where there may be a placebo effect with the use of a device, comparison of the results of use of the device with an ineffective device used under conditions designed to resemble the conditions of use under investigation as far as possible;*
>
> *(c) Active treatment control. Where an effective regimen of therapy may be used for comparison, e.g., the condition being treated is such that the use of a placebo or the withholding of treatment would be inappropriate or contrary to the interest of the patient;*
>
> *(d) Historical control. In certain circumstances, such as those involving diseases with high and predictable mortality or signs and symptoms of predictable duration or severity, or in the case of prophylaxis where morbidity is predictable, the results of use of the device may be compared quantitatively with prior experience historically derived from the adequately documented natural history of the disease or condition in comparable patients or populations who received no treatment or who followed an established effective regimen (therapeutic, diagnostic, prophylactic).*

In addition to the four types of controls identified in the CFR, this guidance also considers a fifth, "Subject Serving as Own Control." In this guidance the term "intervention" will be used instead of "treatment" when describing a control in (a) and (c) above since this term applies to clinical outcome studies for diagnostic interventions as well as to therapeutic and aesthetic interventions.

Each control has advantages and limitations for use in a clinical study. In general, there is less bias associated with study designs that use concurrent controls than with those that use non-concurrent controls.

Table 1 outlines some considerations for each type of control in relation to study bias and resulting level of evidence.

Table 1: Types of Controls for Clinical Outcome Studies

Type of Control	**Subcategory**	**Description**	**Considerations**
Concurrent Control	Active Intervention Control ("Active")	Control group provides another intervention (usually another device or surgery, but possibly a drug or biological product) that delivers a known effect.	• Demonstration of either superiority or non-inferiority to active control. • Choice of an appropriate control is based on the current standard of care for the intended subject population • Extent of knowledge about the effect size of the active control
	Placebo Control ("Sham")	Control group may be another device, simulated procedure or possibly a drug or biological product that is believed to have no therapeutic (or diagnostic) effect.	• A placebo control is useful if there is thought to be a placebo effect. • It may be challenging to construct a placebo control that appears to function like the investigational device but delivers no therapy. • In some cases, it may be unethical to randomize subjects to a placebo that will provide no known effect.
	"No Intervention" Control	Control group provides no intervention (or diagnosis).	• Choice of a "no intervention" control may present a challenge in recruiting subjects who might receive no intervention or keeping subjects enrolled who were randomized to the "no intervention" control group. • Choice of a "no intervention" control has built-in bias because control group subjects expect to receive no benefit, whereas experimental group subjects expect to receive a benefit. • A "no intervention" control may sometimes be standard of care/best medical management which can provide evidence about any incremental benefit or risk, although the control could vary among the different study centers.
	Subject as own control	Subject serves as concurrent control to self (e.g., split face, fellow eye, etc.).	• For therapeutic and/or aesthetic device studies, use of the subject as his/her own concurrent control allows for the advantageous use of the correlation within the subject. • This design is only possible when the experimental device and control intervention effects are local and do not overlap.
Non-concurrent Control	Subject as own control	Subject's outcomes at baseline compared to outcomes at endpoint evaluations.	• Use of baseline outcomes as a comparison for outcome at the endpoint evaluations is inadequate for most therapeutic studies since subjects may improve for reasons unrelated to investigational device (e.g., regression to the mean, placebo effect).

Type of Control	Subcategory	Description	Considerations
Non-concurrent Historical Control	Subject-level data on a parallel group	Control group consists of a different group of subjects treated in the past for whom individual subject-level data are available for same outcomes and same covariates as in current study.	• A significant concern is comparability between the two groups with respect to baseline covariates. The use of a comparator study separated in time can introduce severe and unknown selection bias. • Statistical methods such as covariate analysis or propensity score analysis can potentially address some concerns. • The historical control group may not reflect current practice of medicine and may include a different subject population and/or outcomes than the contemporary study (temporal bias). • This control is challenging when subjective endpoints are used or when all of the necessary endpoints were not previously evaluated or evaluated in different ways. • This control presents a significant challenge in addressing the implications of missing data. Sensitivity and missing data analyses may potentially address some concerns associated with bias.

7.5 Placebo Effect and Other Phenomena

A concern in many clinical outcome studies is that a device may have no actual effect but may still appear to demonstrate effectiveness. To counter this, a placebo device (sometimes referred to as a "sham" device) is used. A placebo device is intentionally designed not to deliver any apparent effect but may nevertheless appear to demonstrate effectiveness. This phenomenon is known as the placebo effect. The placebo effect occurs frequently in studies of pain, function or quality of life and can be quite large. The placebo effect can be observed with objective as well as subjective endpoints, and has been known to last for a period of many months and even years.

There are several well-recognized reasons for the placebo effect.

- **Expectation of benefit** - In a randomized, blinded study, there is an expectation of benefit since a subject could be randomized into either group; this is in contrast to a non-blinded study with a "no intervention" control group in which subjects have no expectation of benefit.
- **Study effect** - Related to the placebo effect is the notion that people tend to behave differently when they know they are being measured in a study. In addition, subjects may receive better or more attentive care in a study. Both of these effects can affect both objective and subjective reported outcomes in any study.

The placebo effect introduces a bias into the simple comparison of improvement from an investigational device versus a control. For this reason, it is desirable to include a placebo control when possible to compare the investigational device to a therapy that is ineffective. If superiority to the placebo can be demonstrated, then it can be inferred that the investigational device is effective. Such studies work best when intervention assignment is blinded to the subjects, investigators and third–party evaluators.

While use of an active control does not allow direct measurement of the placebo effect, in cases where the placebo effect can be assumed to be comparable in both intervention groups, it does allow for adequate comparison of the relative safety and effectiveness between the two groups. Unfortunately, in randomized studies with an active control, there can be a different size of the placebo effect in each group which is approximately proportional to the "ritual" associated with the test procedure (e.g., open surgery has a larger placebo effect than taking an oral pill).

In diagnostic clinical outcome studies, when clinicians at different sites may have different standards of care there can be concern about the interpretation of study results. The closest approximation to a placebo controlled study would be one in which the clinicians are unaware of which group their subjects are in until after the device is used, in order to minimize changes in behavior relative to the standard of care.

There are other related phenomena that can make interpretation of the results of a study difficult.

- Regression to the mean - For any measurement on a subject that has an element of randomness, if that subject has an extreme measurement on entry into a study (e.g., as required by the study's inclusion criteria) then subsequent measurements on that same subject will tend to be closer to the overall mean. So, as a purely statistical phenomenon, subjects who are initially "sicker" will tend to improve more.
- Increased medical attention to subjects in a clinical trial may lead to their improvement.
- Spontaneous remissions - Some subjects may heal naturally or no longer exhibit symptoms during the course of a study.

7.6 Non-Comparative Clinical Outcome Studies

Some clinical outcome study designs are not well-controlled studies since they do not use concurrent (or historical) controls and hence have no direct comparator.

7.6.1 Single-Group Study with Objective Performance Criterion (OPC)

An Objective Performance Criterion (OPC) refers to a numerical target value derived from historical data from clinical studies and/or registries and may be used in a dichotomous (pass/fail) manner by FDA for the review and comparison of safety or effectiveness endpoints. It is important to point out that there are currently very few validated OPCs. An OPC is usually developed when device technology has sufficiently matured and can be based on publicly available information or on information pooled from all available studies on a particular kind of device. An OPC needs to be carefully constructed from a prior meta-analytic review of all relevant sources, and a subject-level meta-analysis is preferred. An OPC will tend to have greater validity if it is commissioned or adopted by a medical or scientific society or a standards organization or is described in an FDA guidance document. An OPC typically cannot be developed by a single company using only their data or based on their review of relevant scientific literature, nor is an OPC typically developed unilaterally by FDA. It is also important to note that an OPC can become obsolete over time as technology matures and improves. Sponsors wishing to utilize an OPC for a pivotal clinical study should have discussion and concurrence with FDA staff prior to study initiation.

7.6.2 Single-Group Study with Performance Goals (PG)

A performance goal (PG) provides a level of evidence that is inferior to an OPC. A PG refers to a numerical value (point estimate) that is considered sufficient by FDA for use as a comparison for a safety and/or effectiveness endpoint. In some instances, a PG may be based on the upper (or lower) confidence limit of an effectiveness and/or safety endpoint. Generally, the device technology is not as well-developed or mature for use of a PG as for an OPC, and the data used to generate a PG is not considered as robust as that used to develop an OPC. A PG might be considered for challenging patient populations or if there is no clinical equipoise for any control. A PG will tend to have greater validity if it has been accepted or developed by a medical or scientific society or a standards organization or is described in an FDA guidance document. It is not generally recommended that a PG originate with a sponsor or be developed unilaterally by FDA for a particular submission. Like OPCs, PGs can become obsolete over time as technology improves and as additional knowledge about the performance of the device is learned. Sponsors wishing to utilize a PG in a pivotal clinical study should have discussion and concurrence with FDA staff prior to study initiation.

PGs need to be used with great care. In particular, an important question to ask is whether there is convincing evidence that any device that achieves a performance goal for safety (or effectiveness) would in fact successfully demonstrate such safety (or effectiveness) in a well-controlled investigation. Achievement of (or failure to achieve) a PG does not necessarily lead to immediate acceptance (or rejection) of the study results. In some cases, the study results need to be explored more qualitatively if they are mixed or if unusual signals within the results are found. FDA might present PMAs using PGs to the relevant advisory panel to obtain outside scientific counsel on interpretation of study results.

7.6.3 Observational Studies or Registries

Examining clinical databases to compare therapeutic effect is fraught with bias. Whereas randomization in clinical trials prevents assignment of therapy based on prognosis, there is no such assurance of this kind of bias control in observational studies and registries that contain clinical outcomes. There are examples in the literature where the outcome from randomized clinical studies differs significantly from what had been reported in observational studies. One explanation for the discrepancy is that intervention assignment in the observational studies may have depended on the subjects' prognoses.

Other designs used in epidemiological research may call for cases with one or more control subjects selected based on matching important covariates. Matching may be problematic because selected cases may be disproportionately chosen from a subset of the overall target population and thus the controls may not also be representative of the target population. This type of observational study is not recommended in a pre-market study, whether diagnostic or therapeutic. However, it can sometimes be useful in post-market studies where the association of a particular event with a specific device could be made.

Meta-analysis

The use of meta-analysis to attempt to demonstrate the safety and effectiveness of a medical device without generation of new clinical data introduces potential bias because studies with

insignificant results or poor outcomes are typically not published. In the rare instance where this study design may be useful, it is critical to employ accurate statistical methods and have predetermined strict quality control for inclusion and rejection criteria for selecting published literature studies to minimize selection bias. A well-accepted methodology for meta-analysis is to identify the criteria for selection of studies (such as randomized clinical studies) for inclusion into the meta-analysis before any analysis is attempted. This approach could be termed a prospective meta-analysis. However, a significant flaw is that the majority of publications do not include subject-level data or sufficient details to allow for independent analysis of the data within each study. Other common concerns include inconsistent inclusion/exclusion criteria across studies, significant differences in the definition of endpoints and differences in the length of follow-up of subjects. It is important to note that meta-analysis should only involve studies of the version of the device the sponsor wishes to market.

7.6.4 Literature Summary

Literature summaries can include well-documented case histories conducted by qualified experts, and reports of significant human experience with a marketed device.

In these reports no new clinical data are generated, but they differ from meta-analyses in that no new analyses are performed. A PMA that includes literature summaries may depend on the analyses that were conducted in the selected published literature, and potentially on the well-documented experiences of specific study investigators. These reports are rarely useful for demonstrating effectiveness as there are even more significant limitations than the use of a meta-analysis.

7.7 Diagnostic Clinical Outcome Studies

For diagnostic devices, the pivotal clinical investigation is often a diagnostic clinical performance study (see Section 8); however, sometimes a clinical outcome study is needed. In a diagnostic clinical outcome study, the diagnostic device result is used during a treatment or management intervention. Device performance is assessed in part by the intervention's effect on a subject's outcome. Diagnostic clinical performance may also be evaluated. Clinical outcome studies can be appropriate if, for example, diagnosis and treatment of diseases or conditions are performed at the same time (e.g., as in some endoscopy procedures), or clinical benefit (improvement in clinical outcome) from accurate diagnosis is not clear. Interventions that are needed solely to collect a specimen, but for which the diagnostic result is not used to determine management in the study, are not considered diagnostic clinical outcome studies in this guidance.

Safety and effectiveness are measured by either appropriate clinical endpoints, diagnostic performance, or both. To fully evaluate safety and effectiveness, a control group is sometimes needed in which the diagnostic result is not used by the clinician. Parallel group or paired designs can be appropriate for comparing the investigational and control groups.

In some controlled diagnostic clinical outcome studies, the clinician cannot be blinded to whether the subject is in the investigational or control group, since the clinician knows if s(he) is using the diagnostic result or not. However, whenever possible, the clinician evaluating the clinical endpoint should be blinded to which group the subject is in.

7.8 Advantages and Disadvantages of Some Clinical Outcome Studies

Determination of an appropriate study design for a given device and desired intended use is dependent on many factors, including characteristics of the device, conditions of use, existence of alternative interventions (or diagnostic tests) for the same intended use, existence of adequate warnings regarding use of the device, and extent of experience with the device. It is also important to consider the ultimate desired labeling claims and directions for use, since the study needs to provide sufficient level of evidence to support the labeling. FDA recommends that sponsors select the study design for an investigational device likely to provide the necessary evidence to demonstrate a reasonable assurance of device safety and effectiveness for its proposed intended use, given the specific constraints and characteristics of the particular device type.

Some study designs have the potential to provide a higher level of evidence than others. Choice of a study design that provides a lower level of evidence may require justification that the design is appropriate, and would adequately control potential biases in a manner to support the intended use. Whenever a sponsor believes it is not appropriate or necessary for a clinical outcome study to be well-controlled, randomized and/or blinded, the sponsor should explain why the possible biases can be ignored. The more that a study is designed to minimize bias, the stronger the level of evidence will be (with everything else being the same).

The following sections describe the advantages and disadvantages of study designs common in clinical outcome studies.

7.8.1 Randomized, Double-Blinded, Controlled, Parallel Group Clinical Study

This study design is generally preferred when contemplating a parallel study design, as it can usually provide the strongest level of scientific evidence and the least amount of bias. Double-blinded indicates that the intervention assignment is not known to the subject or the study staff (including the investigator or any third-party evaluator(s)). This study design provides the highest level of assurance that the subject populations in the investigational and control groups are comparable and avoids systematic differences between groups with respect to known and unknown baseline variables that could affect both safety and effectiveness outcomes. However, there are devices for which this design is neither feasible nor practical.

The control chosen for this study design could be active or placebo (see Section 7.5). Deviation from this study design is especially problematic in situations where there is a possible placebo effect, or when subjective outcome measures are used as study endpoints. While use of a placebo control may be desirable since such a design can provide direct evidence of the benefits and risks of the investigational device, it is often problematic to deprive subjects in the control group of a therapy. Therefore, the choice of an active or placebo control may depend on ethical and practical considerations. When considering an

active control, an important consideration is whether to design the study to demonstrate superiority or, in the case of comparison to an active control, non-inferiority.

7.8.2 Randomized, Subject as Own Control, Paired Clinical Study

In such a study design, the subject could be treated with both the investigational and control interventions at the same time. Examples include situations in which one half of the face is treated with the investigational device and the other half is treated with the control intervention. In this design, the assignment of intervention is randomized (e.g., side of face). This study design is possible when the device effect is only evident locally. It is impossible to evaluate and differentiate systemic safety or effectiveness outcomes when using this study design. The advantage of this study design, when used appropriately, is that the effects of both interventions are measured in the same subject and the variability is smaller so a smaller sample size may be required.

Another type of such a study design is a two-group cross-over design study, where each subject receives the investigational and control interventions sequentially, with a randomly assigned order. Similarly, such a design allows the comparison of the performance of the investigational device and control intervention for each subject. However, with this design one needs to assume that the effects of the first intervention will not carry over into the second intervention period. When this assumption is not appropriate, a longer period between interventions (a "washout period") may have to be incorporated into the study.

7.8.3 Randomized, Non-blinded Study with Concurrent Control (Active, Placebo or "No Intervention")

The primary difference between a randomized, non-blinded study with concurrent control and the study designs in 7.8.1 and 7.8.2 is incomplete blinding or absence of blinding. Incomplete blinding refers to instances where the subject, the investigator or the third-party evaluator is not blinded. When no one is blinded, the study is often referred to as an open-label study. As discussed above, in comparative clinical studies, bias can be minimized if the subjects, investigators, and third-party evaluators are blinded to the intervention assignment. However, with an active or a placebo control, it may not always be possible to blind the subjects or the investigators, and sometimes it may even be a challenge to blind the third-party evaluators (e.g., the investigational device and the device serving as the active control have completely different appearances on imaging).

In instances where the control is "best medical management" or a "no intervention" control, the study is usually non-blinded to both the subjects and to the investigators. Consequently, every subject in the control group knows that he or she is not receiving the investigational device. This knowledge often creates a bias of unknown size.

If study participants are not blinded, it is very difficult to assess the size of the resulting bias, and it can threaten the scientific validity of an otherwise solid study, even when a truly objective endpoint is used. In instances where blinding of any or all of the study participants (subjects, investigators, evaluators) is not possible, a detailed rationale and explanation of proposed means to address concerns related to bias should be provided to FDA.

7.8.4 Non-Randomized Study with Concurrent Control (Active or Placebo or "No Intervention")

In a non-randomized design with a concurrent active control, subjects and investigators are not blinded to the intervention assignment. Consequently, this study design suffers from all the drawbacks of a randomized, non-blinded study with concurrent control design. In addition, because there is no randomization and each subject receives only one of the possible interventions, there is a very real possibility of a bias with unknown size due to intervention assignment.

This design is generally not recommended since it is as labor intensive as a randomized study, but introduces more biases due to likely differences in the groups, sites, and investigators, including unmeasured, but likely confounding, differences. Even if there appears to be a balance between the two intervention groups for the study overall, there is likely no balance for each participating investigator such that there may be an investigator-by-device interaction, in which the advantage of the investigational device appears to differ by investigator.

7.8.5 Single-Group Study Compared to Baseline

In many therapeutic studies, a very important consideration is that although it may be tempting to use a subject's baseline status as a control, it is usually advisable to also have a randomized group with an active or placebo control (or even a "no intervention" control). Such a randomized group in a blinded study will provide a much more stringent control and avoid placebo effect bias as well as temporal bias.

7.8.6 Single-Group Study with Historical Control or Information

A single-group study with a historical control or some historical information may be conducted when a device technology is well developed and the disease of interest is well understood.

- Comparison to a historical control group with subject-level data available:

  If subject-level data including all important variables for each subject in both the historical and current studies are available, it is at least possible to make some statistical comparisons. The challenge is in demonstrating that the historical control is comparable to the group in the current study. It may be possible to use a propensity score model to assess the comparability of the two groups after the current study has been completed; however, there is a significant risk that in the end the data may not be comparable. There is no way to assess comparability until the subjects are enrolled and baseline collected and analyzed so this approach can be risky.

  The obvious bias inherent in the use of a historical control is temporal bias, since the groups are not concurrent. This separation in time introduces concerns about the comparability of the two intervention groups as well as concerns that the practice of medicine has likely changed with resultant changes in the target subject population and expected outcomes. Thus the disadvantage of this design is that the subject outcomes in

a historical control may not be discernible or applicable to the current population being targeted.

- Comparison to an OPC or PG derived from historical information

  If a historical control group is not available, the performance of a device may be evaluated through a comparison to a numerical target value, OPC or PG, pertaining to a safety or effectiveness endpoint. Such a study design shares all of the challenges and limitations of comparison to a historical control. In addition, there is no independent way to assess how comparable the current group may be with the historical groups from which the OPC or PG is derived, and it is impossible to quantify the bias.

  Since there is no control group involved in such studies, comparison to an OPC or PG cannot demonstrate either superiority or non-inferiority.

7.9 IDE Application Considerations

For clinical outcome studies, a sponsor's IDE application must, under 21 CFR 812.20, include the details of the proposed study design and a rationale for the study design chosen. The sponsor should explain why the proposed study design is sufficiently robust. A discussion of any alternative study designs with less potential for bias, which had been considered but rejected, would be helpful for efficient review of the proposed study design.

8 Diagnostic Clinical Performance Studies

For diagnostic devices, the pivotal clinical study is often a diagnostic clinical performance study. In such a study, diagnostic clinical performance of the diagnostic device is characterized by measure(s) that quantify how closely the diagnostic device output is associated with a clinical reference standard[13] (defined in Section 8.2) that is used to assess subjects for the target condition. The choice of appropriate clinical performance measure(s) depends on the intended use of the device, the nature of the diagnostic device output and the clinical reference standard. The goal of a diagnostic clinical performance study is to establish device diagnostic clinical performance and to support a favorable risk/benefit analysis related to the clinical performance of the device in the target population.

Diagnostic clinical performance studies are often preceded by bench studies, non-pivotal clinical studies and/or studies that characterize various aspects associated with device measurement (e.g., analytical studies for IVDs).[14] For example, consider an *in vitro* diagnostic device for detecting

[13] The independent assessment of subjects for the target condition is sometimes referred to as a "gold standard," "ground truth" or "standard of truth"; however, these terms can have other meanings as well so we refrain from using them in this guidance.

[14] Detailed discussion on studies to evaluate the quality of device measurement is beyond the scope of this document. CDRH recognizes numerous standards and guidelines for the evaluation of the quality of device measurement studies. A list of standards recognized by CDRH can be found on FDA's website at

high risk strains of human papillomavirus (HPV) DNA to predict cervical cancer in women 30 years or older with a normal Pap test result. The ability of the HPV test to predict cervical cancer (target condition) is assessed in a clinical performance study, while the ability of the device to measure the high risk strains of HPV DNA (measurement of interest) is assessed in analytical performance studies (including, but not limited to, assessment of measurement bias, precision, limits of quantitation and detection, linearity, interferences, and carry-over).

The safety and effectiveness of a diagnostic device are often not separable. When the result reported by a diagnostic device is incorrect (e.g., the result is either misclassified as a false positive or false negative) or misinterpreted, subjects can be harmed by subsequent inappropriate management or by psychological trauma. For dichotomous (present/absent) target conditions, the safety and effectiveness of a diagnostic device is often captured by its ability to correctly identify the presence or absence of the target condition. In addition to misdiagnosis, a diagnostic device may sometimes introduce safety concerns for subjects during specimen collection or device use. For example, it may expose subjects to radiation or other forms of energy, result in the use of invasive procedures, or result in the administration of therapeutic products. In these situations, risk to the subjects in a diagnostic clinical performance study would also be considered when evaluating the appropriateness of a study design and in determining whether the device poses a significant risk to subjects in the study.[15]

Critical factors affecting the design of clinical investigations for a diagnostic clinical performance study may include the importance of the intended use of the device to define the study design, choice of appropriate study population, and mitigating specific sources of bias. These factors are discussed further in the subsections that follow.

8.1 Consideration of Intended Use

Intended uses for diagnostic devices vary considerably, as do the types of results provided by these devices. Therefore, the designs of diagnostic clinical performance studies vary accordingly. Many diagnostic devices attempt to classify subjects according to presence, absence, or stage of a specific target condition or disease. Other diagnostic devices provide a measurement of a biological quantity (e.g., viral load, blood glucose level, or retinal thickness) as an aid in diagnostic evaluation or for subject monitoring.

The pivotal diagnostic clinical performance study should support the intended use of the diagnostic device. A diagnostic device might be intended as a stand-alone diagnostic, to replace an existing diagnostic device or procedure, or, it might be intended to be used in conjunction with other information (sometimes through use of an algorithm) to assess a subject's target condition. Alternatively, a diagnostic device might provide adjunctive diagnostic information

---

http://www.fda.gov/MedicalDevices/DeviceRegulationandGuidance/Standards/default.htm, accessed February 2012.

[15] Refer to "Information Sheet Guidance For IRBs, Clinical Investigators, and Sponsors Significant Risk and Nonsignificant Risk Medical Device Studies," available at http://www.fda.gov/downloads/RegulatoryInformation/Guidances/UCM126418.pdf; accessed March 2011

(e.g., the additional information does not over-rule recommendations based on an existing device or procedure).

In designing a diagnostic performance study, the device should be evaluated in the context of its intended use, including the following, as applicable:

- what the device measures or detects;
- what the device reports;
- cell, tissue, organ, part, or system examined;
- specimen source(s), specimen type(s), and specimen matrix(-ces)
- how the device is used (per instructions for use);
- when the device is used (conditions of use);
- by whom the device is used (operator or target user);
- for what (target condition);
- on whom (target population) device is used.

Changes to the way an approved or cleared device is used may warrant a discussion with FDA to determine what type of evaluation is needed.

8.2 Clinical Reference Standard for the Target Condition

Ideally, characterization of the clinical performance of a diagnostic device requires independent knowledge of the subject's true status with respect to a target condition (refer to Section 4.1). In a diagnostic clinical performance study, each subject is assessed for the target condition, using a clinical reference standard.[16] A clinical reference standard, for regulatory purposes, is the best available method for establishing a subject's true status with respect to a target condition (e.g., the pathological result of a biopsy to determine the presence of breast cancer). It can be a single method or a combination of methods and techniques, including clinical follow-up, but it should not consider the investigational device output.

As described in other FDA guidance,[17] the "best available method" is established by evidence from current practice within the medical, laboratory, and regulatory communities. Sometimes one of several possible methods is chosen. In other situations a composite is constructed or a

---

[16] This definition does not restrict the target condition to be dichotomous (present/absent); otherwise, this definition is identical to that for "reference standard" as described in FDA's "Statistical Guidance on Reporting Results from Studies Evaluating Diagnostic Tests," March 13, 2007, available at: http://www.fda.gov/MedicalDevices/DeviceRegulationandGuidance/GuidanceDocuments/ucm071148.htm, and Bossuyt et al and diagnostic accuracy criteria (see CLSI Harmonized Terminology Database, at http://www.clsi.org/Content/NavigationMenu/Resources/HarmonizedTerminologyDatabase/Harmonized_Terminolo.htm; accessed February 2011).

[17] Guidance for Industry and FDA Staff: Statistical Guidance on Reporting Results from Studies Evaluating Diagnostic Tests (March 13, 2007), see http://www.fda.gov/MedicalDevices/DeviceRegulationandGuidance/GuidanceDocuments/ucm071148.htm, and Statistical Guidance on Reporting Results from Studies Evaluating Non Diagnostic Medical Devices, see http://www.fda.gov/MedicalDevices/DeviceRegulationandGuidance/GuidanceDocuments/ucm106757.htm.

consensus of a panel of experts is used. Before the study begins, the clinical reference standard should be pre-specified. It should also be described in a fully detailed manner, including all of the information needed to determine its result. To the extent possible, this requisite information should be recorded on every subject in the study. Ideally, the reference standard result should be determined for every subject.

Sometimes, a clinical reference standard does not exist, is not available, or cannot be used in a clinical study due to its invasive nature.[18] In such cases a comparator that is not a clinical reference standard (a non-reference standard) may be specified and used, and these results compared to the investigational device output. For example, investigational hepatitis B assays are typically compared to the results of multiple FDA-approved HBV marker assays rather than to a clinical reference standard for hepatitis B virus. Sponsors should consult with FDA prior to planning a study that does not use a clinical reference standard to ensure that the study will support the intended use of the device. Diagnostic clinical performance studies that do not use a clinical reference standard to assess the target condition are called agreement studies. In agreement studies, the "correctness" of the diagnostic device cannot be estimated directly; an investigational device may agree with the non-reference standard comparator, but neither may correspond to what would have been the result of the clinical reference standard had it been available. Concerns regarding the interpretation of agreement measures are discussed in in other FDA guidance in the context of comparing results from two diagnostic devices.[19]

Since a clinical reference standard can evolve over time as knowledge increases and medical systems advance, measures of diagnostic clinical performance should be reported with, and interpreted in the context of, the clinical reference standard that is used. When a test developer believes that there will be discordances between the new device and the clinical reference standard due to errors in the clinical reference standard, sponsors should consult with FDA prior to conducting such a study.

8.3 Study Population for Evaluation of Diagnostic Performance

Sites from which subjects or samples are chosen for studies that support the intended use of the device should be representative of the types of sites where the device is intended to be used. Subjects or samples should also represent the proposed target population. Estimates of overall performance from non-representative sites or subjects may suffer from selection bias. The actual method of selecting subjects or samples for a study must be specified in the study protocol (See 21 CFR860.7(f)(1)(a)). Different selection methods along with advantages and disadvantages are described earlier in Section 6.3.

---

[18] An example is the difficulty of using laparoscopy/biopsy to establish presence or absence of ovarian cancer in a population of apparently healthy subjects.

Guidance for Industry and FDA Staff: Statistical Guidance on Reporting Results from Studies Evaluating Diagnostic Tests (March 13, 2007), see http://www.fda.gov/MedicalDevices/DeviceRegulationandGuidance/GuidanceDocuments/ucm071148.htm.[20]

Informed Consent for In Vitro Diagnostic Device Studies Using Leftover Human Specimens that are Not Individually Identifiable, April, 2006, at https://www.fda.gov/downloads/MedicalDevices/DeviceRegulationandGuidance/GuidanceDocuments/ucm071265.pdf.

Subjects enrolled in the study should represent the target condition spectrum. When the subjects enrolled do not match the target condition spectrum, estimates of diagnostic clinical performance are subject to a spectrum effect. For example, if only subjects from the extreme ends of the target condition are sampled (e.g., either healthy normal subjects or subjects with advanced stage disease), then performance can appear to be better than it truly is. This is because subjects in the middle of the target condition spectrum that are omitted tend to be more difficult to diagnose correctly.

Sometimes the target population includes subjects with a rare condition such that recruiting subjects with the rare condition can be difficult and expensive. Designs that over-represent the rare condition in the subject population, compared to the proportion in the target population, might sometimes be appropriate. However, estimates of overall performance from such a design may have the potential for bias, so this potential bias should be considered in the Statistical Analysis Plan.

8.4 Study Planning, Subject Selection and Specimen Collection

Diagnostic devices may test a subject directly to yield subject specific data, or may test specimens collected from subjects. Specimens may be collected and tested immediately, or under certain circumstances, may be collected and stored prior to being tested. Specimens or subject data are said to be prospectively obtained when a pre-specified protocol is used, and only specimens or subject data from subjects meeting the protocol criteria are obtained. Specimens that are obtained from collections that are assembled without a pre-specified intent of use or were part of a pre-specified protocol for a different study, e.g., biobanks, are not considered to be prospectively obtained but can be used in retrospective studies. Similarly, subject data collected from devices that test a subject directly (e.g., ECG, EEG, image set) can be stored for later selection and analysis; this is another type of retrospective study.

In a prospectively planned study a pre-specified protocol is used. Such a protocol pre-specifies study design, including inclusion/exclusion criteria, method of subject recruitment and selection, testing protocol, and analysis methods to be used. Subjects meeting inclusion/exclusion criteria are selected over the study duration. Well-executed prospective planning can help ensure that the study population provides an adequate representation of the target population so that the study provides evidence to support the intended use.

In certain situations it may be acceptable to supplement a prospective study with bank specimens or previously collected subject data (e.g., when the target condition is very rare and it is very difficult to obtain a sufficient number of subjects with the target condition in a prospective manner), or to use only banked specimens or subject data to assess the performance of the device, provided that the potential for bias and other concerns discussed in this guidance can be adequately addressed. When bank specimens or previously collected data or images are added to a study, the person performing the test or interpreting the test results should not be able to differentiate the added specimens or data from those obtained prospectively.

Retrospective selection of previously archived specimens or data can introduce additional issues. In some retrospective study designs, investigators search for subjects with available data, specimens, images, or other stored media or information used by the device. Examples of retrospective selection include going to a tertiary care center to obtain specimens or using

registry data from previous studies that involve long term follow-up. In general, for specimens or subjects selected in a prospective manner, the selection process is under the control of the investigator(s). In contrast, retrospective subject or sample selection may be limited to, for example, subjects with stored specimens and with a clinical reference standard result. One concern is that the retrospectively selected specimens or subject data may be non-representative of the target population (e.g., retrospective specimens or data may represent only extreme cases of the target condition). Another concern is that a convenience sample of retrospective specimens may confound diagnostic device clinical performance with tissue storage or handling, or covariates predictive of the target condition. The use of retrospectively obtained specimens and subject data without prospective planning thus raises a number of possible issues, including the purpose for which the specimens or subject data in the archive were collected (with respect to representativeness to the current target population), possible degradation of specimens or change of technology used to acquire and store subject data over time, and non-random depletion of archival specimens. For more information see FDA's guidance concerning informed consent using leftover human specimens.[20] Sponsors should consult with FDA to determine if available specimens or subject data are appropriate to support a diagnostic device's intended use.

When designing any type of diagnostic clinical performance study, protocols for acquisition of specimens or subject data are important. For IVDs, subject preparation, specimen collection, storage and handling procedures are critical components that should be fully described in the study protocol. For diagnostic devices other than IVDs, the measurement or data acquisition procedure is a critical component. The study protocol should describe how a subject measurement or result should be acquired including specific instructions (e.g., specific stimulation procedure, specific electrode placement, specific subject condition while data are acquired).

8.5 Diagnostic Clinical Performance Comparison Studies

The goal of a diagnostic clinical performance study is to establish the performance of an investigational device. Comparative studies that compare the diagnostic clinical performance of an investigational device with the diagnostic clinical performance of an established device or method are only possible when a clinical reference standard is used. It is recommended that sponsors designing such studies consult with the appropriate FDA review division at the design stage.

When a clinical reference standard is unavailable, the investigational device is sometimes compared with another device in an agreement study. A very high level of agreement can indicate that the investigational device is non-inferior to the established device. However, a high level of agreement is only meaningful if the established device is already known to have an acceptable level of performance.

---

[20] Informed Consent for In Vitro Diagnostic Device Studies Using Leftover Human Specimens that are Not Individually Identifiable, April, 2006, at https://www.fda.gov/downloads/MedicalDevices/DeviceRegulationandGuidance/GuidanceDocuments/ucm071265.pdf.

8.6 Blinding (Masking) in Diagnostic Performance Studies

Clinical studies for diagnostic devices can involve multiple evaluations and users/readers. For instance, a clinical study for diagnostic performance could involve the user/reader of the investigational device, a person obtaining the clinical reference standard result, and sometimes a user/reader of an established device used in a comparison study. The user of the investigational diagnostic device should not be aware of (and so should be blinded to) the result from the clinical reference standard or the results from other diagnostic evaluations, and vice versa.

8.7 Skill and Behavior of Persons Interacting with the Device (Total Test Concept)

Use of diagnostic devices often requires multiple activities performed by persons with differing levels of training or skills, e.g., layperson, phlebotomist, laboratory technician, pathologist, radiologist. These activities might include collecting and preparing samples, positioning a device on the subject, and interpreting visual outputs. When the task requires skill through training, subject knowledge, aptitude for reading images and/or wave forms, and experience, differences in human performance are not unusual and can affect the device performance. Therefore, when evaluating the clinical performance of a diagnostic device, the clinical study protocol should account for variability in the performance of persons interacting with the device. Sometimes additional studies are necessary to examine specific device performance in the hands of different persons interacting with the device. The sponsor may need to document the training given to persons in the study and provide training materials to be marked for review by FDA.

In clinical performance comparison studies of two diagnostic assessments applied to the same subject when the assessments being compared are read or interpreted by the same trained person, a reading order bias can be introduced. In such studies, since readings from the various outputs (e.g., images, slides) cannot be done at the same time, they are done in some pre-specified sequence. When two different assessments are made on the same subject or sample by sequential reading, the knowledge of one assessment may influence the other assessment. The effect on the second assessment may also be potentially confounded by simply having additional assessment time. One way to mitigate reading order bias is to have a long period of time between assessments ("wash-out" period) to eliminate reader memory of the first assessment. Other mitigations are possible and we recommend sponsors consult with the FDA review division for further information.

If the context in which a diagnostic clinical performance study is conducted is not reflective of medical practice, clinical performance estimates can be subject to bias. For example, the prevalence of a target condition may vary according to a given setting and may therefore affect estimates of the diagnostic device performance. Readers/interpreters may consider investigational device results to be positive more frequently in settings with higher disease prevalence, thereby also affecting estimates of diagnostic device performance.
Sponsors should consider how these types of bias can affect estimates of the performance of their device, and attempt to ensure that they are controlled as well as possible.

8.8 Common Types of Bias in Diagnostic Clinical Performance Studies

In Section 6.2, the importance of minimizing bias was discussed and some types of bias that can affect diagnostic clinical performance studies have been introduced in previous sections. These and some additional types of bias (not an exhaustive list) are further described below. These should be recognized and mitigated or eliminated where possible.

- **Selection Bias:** Systematic error in choosing a study population, so that it is not representative of the intended use population.
    - Spectrum Bias: A type of selection bias in which a study of diagnostic clinical performance fails to account for the variation or heterogeneity of the test performance across population subgroups. Bias may occur when performance varies across subgroups of the intended use population and the study does not adequately represent all subgroups.

- **Verification bias:** Typically, a single clinical reference standard is applied to all subjects in the study. When the clinical reference standard is applied to only a subset of study subjects then performance estimates have to be adjusted accordingly or they will have the potential for verification bias. For example, verification bias is produced when frequency of application of the clinical reference standard depends on the investigational device results (e.g., because the clinical reference standard is less frequently applied when the investigational device result is negative).

- **Disease progression/regression bias**: Disease progression/regression bias occurs when (a) the results of the investigational device and the clinical reference standard are not collected on the same subject at the same time and (b) spontaneous recovery or progression to a more advanced stage of disease takes place in the interim. For an investigational device that determines a present state of health, the investigational device and the clinical reference standard or other diagnostic devices used in the study should be applied to a subject at nearly the same time.

- **Lead-time bias**: Subjects who are screened with a diagnostic device can falsely appear to benefit from diagnostic testing because of a bias known as lead-time bias. Subject survival from the time of testing may be no better when a test result is known than when it is not, but can appear to be better because earlier detection adds to the survival time relative to detection at a later time. This can be a particular problem when screening intervals differ across areas of clinical practice.

- **Length-time selection bias (survivor bias)**: Subjects who have a target condition for a long period of time are more likely to be included in clinical studies that those who have the target condition for a short period of time. These subjects usually have a better prognosis and may not represent subjects in the target population. As a result, estimates of survival times for study subjects can be longer than survival times in the target population.

- **Extrapolation bias:** The conditions or characteristic of the population in the study are different from those in which the test will be applied.

- **Reading order bias:** When comparing two or more diagnostic assessments, the reader's interpretation is affected by memory of the results from the competing assessment.

- **Bias due to lack of independent evaluation**: Performance of a diagnostic device is likely to be inflated if it is evaluated in a study that was used to develop or refine one or more aspects of the device. Once all aspects of a diagnostic device have been finalized, the performance of the device should be evaluated in a new study that is independent of any preliminary studies that were used to develop the device. For example, the diagnostic clinical performance of a device that includes an algorithm that is developed ("trained") using one data set should be evaluated ("tested") using different, independent data in order to avoid the potential for bias ("overfitting").

9 Sustaining the Quality of Clinical Studies

This section provides information on plans and techniques that sustain the level of evidence of clinical studies. It applies to both clinical outcome studies as well as diagnostic clinical performance studies.

The evidence generated by a clinical study permits scientifically valid evaluation of the safety and effectiveness of the medical device. A key factor that contributes to the generation of this evidence is the selection of study design, which will hopefully also reduce the sources of bias. The use of sound scientific methods to carefully conduct the study and analyze the data should maximize how informative the study will be. Poorly-conducted or inappropriately-analyzed studies reduce the ability to rely on the evidence generated to evaluate the safety and effectiveness of the device. Quality study conduct helps to sustain the level of evidence of the clinical study for the device in question.

Sponsors are responsible for ensuring proper monitoring of the investigation (21 CFR 812.40). Plans and techniques should be put into place at the design development stage to optimize the reliability and usefulness of data and information generated in the clinical study. These plans and techniques should address the various aspects of the clinical study, such as, handling clinical data, conducting the clinical study, planning the analysis strategy, and prospectively accounting for changes that may occur during the course of the study. These aspects are further discussed below.

9.1 Handling Clinical Data

Title 21 CFR 812.150(a) requires sponsors to prepare and submit complete and accurate reports of certain information and 21 CFR 812.140(a) requires accurate, complete and current records of certain information, such as each subject's case history. FDA strongly encourages study sponsors to establish at the onset of the clinical study both a data management plan and a

comprehensive training program for handling clinical data that follow the principles of Good Clinical Data Management Practices (GCDMP).[21] With more clinical studies being conducted, reported, recorded, and analyzed in electronic environments, a data management plan is critically important to establishing the level of evidence, minimizing bias, and generating reliable, useful data. While FDA regulations do not require submission of a data management plan to FDA for review, FDA encourages sponsors include an executive summary of the data management plan in the clinical study protocol. It would also be in sponsors' interests to request a Pre-Submission meeting that includes staff from the Division of Bioresearch Monitoring to discuss best practices for handling clinical data.

Study data should be collected in a consistent format and structure so that they may be easily interpreted, understood, and evaluated. Maintaining an efficient standard method of data collection across studies, sites and investigators can help to ensure high-quality data across the studies. Further, it can facilitate the interpretation of protocol designs across studies by comparing the associated metadata. Utilizing standard vocabularies and requirements for data collection is encouraged as it will optimize data collection and improve data quality and predictability.[22]

9.2 Study Conduct

FDA carefully reviews progress reports for clinical studies conducted under an IDE and has the authority to disqualify any investigators from further participation in clinical studies if they do not conduct studies in a manner consistent with GCPs. *See* 21 CFR 812.119. Further information on FDA's regulations regarding the conduct of clinical studies, including guidance on GCPs and adequate human subject protection, is available.[23]

FDA's guidance on Data Monitoring Committees (DMCs) provides information to assist clinical study sponsors in determining when a DMC may be useful for study monitoring as well as information on how a DMC should operate.[24]

Following these recommendations for planning and managing a clinical study will improve the quality of data from the study and increase the likelihood that an investigation will be adequate to support approval of a device:

- The randomization code and procedure should be carefully preserved. If adaptive randomization is used, the algorithms and data used to create the probability assignments should be preserved.

---

[21] See http://www.scdm.org/gcdmp/.

[22] The CDISC (http://www.cdisc.org) and HL7 (http://www.hl7.org) standards groups have more information on data standards.

[23] See http://www.fda.gov/ScienceResearch/SpecialTopics/RunningClinicalTrials/default.htm.

[24] Guidance for Clinical Trial Sponsors: Establishment and Operation of Clinical Trial Data Monitoring Committees, at http://www.fda.gov/downloads/RegulatoryInformation/Guidances/UCM127073.pdf

- The study blind should be strictly maintained and the integrity of the blind should be evaluated. We suggest that sponsors keep a log of perceived unblinding events.
- The study protocol should be strictly followed and all types of protocol deviations, including those deemed minor, should be minimized. The protocol should define the types of deviations that are considered minor or major. All protocol deviations should be reported in detail. An unacceptable rate of major protocol deviations might make it impossible to generalize the study results.
- Study subjects should be consistently and completely followed according to the study protocol. Great effort should be made in the study design and conduct phases to reduce the occurrence and impact of missing data due to subject loss-to-follow-up. For example, the protocol might identify procedures to follow-up missed visits or dropped contacts, including continued safety follow-up on subjects who refuse further treatment or efficacy evaluations. Although statistical techniques may be used to address issues of loss-to-follow-up and missing data, these techniques often employ major assumptions that cannot be fully validated for a particular study. Therefore, the best way to address issues of missing data due to loss-to-follow-up is to plan to minimize its occurrence during the planning and management of the clinical study. Nevertheless, the study protocol should pre-specify appropriate statistical data analysis methods, in addition to sensitivity analyses, for handling missing data.
- Vigilant data monitoring should be conducted to ensure reliable, accurate data and minimize missing data. Clinical research associates or study monitors should be selected based upon training and experience to monitor the clinical study and to ensure the quality, reliability, and integrity of the study including the rights, safety, and well-being of human research subjects. Study monitors should not be involved in the actual conduct of the study (e.g., subject assessment or recording of study data). A clinical quality assurance program should be implemented to ensure that the study is conducted as designed and intended.
- Consistent adherence and/or commitment to optimal clinical care (e.g., medication strategies, use of operators with appropriate training and expertise in use of the device or the control, consistent follow-up procedures and strategies) should be maintained.
- The study data should be carefully protected to prevent biases due to early looks unless explicitly pre-planned in the Statistical Analysis Plan. This also applies to open label studies.
- Measures should be in place to avoid premature discontinuation of the study unless a planned interim analysis or stopping rule is pre-defined in the study protocol or the discontinuation decision is based on safety concerns. Studies can of course be discontinued prematurely with no intention of FDA submission but studies prematurely discontinued for lack of funding will have failed their primary analysis. However, the sponsor would continue to have the obligation of subject safety-related monitoring for such discontinued studies.

- Sponsors must select investigators qualified by training and experience to investigate the device 21 CFR 812.43(a). All study site personnel (clinicians, study coordinators, etc.) should be adequately trained. Training of investigators should be properly documented.

The clinical study design and protocol should include sufficient procedures to address, optimize and mitigate all of the above considerations. If adherence to any clinical study standards (e.g., ISO 14155:2011) is planned, the standards should be indicated in the study protocol. Investigators and study monitors should be trained regarding the proper implementation of such a standard if used.

With respect to protocol deviations, FDA has found that some participating clinical investigators do not follow an approved protocol because they do not agree with some aspects of the study design. FDA encourages study sponsors to engage prospective clinical investigators in discussions throughout the development of the study protocol so that possible issues with the protocol and potential deviations may be resolved prior to the establishment of a final protocol. These discussions may lead to improvements in the study design that otherwise might have resulted in protocol deviations, which would have been problematic for study analysis and poolability of data. In addition, sponsors must obtain from each participating investigator a signed agreement that includes the investigator's commitment to conduct the investigation in accordance with the agreement, the investigational plan, 21 CFR Part 812, and conditions of approval imposed by the reviewing IRB or FDA. 21 CFR 812.43(c)(4)(i).

9.3 Study Analysis

Poorly performed, inappropriate, and/or *post-hoc* analyses may adversely affect the usefulness of the evidence to support the safety and effectiveness of a device. Thus, the study protocol should have a detailed, pre-specified Statistical Analysis Plan that includes plans to evaluate, to the extent possible, key assumptions that were made in the design of the study (e.g., assessment of carry-over effects in a crossover study design, proportionality of hazards in a survival analysis, or pooling analysis across clinical sites or geographic regions). This predefined SAP should be adhered to in analyzing the data at the completion of the study to support the usefulness of the evidence generated by the study.

Unplanned *post-hoc* analyses and deviation from the analysis populations specified in the protocol should generally be avoided. Examples of *post-hoc* analyses include the use of a different statistical analysis without proper justification, change in the primary endpoint, or the use of a subgroup for analysis that was not pre-specified. These *post-hoc* analyses can inflate the experiment-wise type I error rate and endanger the scientific validity of an otherwise well-designed and well-conducted study. The protocol should pre-specify sensitivity analyses to demonstrate that inferences are robust to any known sources of bias. It is also important to critically analyze the impact of missing data on the conclusions drawn from the study.

In some cases, *post-hoc* analyses may complement pre-specified analyses, as long as they are clearly described and interpreted with the appropriate degree of skepticism that comes with this type of analysis.

9.4 Anticipating Changes to the Pivotal Study

In some cases, the results of an interim analysis or the occurrence of adverse safety events could necessitate a change to device design to improve device safety and/or effectiveness during the course of a pivotal clinical study. For example, changes to the device design can be significant enough to require that study subjects treated with different versions of the device be considered as separate strata and analyzed separately, calling into question whether the data can be pooled across strata. A proposal for consideration of the different intervention groups should be discussed with FDA. To reduce the incidence of device design changes late in device development, a sponsor should take advantage of a robust exploratory stage prior to investment in more resource-intensive pivotal studies.

In contrast, changes to the study design midstream might be planned such that the studied subject populations may be pooled. Some adaptations can be planned in advance and built into the study design. Specifically, interval modifications to a study design (e.g., change in sample size, randomization modification) can be prospectively incorporated in a protocol to maintain the statistical integrity of the study either by a Bayesian approach[25] or by various methods for frequentist interim analyses. It is possible to plan an adaptive design in advance that provides for specific modifications to the study depending on results within the study. Such designs should address who will decide whether the modifications will be implemented once the interim results are available (e.g., the DMC an independent statistician). If sponsors are considering a Bayesian or adaptive trial design, they should seek FDA input as early as possible. Adaptations that are not pre-planned can severely weaken the scientific validity of the pivotal study.

In appropriate instances, FDA may also grant approval in stages or with conditions for a subset of the planned subject cohort while certain outstanding questions are answered in parallel with enrollment. The sponsor will be permitted to expand enrollment once an IDE supplement containing the necessary additional information has been submitted to FDA and allowed to proceed. In such cases, the study design may need to provide accommodation for looks at interim safety data prior to allowing enrollment expansion with minimal interruption in the study.

10 The Investigational Plan or Protocol

The investigational plan or study protocol is a written document that provides the detailed plan for the design, conduct and analysis of the clinical study (See 21 CFR 812.25 and 21 CFR 860.7(f)(1)). The protocol should, include the following:

- scientific rationale for the study;
- definition of the subject populations to be evaluated (including the inclusion/exclusion criteria);

---

[25] Guidance for Industry and FDA Staff: Guidance for the Use of Bayesian Statistics in Medical Device Clinical Trials (February 5, 2010) at http://www.fda.gov/MedicalDevices/DeviceRegulationandGuidance/GuidanceDocuments/ucm071072.htm.

- identification of the proposed intended use for the device;
- listing of the study endpoints;
- statement of the procedures (treatments and tests) that will be applied to study subjects;
- a summary of the methods of analysis and an evaluation of the data derived from the study, including any appropriate statistical methods utilized.

It is also recommended that the protocol include, for therapeutic and aesthetic devices, sufficient statistical detail about the statistical analysis of the primary endpoint(s) to provide justification for the sample size calculation. The protocol should include a complete Statistical Analysis Plan that clearly describes the precise strategy to analyze the data. The complete Statistical Analysis Plan can be either included in the protocol or attached as a separate document, but should reference the relevant protocol. In any case the complete Statistical Analysis Plan should be finalized before any outcome data are available in order to preserve the scientific integrity of the study.

Documentation of the rationale for decisions made about the study protocol, especially with regard to the selected clinical study design and the clinical endpoints will facilitate FDA's review of the clinical study by providing an explanation to support the proposed study design and endpoints. The rationale why other endpoints or alternative study designs with less potential for bias were not selected may be helpful for review of the study.

FDA welcomes the opportunity to provide informal advice and feedback during the development of the pivotal study design through the pre-submission process. It is also advisable that investigator input be sought during the study design phase. Clinical data managers play a critical role in providing input into study design and case report form design based on past experiences running similar clinical studies.

11 Glossary

In this glossary, terms are defined according to their specific interpretation as used in this particular guidance.

Active Control Investigation (Active Treatment Control Investigation)

A study that uses an intervention whose effectiveness has been previously established. In a device investigation, the active control could be a device (drug or biological product) approved or cleared for that indication or a surgical procedure.

Aesthetic Device

Device intended to provide a desired change in visual appearance in the subject through physical modification of the structure of the body.

Agreement Study

A diagnostic clinical performance study in which the diagnostic device result is compared with a result that is not from a clinical reference standard.

Benefit-Risk Assessment

The probable benefit to health from the use of a device weighed against any probable injury or illness from such use.

Bias

The introduction of systematic errors from the truth. In a clinical investigation, bias can lead to incorrect conclusions about what the study shows.

Blind (Mask)

A condition placed on an individual or group of individuals to keep them from knowing the intervention (or test) assignment of the subjects or subject specimens. For ophthalmic device studies, the term "blind" to describe this condition is inappropriate.

Clinical Investigation

See Clinical Study.

Clinical Outcome Study

A study in which subjects are assigned to an intervention and then studied at planned intervals using validated assessment tools to assess clinical outcome parameters or their validated surrogates to determine the safety and effectiveness of the intervention.

Clinical Reference Standard (CRS)

Best available method for establishing the true status of a subject's target condition; it can be a single method or combination of methods and techniques including clinical follow-up, but it should not consider the investigational device output.

Clinical Study

Systematic study conducted to evaluate the safety and effectiveness of a therapeutic, aesthetic or diagnostic device using human subjects or specimens (see also Clinical Investigation).

Comparator

A procedure or another medical product that serves to assess the level of performance of the investigational device. Often the comparator is another medical device.

Composite Endpoint

A prespecified combination of several endpoints.

Condition of Interest

See Target Condition.

Concurrent Control

A control based on data collected over the same time period as the investigational device.

Control

A device, drug, biological product, medical procedure, or "no intervention" that is used to compare with the investigational device.

Control Group

In a clinical outcome study, the group of subjects or specimens who receive the control.

Controlled Clinical Study

A clinical study that evaluates the performance of the investigational device by comparison with results from a control group, a comparator device or a clinical reference standard.

Cross-over Design (Cross-over Study)

A study in which subjects receive a sequence of different interventions (or diagnostic tests). In the simplest case of a cross-over design study, each participant receives either the investigational device or the control in the first period, and the other in the succeeding period. When necessary, the two periods are separated by a suitably long "washout" period to mitigate the effect of the earlier intervention(s). The order in which investigational device or control is given to each subject is usually randomized.

Data Monitoring Committee (DMC)

A group of individuals with pertinent expertise that reviews on a regular basis accumulating data from one or more ongoing clinical studies. A DMC may recommend that a study be stopped if there are safety concerns or if the study objectives have been achieved. Also sometimes called a Data Safety and Monitoring Board (DSMB).

Diagnostic Clinical Performance Study

A study in which the performance of a diagnostic device is characterized by measure(s) that quantify how well the device result of a subject is associated with the clinical condition of the subject as determined by a clinical reference standard.

Diagnostic Device

A device that provides results that are intended to be used alone or in the context of other information to help assess a subject's target condition.

Exploratory Stage

A medical device clinical development stage that includes initial development, evaluation, first-in-human and other feasibility studies.

Feasibility Study

A preliminary clinical study to see, for example, if a pivotal study is practical or to refine the study protocol for the pivotal study. A feasibility study is sometimes also called a pilot study.

Good Clinical Practice (GCP)

A standard for the design, conduct, performance, monitoring, auditing, recording, analyses, and reporting of clinical studies that provides assurance that the data and reported results are credible and accurate, and that the rights, safety, well-being, integrity, and confidentiality of study subjects are protected.

Good Clinical Data Management Practices or GCDMP

Current industry standards for clinical data management that consist of best business practices and acceptable regulatory standards.

Historical Control Group

A control group of subjects who were observed prior to the pivotal study. Data collected from this control group is used to compare the performance of the investigational device.

Intervention Assignment Mechanism

Method that assigns the study subjects to investigational or control groups.

Investigational Device

1) An unapproved new device or a currently marketed device being studied for an unapproved use in a clinical investigation or research involving one or more subjects to determine the safety or effectiveness of the device. 2) A device, including a transitional device, that is the object of an investigation, where a *Transitional device* means a device subject to Section 520(l) of the FD&C Act, that is, a device that FDA considered to be a new drug or an antibiotic drug before May 28, 1976 (See 21 CFR 812.3(g) and (h)).

***In Vitro* Diagnostic (IVD) Device**

A diagnostic device that is intended for use in the collection, preparation and examination of specimens taken from the human body.

Learning curve

A graphical representation of the change in the rate of learning in the use of a medical device or, for a surgical implant, in the implantation procedure of the device. It can be measured in terms of the time taken to achieve desired outcomes or in the number of procedures until successful outcomes are assured.

Level of Evidence

The collective level of confidence about the validity of estimates of benefits and harms for any given intervention or diagnostic test.

Medical Device

An instrument, apparatus, implement, machine, contrivance, implant, *in vitro* reagent, or other similar or related article, including any component, part, or accessory, which is (1) recognized in the official National Formulary, or the United States Pharmacopeia, or any supplement to them,

(2) intended for use in the diagnosis of disease or other conditions, or in the cure, mitigation, treatment, or prevention of disease, in man or other animals, or (3) intended to affect the structure or any function of the body of man or other animals, and which does not achieve its primary intended purposes through chemical action within or on the body of man or other animals and which is not dependent upon being metabolized for the achievement of its primary intended purposes. (See Section 201(h) of the FD&C Act.)

Meta-analysis

A statistical synthesis of the data from separate but similar (i.e., comparable) studies, leading to a quantitative summary of the pooled results.

Non-Inferiority Study

Study designed to demonstrate that the safety or effectiveness of an investigational device is not worse than the comparator by more than a specified margin.

Non-Blinded Study

A study in which there is no blinding; also called an open-label study (see also Open-Label Study).

"No Intervention" Control

A control in which no intervention (including a placebo) is used on the subject. In a treatment study, this could also be referred to as a "no treatment" control.

Objective Performance Criterion (OPC)

A numerical target value derived from historical data from clinical studies and/or registries and may be used by FDA for the comparison of safety or effectiveness endpoints.

Observational Study

Study that draws inferences about the possible effect of an intervention on subjects, but the investigator has not assigned subjects into intervention groups.

Open-Label Study

A clinical study in which the participant, health care professional, and others know which intervention or diagnostic test under study is being given (see also Non-Blinded Study).

Paired Design

The application of two or more interventions or diagnostic tests at the same point in time to the same subjects or subject specimens. This design may be not appropriate if the interventions or test interfere with each other.

Parallel Group Design

An (unpaired) design in which each study subject or subject specimen is assigned only one of several interventions or diagnostic tests being studied.

Performance Goal

A numerical value that is considered sufficient by FDA for use as a comparison of the pivotal study results with a safety endpoint, an effectiveness endpoint, or, in a diagnostic clinical performance study, a diagnostic performance measure.

Pilot Study

See Feasibility Study.

Pivotal Stage

Clinical development stage for medical devices during which the evidence is gathered to support the evaluation of the safety and effectiveness of the medical device. The stage consists of one or more pivotal studies.

Pivotal Study

A definitive study during which evidence is gathered to support the safety and effectiveness evaluation of the medical device for its intended use.

Placebo (Sham)

A device that is thought to be ineffective. In clinical studies, experimental interventions are often compared with placebos to assess the intervention's effectiveness (see placebo control study).

Placebo Control Study

A comparative investigation in which the results of the use of a particular investigational device are compared with those from an ineffective device used under similar conditions.

Placebo Effect

A physical or psychological change, occurring after an ineffective device is used, that is not the result of any special property of the device. The change may be beneficial, reflecting the expectations of the participant and, often, the expectations of the person using the device.

Protocol (Study Protocol)

A study plan on which the clinical study is based. A protocol describes, for example, what types of people may participate in the study, the schedule of tests, procedures, medications, and dosages; and the length of the study.

Randomization

The process of assigning participants to groups such that each participant has a known, and usually an equal, chance of being assigned to a given group.

Randomized study

A study in which participants are randomly (i.e., by chance) assigned to one of two or more interventions (or diagnostic tests) of a clinical study.

Selection Bias

Bias due to systematic differences between those selected for the study population and those for the intended use population. For example, in a clinical outcome, parallel group study, study results can be subject to selection bias when the investigational and control groups are chosen so that they differ from each other in ways that affect the study outcome.

Specimen

The discrete portion of a body fluid or tissue taken for examination, study, or analysis of one or more quantities or characteristics.

Specimen matrix

Medium or milieu in which the analyte of interest may be contained (e.g., cerebrospinal fluid, serum, blood, other tissue, or viral transport media). The discrete portion of a body fluid or tissue taken for examination, study, or analysis for one or more quantities or characteristics.

Stratification

The division of a population into mutually exclusive and exhaustive sub-populations (called strata), which are thought to be more homogeneous, with respect to the characteristics investigated, than the total population.

Stratified (Subgroup) Design

Design in which the target population is divided into subject subsets (or strata) and subjects are selected separately from each subset (or stratum).

Study Endpoint

A primary or secondary outcome used to judge the effectiveness of an investigation.

Superiority study

Study designed to demonstrate that the safety or effectiveness of the investigational device is superior to that of the comparator.

Target Condition

The condition for which the device is to be used. In the context of diagnostic devices, a past, present, or future state of health, disease, disease stage, or any other identifiable condition within a subject; or a health condition that should prompt clinical action such as the initiation, modification or termination of treatment.

Test Device

See Investigational Device.

Therapeutic Device

Devices intended for use in the treatment of a specific condition or disease.