Patient-Focused Drug Development: Collecting Comprehensive and Representative Input

Guidance for Industry, Food and Drug Administration Staff, and Other Stakeholders

DRAFT GUIDANCE

This guidance document is being distributed for comment purposes only.

Comments and suggestions regarding this draft document should be submitted within 90 days of publication in the *Federal Register* of the notice announcing the availability of the draft guidance. Submit electronic comments to https://www.regulations.gov. Submit written comments to the Dockets Management Staff (HFA-305), Food and Drug Administration, 5630 Fishers Lane, Rm. 1061, Rockville, MD 20852. All comments should be identified with the docket number listed in the notice of availability that publishes in the *Federal Register*.

For questions regarding this draft document, contact (CDER) Meghana Chalasani at 240-402-6525 or (CBER) Office of Communication, Outreach and Development at 800-835-4709 or 240-402-8010.

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

June 2018
Procedural

Patient-Focused Drug Development: Collecting Comprehensive and Representative Input

Guidance for Industry, Food and Drug Administration Staff, and Other Stakeholders

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

and/or

Office of Communication, Outreach and Development
Center for Biologics Evaluation and Research
Food and Drug Administration
10903 New Hampshire Ave., Bldg. 71, Room 3128
Silver Spring, MD 20993-0002
Phone: 800-835-4709 or 240-402-8010
Email: ocod@fda.hhs.gov
http://www.fda.gov/BiologicsBloodVaccines/GuidanceComplianceRegulatoryInformation/Guidances/default.htm

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

June 2018
Procedural

TABLE OF CONTENTS

TABLE OF FIGURES

TABLE OF TABLES

Patient-Focused Drug Development: Collective Comprehensive and Representative Input
Guidance for Industry, Food and Drug Administration Staff, and Other Stakeholders[1]

This draft guidance, when finalized, will represent the current thinking of the Food and Drug Administration (FDA or Agency) on this topic. It does not establish any rights for any person and is not binding on FDA or the public. You can use an alternative approach if it satisfies the requirements of the applicable statutes and regulations. To discuss an alternative approach, contact the FDA staff responsible for this guidance as listed on the title page.

I. INTRODUCTION

A. Overview of the Series of FDA Guidance for Enhancing the Incorporation of the Patient's Voice in Drug[2] Development and Regulatory Decision Making

This guidance (Guidance 1) is the first of a series of four methodological ***patient-focused drug development*** (PFDD) guidance documents[3] that FDA is developing to address, in a stepwise manner, how stakeholders (patients, researchers, medical product developers and others) can collect and submit ***patient experience data***[4] and other relevant information from patients and caregivers for medical product[5] development and regulatory decision making.

This series of guidance documents builds on learnings from the disease-specific PFDD meetings[6] that FDA conducted under the fifth authorization of the Prescription Drug User Fee Act (PDUFA V) as an enhancement of the Agency's implementation of a more structured approach to ***benefit-***

[1] This guidance has been prepared by the Office of New Drugs and the Office of Translational Sciences in the Center for Drug Evaluation and Research (CDER), in cooperation with the Center for Biologics Evaluation and Research (CBER), at the Food and Drug Administration.

[2] For the purposes of this guidance, all references to *drugs* include both human drugs and therapeutic biological products unless otherwise specified.

[3] The four guidance documents that will be developed correspond to commitments under section I.J.1 associated with PDUFA VI under Title I of the FDA Reauthorization Act of 2017. The projected timeframes for public workshops and guidance publication reflect FDA's published plan aligning the PDUFA VI commitments with some of the guidance requirements under section 3002 of the 21st Century Cures Act. https://www.fda.gov/downloads/forindustry/userfees/prescriptiondruguserfee/ucm563618.pdf

[4] 21st Century Cures Act: https://www.congress.gov/bill/114th-congress/house-bill/34

[5] A drug, biological product, or medical device.

[6] https://www.fda.gov/forindustry/userfees/prescriptiondruguserfee/ucm347317.htm

risk assessment.[7] The PFDD meetings conducted to date have given FDA a deeper appreciation for the expertise that patients and caregivers can bring to the process and the value of incorporating their voice. This series of guidance documents is intended to facilitate the advancement and use of systematic approaches to collect and use robust and meaningful patient and caregiver input that can better inform medical product development and regulatory decision making.

Focusing on practical approaches and methods, this series will inform stakeholders of FDA's current thinking about methods that could be used bridge from important early-stage efforts to gain patients' narrative perspectives on the clinical context (e.g., meetings with patients), to development and use of ***methodologically-sound*** data collection tools in clinical trials. These guidance documents will also address Agency expectations regarding what sort of analyses might be conducted as part of this work and what sort of documents might be produced, and when appropriate, submitted to FDA.

The topics and questions that each guidance document will address are described below.

Guidance 1: Whom do you get input from, and why? How do you collect the information?

Guidance 1 will discuss sampling methods that could be used when planning to collect patient input. It will also provide a general overview of the relationship between potential research question(s) and method(s) when deciding from whom to get input (including defining the target population and development of the sampling strategy).

Guidance 2: What do you ask, and why? How do you ask non-leading questions that are well-understood by a wide range of patients and others?

Guidance 2 will discuss methods for eliciting information from individuals identified in Guidance 1, gathering information about what aspects of symptoms, impacts of their disease, and other issues are important to patients. It will discuss best practices in how to do qualitative research including conducting interviews, development of interview guides, selection of types of survey questions, and considerations for collecting demographics and survey information. It will also discuss survey methods and qualitative research topics to help avoid misleading results such as inadvertently priming patients in ways that can lead to results that poorly represent what is important to patients.

Guidance 3: How do you decide what to measure in a clinical trial and select or develop ***fit-for-purpose clinical outcome assessments (COAs)***?[8]

Guidance 3 will address refining the list of important impacts and concepts from patients to develop potential study instruments. Given that not everything identified as important by patients, caregivers, and clinicians can demonstrate change in a specific treatment trial or is

[7] https://www.fda.gov/forindustry/userfees/prescriptiondruguserfee/ucm326192.htm

[8] Words or phrases found in the Glossary appear in bold italics at first mention within the body of text in this document.

measurable, how will you select what to measure in a medical product development program to show clinical benefit? How will you identify or develop fit-for-purpose COAs to assess outcomes of importance to patients?

Guidance 4: Once you have a COA measurement tool and a way to collect data using it, what is an appropriate clinical trial endpoint?

Guidance 4 will address topics related to COA-related endpoint development and interpretation, including topics related to instrument administration and meaningful within-patient score changes.

This series will discuss methods and approaches for collecting information that can be applied for different types of patient input. For example, in addition to work related to planning for use of fit-for-purpose COAs, other research questions may include: What aspects of clinical trial conduct (e.g., informed consent, oversight by an institutional review board (IRB), enrollment, frequency of assessments, assessment burden, patient follow-up) can be better tailored to address the needs and concerns of the patients? What steps can be taken to minimize patient burden due to research participation? In all cases, the level of rigor of the methods and approaches applied should be appropriate for the questions the study wants to address and the potential impact of incomplete or misleading results.

The science of patient input is constantly evolving, and gathering robust and meaningful patient experience data to inform medical product development is a collaborative process. This document is intended to serve as a basis for dialogue. Stakeholders around the world have developed and are developing templates, checklists, and guidelines for different aspects of gathering and interpreting patient experience data. As these projects and documents mature, we will be updating our approaches. If you are considering collecting patient experience data, FDA encourages you to have early interactions with FDA and obtain feedback from the relevant FDA review division on appropriate research design.

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

B. Purpose and Scope of Guidance 1

The purpose of Guidance 1 is to present sampling methods for collecting information on the patient experience that is representative of the intended population to inform the development and evaluation of medical products throughout the medical product lifecycle. In addition, this document discusses methods on how to operationalize and standardize the collection, analysis, and dissemination of patient experience data.

Hypothetical case examples, which provide practical supplemental information that illustrate important concepts presented in this guidance, will be posted on the CDER PFDD webpage.[9]

Guidance 1 also includes a glossary of terms that will be used in one or more of the four guidance documents[10].

In addition to standardizing terminology for an identified disease area, the information in Guidance 1 should help the user develop a plan that will:

- Identify approaches and methods to collect information from patients and caregivers
- Identify approaches to sampling will ensure that the input to be collected is sufficiently representative of the range of clinically relevant diversity in the patient population
- Identify methods and necessary steps to develop a plan for analysis and reporting of the information that will be collected

The level of rigor needed for generating patient experience data can vary across studies and will depend on the intended use. However, there are certain common elements to all studies such as a protocol, structured data collection, and analysis.

This document is intended to serve as a focus for continued discussion among FDA, patient partners, medical product developers, researchers, and others.[11] It is anticipated that this document will provide a foundation for FDA and external stakeholders in the development of subsequent relevant guidance(s) on patient-focused medical product development. Although this document presents methods and approaches for collecting patient experience data, it does not fully address methods for collecting and analyzing COAs or ***patient preference information***. Some of those issues are addressed in the following guidance for industry[12]:

- *Patient-Reported Outcome Measures: Use in Medical Product Development to Support Labeling Claims*
- *Patient Preference Information—Voluntary Submission, Review in Premarket Approval Applications, Humanitarian Device Exemption Applications, and De Novo Requests, and Inclusion in Decision Summaries and Device Labeling.*

If you are considering collecting patient experience data, FDA encourages you to have early interactions with FDA and obtain feedback from the relevant FDA review division.

---

[9] https://www.fda.gov/Drugs/DevelopmentApprovalProcess/ucm579400.htm.

[10] The draft glossary of terms has been shared as an attachment to this guidance.

[11] In addition to consulting guidances, stakeholders are encouraged to contact the appropriate FDA office to discuss specific issues that arise during drug development.

[12] Guidances are updated periodically. For the most recent version of a guidance, check the FDA Drugs guidance web page: http://www.fda.gov/Drugs/GuidanceComplianceRegulatoryInformation/Guidances/default.htm.

C. Patient Experience Data

What is patient experience data? Patient experience data is defined in Title III, section 3001 of the 21st Century Cures Act (Pub. L. 114-255), as amended by section 605 of the FDA Reauthorization Act of 2017 (Pub. L. 115-52) (FDARA), to include data intended to provide information about patients' experiences with a disease or condition, including the impact (including physical and psychosocial impacts) of a disease or condition, or a related therapy or clinical investigation on patients' lives; and patient preferences with respect to treatment of their disease or condition.[13] Patient experience data can be understood as including (but is not limited to) the experiences, perspectives, needs and priorities of patients related to:

- the symptoms of their condition and its natural history;
- the impact of the condition on their functioning and quality of life;
- their experience with treatments;
- input on which outcomes are important to them;
- patient preferences for outcomes and treatments; and
- the relative importance of any of these issues as defined by patients.

Others have defined patient experience in similar ways. The patient experience in a medical product development context incorporates the patient's journey throughout the course of their disease or condition including patient views, feelings, needs, actions, preferences, interactions (e.g., clinical trials, home life, social life, etc.) with respect to their disease and its treatment (McCarthy *et al.*, 2016; Wolf, Niederhauser, Marshburn, & LaVela, 2014).

The patient's journey should be defined from the patient perspective. An understanding of that perspective may be enriched or informed by input from patient partners and clinicians. A ***patient partner*** may be an individual patient, caregiver or patient advocacy group that engages other stakeholders to ensure the patients' wants, needs and preferences are represented in activities related to medical product development and evaluation (Wilson *et al.*, 2018). **Table 1** describes types of patient partners.

[13] The definition is codified at section 569C(c)(4) of the Federal Food, Drug, and Cosmetic Act (FD&C Act), and applies to section 3002 of the 21st Century Cures Act, which directed FDA to issue certain guidance documents regarding the collection of patient experience data, see section 3002(b).

Table 1. Types of Patient Partners

- A **patient** is any individual with or at risk of a specific health condition, whether or not they currently receive any therapy to prevent or treat that condition. Patients are the individuals who directly experience the benefits and harms associated with medical products.
- A **caregiver** is a person who helps a patient with daily activities, health care, or any other activities that the patient is unable to perform himself/herself due to illness or disability. This person may or may not have decision-making authority for the patient and is not the patient's healthcare provider.
- A **patient advocate** is an individual or group that advocates for patients' health or healthcare. The advocate may or may not be part of the target population, and may work to influence healthcare policies or practices.

There are different parts of the patient experience to collect and/or measure in medical product development, which may include but are not limited to:

- Impact of the disease and its treatment on the patient
    - signs/symptoms of disease or condition
    - chief complaints (most bothersome signs/symptoms)
    - burden of living with a disease or condition
    - burden of managing a disease or condition
    - burden of participating in clinical studies
    - impacts from disease or condition on activities of daily living and functioning
    - impacts from treatment on activities of daily living and functioning
- Patients' perspectives about potential and current treatments
    - minimum expectations of benefits
    - tolerance for harms or risks
    - acceptable tradeoffs of benefits and risks (i.e., patient preference)
    - attitudes towards uncertainty
- Views on unmet medical needs and currently available treatment options
- Enhanced understanding of the natural history of the disease or condition, including progression, severity, chronicity

Information collected on patient experience will be referred herein as patient experience data.

Can data be collected from other experts as well? Where appropriate to supplement patient experience data, FDA recommends also gathering input from clinicians and other experts in the given disease area to ensure important clinical outcomes are studied.

Who can collect and submit patient experience data? Patient experience data can be collected by any persons including (but not limited to): patients, family members and caregivers of patients, patient advocacy organizations, disease research foundations, researchers, and drug manufacturers. It should be clear in any submission to FDA which person or group has collected the data.

Why is it important to collect patient experience data? Patients are experts in their own experience of their disease or condition and the ultimate consumers of medical products. The collection of patient experience data is important because it provides an opportunity to inform medical product development and enhance regulatory decision making to better address patients' needs.

When do you collect patient experience data? Patient experience data may be collected throughout medical product development, beginning early in development (e.g., discovery) or independent of any specific medical product development program (precompetitive setting). Patient experience data can be used to help identify unmet medical needs and important clinical outcomes to be studied, as well as inform the design of future clinical trials. Further, patient experience data can help inform COA development and selection, as well as analyses and communication of benefit-risk.

When should patient stakeholders be involved in product development? Patients (including patients serving as advisors) should be meaningfully involved throughout the medical product development process—not only as study subjects but as partners. Engaging patients actively in the development process can potentially improve rates of trial enrollment and retention and increase applicability to patients (Bower *et al.*, 2014).

How do you collect patient experience data? Qualitative, quantitative, or mixed methods may be appropriate to collect robust and meaningful patient experience data depending on study goals and the research questions. These methodological approaches are discussed in **Section III** of this document and **Appendix 1**. Some general distinctions between each method are shown in **Table 2**. Factors to consider when selecting an appropriate methodological approach are discussed in **Section II**.

Patient experience data can be collected in a variety of research contexts, including (but not limited to): clinical trials, observational studies, advisory boards, public meetings, and other novel settings (e.g., online patient communities). The level of rigor needed for patient experience data generation can vary across studies and will depend on the intended use. As such, it is important to begin discussions with FDA early to determine which approach should be used. Methods for generating patient experience data will be discussed in more depth in Guidance 2.

Table 2. Methodological Distinctions for Collecting Patient Experience Data

	Research Approaches		
	Qualitative Research	**Quantitative Research**	**Mixed Methods Research**
Common Research Objectives	• Description, understanding, and exploration/confirmation	• Numerical description, causal explanation, and prediction	• Multiple objectives; provide complex and fuller explanation and understanding; understand multiple perspectives
Common Study Characteristics	• Attempt to understand participant views, perspectives and meanings of concepts; study groups and individuals in natural or controlled settings	• Study behavior under controlled conditions; isolate the causal effect of single variables	• Study multiple contexts, perspectives, or conditions; study multiple factors as they operate together
Data Collection	• Qualitative data such as in-depth interviews, participant observations, field notes, and open-ended questions	• Quantitative data generated using structured data-collection instruments	• Both qualitative and quantitative data
Data Characteristics	• Words, images, categorizations	• Quantifiable variables	• Mixture of quantifiable variables, words, categorizations, and images
Data Analysis	• Use descriptive analysis to identify patterns, themes, and holistic features of qualitative data	• Identify statistical relationships among variables	• Quantitative and qualitative analysis used separately and in combination
Examples	• A group of patients are interviewed to describe their experience with the disease or condition	• A group of patients are surveyed about their experience with the disease or condition symptoms with a questionnaire that uses closed-ended questions with distinct response options to quantify information	• A group of patients are given a survey or questionnaire with both open-ended and closed-ended questions

How can external stakeholders submit patient experience data to FDA? It is important to remember that patient experience data informs development and evaluation of medical products throughout the medical product lifecycle. There are various pathways to (a) submit patient experience data to FDA and (b) engage with FDA for discussion. Additional FDA guidance on how to submit patient experience data is under development. Depending on the type of patient experience data and the intended purpose of the data with respect to medical product development, different content and formats may be appropriate for submission. At the minimum, when patient experience data are submitted to the agency, a study report and protocol from the research study should be submitted to FDA, as well as additional information including the primary data capture (see **Section IV** and **Appendix 2**).

Specific criteria defining what is most informative and useful for FDA submission should be discussed early with the appropriate FDA review division(s), as the level and type of criteria might vary based on how the data will be used. However, in all cases the intended purpose of the patient experience data being submitted to the Agency (i.e., how the data are intended for use in supporting medical product development and regulatory decision making) should be made clear in the submission.

Many existing FDA regulations, guidances, and other standards and requirements pertaining to the capture/collection, transmission, processing, storage, archiving, retention, and submission of data from clinical studies conducted to support a regulatory medical product application (e.g., an investigational new drug application (IND), new drug application (NDA), or biologics license application (BLA) or medical product labeling language also apply to patient experience data generated in such studies. See **Appendix 2** for a partial list of such regulations, guidance(s), standards, and requirements.

How is patient experience data used for regulatory purposes? Patient experience data is used to help inform clinical trial design, trial endpoint selection, and regulatory reviews including benefit-risk assessments as well as potential labeling (or other communications). FDA encourages stakeholders considering collecting and submitting patient experience data to FDA to have early interactions with FDA during the design phase of such studies and obtain feedback from the relevant FDA review division.

FDA values the use of patient input to help foster the development and availability of safe and effective medical products. The collection of patient input helps FDA gain a better understanding of the patient experience and expected clinical benefit.

II. GENERAL CONSIDERATIONS FOR COLLECTING PATIENT EXPERIENCE DATA

A. Overview

The selection of people from whom to collect input depends upon the specific questions and issues to be addressed. Thus, the selection process starts by considering the research question: what are the specific objectives to be addressed by collecting patient input? Are the objectives

focused on understanding the most burdensome symptoms, the impact of current therapies, their tolerance for risk, or the course of the disease over time? Each of these may require different approaches to patient selection and input collection.

How do you select a research approach? The research approach should be determined during the study design phase, prior to study implementation, and should be comprised of the plans for your research as well as the steps to implement those plans. While selecting the appropriate study methods, you should consider the broad research assumptions underlying your study design as well as the detailed elements that should be incorporated into the methodology to meet those assumptions and achieve success (Johnson & Christensen, 2017; Teherani, Martimianakis, Stenfors-Hayes, Wadhwa, & Varpio, 2015).

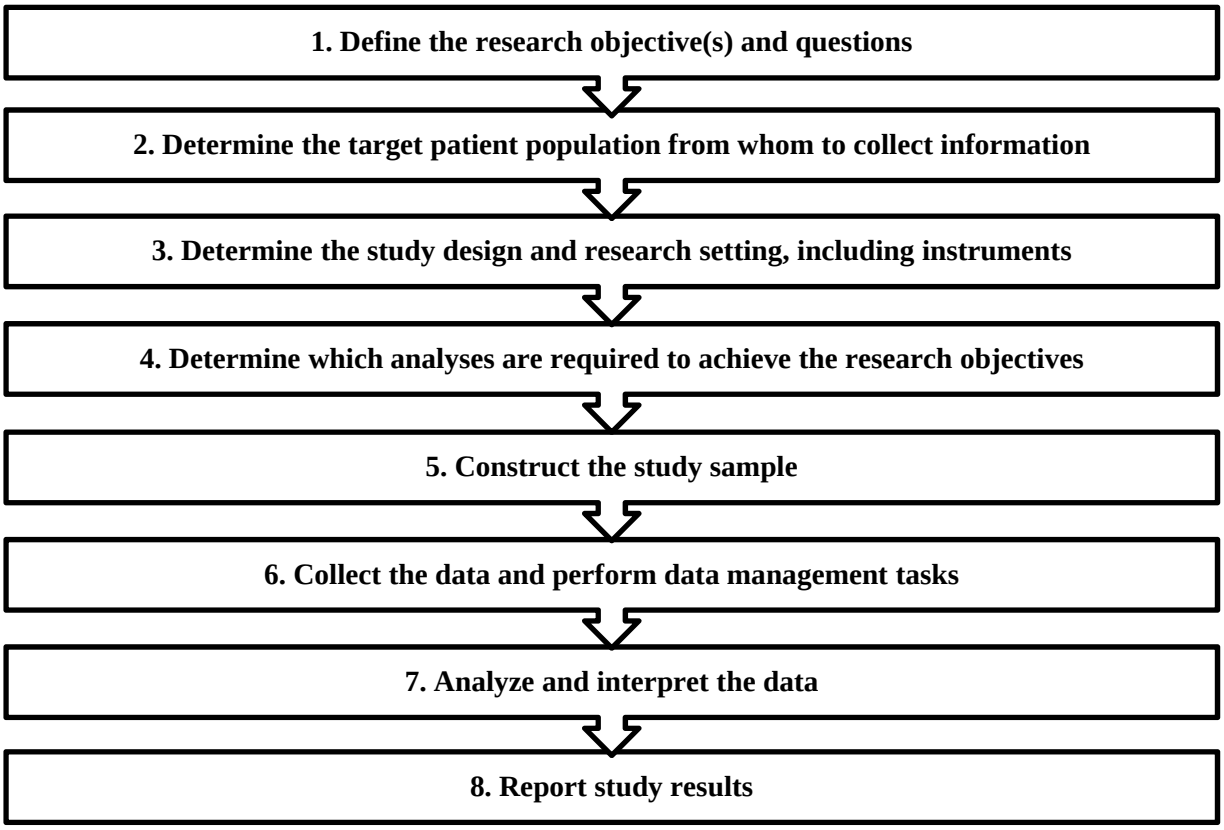
Some factors that are important to consider when selecting a research approach include:

- Research goals or questions to be addressed
- Target population and availability of people in that population
- Most valuable information that should be generated through the study to achieve your research goals or answer your questions
- Expected short-term and long-term impacts of the information you intend to gather through the study
- Amount of time to conduct your studies
- Study budget (including staffing, travel time, facilities costs, remuneration, data storage, management, and analysis)

If the research objective changes within the study, the research approach should be adjusted accordingly. You may leverage and build upon existing literature and data to fit the specific needs of the research question(s) (See **Section II.F.3**).

What steps should be used to collect patient experience data? FDA generally recommends that stakeholders follow the general steps listed in **Figure 1** for studying patient experience to the extent possible. The subsequent sections provide additional details.

Figure 1. General Steps for Conducting Studies about Patient Experience



B. Defining the Research Objectives and Questions

How do you define research objectives and questions? In general, your research objectives are defined by the research questions you are trying to answer. When formulating your research objectives, you should be specific. It may be useful to break down a broader research goal into specific research objectives and questions. Your research objectives and questions should inform which methodological approaches you use in your research.

When drafting your research questions, you should consult previously conducted studies and other relevant research literature along with subject matter experts (e.g., clinicians, social scientists). This will help to determine the most appropriate questions that will guide your study procedures. A carefully conducted review on your topic of interest coupled with expert consultation early in the study planning phase will help you clearly identify objectives and questions that will inform:

- which methods are better suited to meet your research goals and provide evidence to support your research questions; and

- the design of study materials (e.g., study protocol, interview guides, coding dictionary).

Example:
Research objective: To explore the attitudes toward treatment of U.S. teenage patients with human immunodeficiency virus (HIV)

Research questions:

1. How does HIV treatment impact patients' daily lives?
2. Why might HIV patients not use certain treatments?
3. What do patients look for in an ideal treatment for HIV?

Next steps: After defining your research objective and questions, you can start thinking about what research method to choose to meet your goal. If patients feel uncomfortable asking or answering questions or sharing concerns about living with HIV, it might be more suitable to engage them in one-on-one interviews over the telephone or other methods to provide them with a more comfortable interview setting rather than in group discussions or even administering a survey.

C. Who to Collect Information From

1. *Defining the Target Population*

How do you define the target population*?** The group of patients whose experience you wish to learn about is the ***target population. Characteristics of the target population should inform both the type of research methodology including the data collection mode that you choose for your study. It is important to tie the target population characteristics to the study sample and inclusion criteria. It also is useful to talk with FDA about how the target population could inform future medical product development and regulatory decision making.

Example: If you wish to understand the views and preferences of all individuals with Parkinson's disease (PD) in the world, then the target population could be defined as the set of all individuals who have been diagnosed with PD. If you are interested in a subset of PD patients, such as patients diagnosed within the last 5 years, then the target patient population could be restricted accordingly. The target population may also be restricted to a certain geographic area, such as PD patients in the US or the state of California. However, more restrictive patient inclusion criteria (e.g., limiting patients to specific geographic regions), the less likely it is the information is generalizable to a broader sample. PD patients in California may have different views and preferences than those in another country or even another part of the US.

2. *Determining Who Will Be Providing Patient Experience Data*

Who should provide the patient experience information? FDA generally recommends that the patient directly report their experience, unless the patient cannot reasonably be expected to reliably self-report (e.g., young children, individuals with cognitive problems, such as Alzheimer's disease). In such cases, a clinician or other trained health care professional and/or

primary caregivers, may report on patient experience if it is observable (e.g., signs of disease or condition, functioning) (FDA, 2015). Patient partners can also provide valuable information about the patient experience.

The ***reporter*** (the person who will be providing the patient experience information) may vary from patient to patient within the target population.

Factors to consider when deciding if self-report is feasible include but are not limited to:

- Age
- Level of cognitive development or function
- Communication (e.g. linguistic, numeracy) skills
- ***Health literacy*** (including basic ***literacy***)
- Insight
- Health state
- Co-morbidities

Prior to study initiation, it is important to set the criteria for determining the reporter. This may include, for example,

- What is the minimal age limit at which children can provide reliable information?

- What is the minimal cognitive function at which individuals can provide reliable responses?

- What are the scenarios under which multiple reporters may be required?

The reporter should be recorded for each individual in your study.

Example: If you are studying asthma in patients aged 4-17 years old, then the reporter might be (a) the patient's primary caregiver or parent for young children who cannot provide a reliable response and (b) the patient themselves (if determined they are of age to provide a reliable response).

FDA recommends stakeholders engage with subject matter experts (e.g., clinicians, social scientists) in the specific disease area of interest when determining the appropriateness of self-report in the target population.

3. *Subgroups*

All subgroups of interest should be pre-specified at the study design stage whenever possible. Care should be taken with the number of subgroups being proposed for analysis and inference. Subgroups of interest may be based on reporter type (e.g., patients versus primary caregivers) and/or socioeconomic, demographic, cultural, linguistic, clinical, or other factors pertinent to the disease/condition of interest.

D. Determining the Study Design and Research Setting

How do you determine the research study design and setting?

Some study features that are important to consider when determining your research study design and setting include:

- Study Type (e.g., a clinical trial/study, observational study, survey study
- Methodological Approach (e.g., qualitative, quantitative, or mixed methods)
- Sampling Method
- Including sample size
- Patient Selection
- Including methods for diagnosis determination (e.g. self-report, clinician report, other source)
- Subgroups
- Whether special considerations are needed for subgroups of interest

In general, these study features are determined by your research objectives and questions, and to some extent, they are determined by your resource constraints.

1. *Sampling Methods*

Prior to eliciting patient experience data, it is important to determine how individuals are to be selected to participate in your study. This is sometimes referred to in the statistical literature as the sampling scheme. Understanding how patients are sampled into your study determines whether your research objectives and questions can be answered by the patient experience data that you will collect.

There are many sampling approaches, each varying in complexity, the use of which depends on your research objectives and resource constraints. FDA recommends stakeholders engage with subject matter experts (e.g., statisticians, psychometricians) when determining the appropriateness of sampling methods to use.

Table 3 provides a listing of sampling approaches that are used to obtain patient experience data. They can be classified under two broad categories:

- probability and
- non-probability.

Example: If you intend to conduct a study that is exploratory or hypothesis generating, with a view towards gaining insight into patient experience (see Section III.A), then a non-probability sampling approach may suffice. See Table 3 for the different types of non-probability sampling approaches.

The necessary components for probability sampling include:

- Well-defined target population
- Listing of individuals within the target population
- Random device such as a random number generator

The listing of individuals is often referred to as the sampling frame. Ideally, the sampling frame should enumerate all individuals in the target population. A random number generator can be used to randomly sample individuals from the sampling frame which in principle produces a sample of patients whose experiences can be interpreted as being representative of the target population.

Example: Suppose the target population consists of 100,000 Parkinson's disease (PD) patients alive in the US and each individual is enumerated in a sampling frame with a label of 1 to 100,000. A sample of 2000 patients is randomly selected from among the 100,000 patients and their experiences are ascertained. Random sampling provides a mechanism for extending statements made about patient experience based on the individuals in the sample to the entire PD population. In practice, note that additional steps such as stratification may be needed to induce a sample having the desired characteristics.

Non-probability sampling, however, does not require a listing of the entire target population nor does it require a random device to sample individuals. Note also that in some cases, probability sampling can be accomplished without the availability of a formal sampling frame prior to study initiation as it may be constructed as part of the study.

It is beyond the scope of this document to discuss these sampling schemes in any detail. However, more in-depth discussions with respect to advantages and disadvantages can be found in the literature (For example, Fricker, 2008; Groves *et al.*, 2009; Heckathorn, 1997; Johnson & Christensen, 2014; Johnson, 2015; Korn & Graubard, 1999; Levy & Lemeshow, 2013; Rothenberg, 1995; Valliant, Dever, & Kreuter, 2013).

As noted earlier, the appropriate sampling scheme is that which

- enables you to answer your research objectives and questions, and
- can be implemented within the scope of your resource constraints.

543 **Table 3. Types of sampling, examples, and some potential limitations**
544

Types of Sampling	Selection Strategy	Examples	Potential Limitations
Probability Sampling			
Simple Random Sampling (SRS)	A sample drawn by a procedure in which every member of the population has an equal chance of being selected.	A simple random sample is taken from a population of patients admitted to a hospital in the first six months of 2015.	• Can be expensive or infeasible to conduct. • SRS samples can fail to reflect the heterogeneity in the target population.
Stratified Random Sampling	A sample drawn by dividing the population into mutually exclusive groups and then selecting a random sample from within each group.	Population of prisoners admitted to California prisons are stratified by race and gender and a SRS is taken for each race and gender combination.	• Requires the stratification factors to be known.
Multiplicity Sampling	A sample drawn by first taking a probability sample from the target population followed by drawing a sample from the set of individuals who belong to the network of those initially sampled	Current Population Survey Immigration-Emigration Supplement probability samples households each month. Includes question about immediate relatives who had previously lived in the US but are currently living abroad. Enables estimation of emigration rate. (Jensen, 2013)	• The initial probability sampling phase may not be feasible. • Relies on the initial respondents to identify members in their network.

Types of Sampling	Selection Strategy	Examples	Potential Limitations
Cluster Sampling	A sample drawn by which clusters (i.e., a collective type of unit that includes multiple elements, such as clinical sites in different geographic areas) are randomly selected and either complete- or sub-sampling of individuals within the selected clusters are taken.	A probability sample of hospitals in a state is taken, from which a probability sample of patients from each hospital is taken.	• Often requires information about cluster size as selection probabilities can depend on such information. • Units within cluster tend to be homogeneous.
Multistage Probability Sampling	Generalization of cluster sampling to include multiple levels/stages of cluster sampling.	CDC Medical Monitoring Project. • Stage 1, a probability sample of states. • Stage 2, a probability sample of facilities within each sampled state. • State 3, a probability sample of HIV patients from each sampled facility.	• Often requires information about cluster size as selection probabilities can depend on such information. • Units within cluster tend to be homogeneous.
Non-Probability Sampling			
Clinical Trials	A sample that consists of patients who volunteer to participate in a clinical trial.	Patients with iron deficiency anemia are recruited to participate in a clinical study that compares the efficacy of an experimental therapy against a standard of care.	• Trial results may not be generalizable to the population of all iron deficiency anemia patients for whom the therapy is indicated.
Convenience Sampling	A sample drawn by including people who are available, volunteer, or can be easily recruited in the sample.	Patients who can travel to attend Patient-Focused Drug Development (PFDD) meetings	• Study results may not be generalizable to the target population.

Types of Sampling	Selection Strategy	Examples	Potential Limitations
Purposive sampling	A sample drawn by which the researcher specifies the characteristics of the population of interest and locates individuals with those characteristics.	Researcher is interested in studying adult females with acne and selects their sample based on these characteristics (i.e., adult females with acne) and objective of study.	• Study results may not be generalizable to the target population.
Quota Sampling	A sample drawn by which the researcher determines the appropriate sample sizes or quotas for the groups identified as important.	Researcher chooses their sample to consist of 45% females and 55% males to maintain the correct proportions representative of the target population.	• Study results may not be generalizable to the target population.
Respondent-driven Sampling	Similar to snowball sampling. The chain of referrals is often longer than snowball sampling and under certain conditions, estimates can be generalizable to target population.	A convenience sample of individuals with substance use disorders (SUDs) is recruited. Each individual in this initial sample is provided a fixed number of coupons which he/she uses to recruit others in his/her network. The 2nd set of individuals recruited via coupons by the first set of individuals are also given a fixed number of coupons which they use to recruit individuals in their network. This is repeated for a fixed number of cycles after which recruitment terminates. The coupons serve as financial incentives for the recruited to	• Requires long recruitment chain and socially-networked population. • Study results may not be generalizable to the target population unless assumptions, which are not verifiable, are valid.

Types of Sampling	Selection Strategy	Examples	Potential Limitations
		recruit others in the network.	
Snowball Sampling (chain-referral)	A sample drawn by which each research participant is asked to identify other potential research participants. The initial sample of individuals is often obtained via non-probability sampling; subsequent samples are obtained by chained referrals from the previous sample.	Patients with sickle cell disease participate in focus groups to discuss symptoms of the disease and impacts of the medications taken. Focus group participants are asked to identify other people they know with sickle cell disease who may be potential research participants so study staff can invite them to join the research study.	• Study results may not be generalizable to the target population.
Web-based Sampling	A sample drawn by the contact mode (i.e., how the respondents are contacted, such as the web) which can involve multiple sampling strategies (e.g., systematic sampling, multiplicity sampling, list-based, entertainment polls, un-restricted self-selected surveys, volunteer (opt-in) panel).	Researcher selects patients from a web-panelist (e.g., online polling panel) to include in study	• Limited by pre-registered panelists • Study results may not be generalizable to the target population. • Potential response bias (e.g., measurement error, misclassification)

2. *Representativeness*

What is representativeness? When studying patient experience, it is important to obtain patient experience data that are not only relevant, objective, and accurate, but also representative of the target population. This is because it is usually impractical or impossible to select or study all patients in your target population. In this document, the term *representative* or *representativeness* can be interpreted in the following ways.

(1) **A sample is representative of the target population if statements made about patient experience based on data from the sample of patients are generalizable to the target population.** In principle, probability sampling schemes enable you to obtain such representative samples and often arise in the context of quantitative studies. However, if there are subgroups of patients from the target population that are not adequately represented in your study sample, your ability to generalize your research findings to the target population may be limited, even if you use a probability sampling scheme.

(2) **A sample is representative of the target population to the extent that patients in the study sample consists of individuals of various characteristics that to some degree approximate the heterogeneity of characteristics in the target population.** For example, your sample might consist of individuals from all levels of disease severity but the severity distribution in your sample does not necessarily resemble the severity distribution in your target population. This implies that statements made about patient experience based on data from the study are not necessarily generalizable to the target population. Whether this is acceptable depends on the research objectives. If your research objective is concept elicitation or hypothesis generation or instrument development, then this interpretation of representativeness is sufficient.

3. *Sample Size*

Sample size estimates are driven by:

- research objectives
- type of outcomes/endpoints under consideration
- study design
- planned methods of analysis
- whether the study is quantitative or qualitative in nature.

Having an insufficient sample size may produce unreliable and/or imprecise results. FDA recommends that if the sample size is limited due to practical considerations (e.g., rare diseases), the research objectives should be adjusted accordingly and noted as a limitation in the study report. Other practical considerations include:

- Small amounts of data

  The number of sampled individuals completing the study may be small. In this case, or when missing data/non-response are impactful, estimated sample size may need to be adjusted upward to maintain the desired level of statistical information.

- Subpopulations of interest

  There may be specific interest in one or more subpopulations. In this case, the sample size should be determined to ensure there is sufficient information to make statements about the subpopulation of the target population. If the goal of the study emphasizes both

the target population and a subpopulation within the target population, then the sample size should be determined to ensure that there is sufficient information to make statements about the subpopulation and the target population.

- Study design and analysis sample size needs

  The appropriate analysis methods for a study design may be unstable at certain sample sizes. Sample size calculations should take these features into consideration.

a. Studies Using Qualitative Methods

For qualitative studies, sample size determination is often less formal and based on the concept of saturation, which roughly means no new relevant or important information (e.g., new concepts of importance and relevance to subjects and research question) is gained by recruiting additional patients (Dworkin, 2012; Francis *et al.*, 2010) and the group of patients thus far recruited appears to be representative. As such, sample size formulae for such studies are often unavailable. Although sample size determination for qualitative studies is usually subjective, there is some guidance in the literature (Dworkin, 2012; Francis *et al.*, 2010; Sandelowski, 1995).

b. Studies Using Quantitative Methods

For quantitative studies, the criteria for sample size calculation are usually quantifiable.

Example: In efficacy superiority clinical trials comparing two or more arms, some of the common statistical specifications for determining sample size are:

- attaining a pre-specified power (e.g., sensitivity to detect a treatment effect of at least 80%, if the effect exists), and
- minimizing the chance of false positive results (e.g., type I error at most 5%).

For studies focusing on a single population, sample size calculation may be based on a precision criterion. For example, a study may require a sample size that is sufficiently large such that the estimated prevalence has a margin of error of at most 5% (roughly, precise to within 5%). Sample size calculations for different sampling types, study types, and data types can be found in the literature (Chow, Wang, & Shao, 2008; Levy & Lemeshow, 2013; Thompson, 1987). For complex designs where sample size formulae are intractable to obtain, simulation could be used.

E. Constructing a Sampling Frame

The existence of a sampling frame facilitates probability sampling. Without a sampling frame, it is potentially difficult or infeasible to randomly sample from the target population. To the extent that disease registries are inclusive and regularly-updated, they may provide a natural sampling frame. The scope of registries may vary, with some defined at the

- national level

- state level
- some local to an organization such as a hospital or a chain of hospitals owned by a particular organization or part of a network.

For many disease areas, however, registries may not exist or may not be inclusive or well-maintained. In such cases, resources may have to be devoted to constructing the sampling frame.

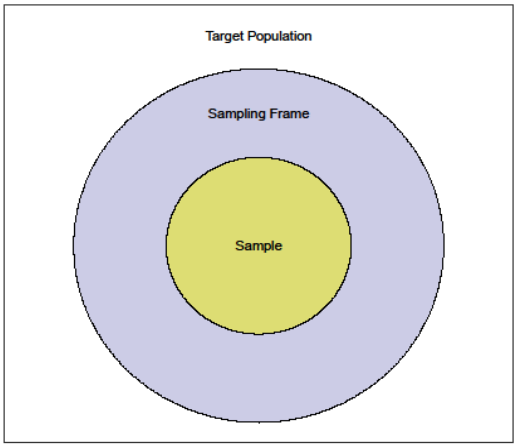
> **Example**: In the United States, physician listings such as the AMA Masterfile or state licensing board files have the potential to be used to create a sampling frame for the target population in the sense that a sample of physicians from these sources may be used to elicit members of the target population.
>
> It is important to note that unless all physicians treating patients are sampled, and all relevant patients under the care of each physician are identified, the resulting sampling frame may exhibit undercoverage in the sense that not every member of the target population is counted in the frame.

Figure 2 illustrates the concept of undercoverage. The target population of interest is depicted as the outer square. Undercoverage occurs because a proportion of members of the target population is not included in the sampling frame, the large circle. In general, undercoverage may not be problematic if:

- members excluded from the frame could be reasonably viewed as not being substantially different from those enumerated in the frame, and
- the primary goal of the study is to understand the distribution of the patient experience in the target population, rather than to estimate total number of people.

Regardless, attempts should be made to minimize undercoverage so that the patient population in the frame is not substantially different from the target patient population. In some cases, it may be possible to conduct a screening study to identify members of the target population and create a sampling frame.

Figure 2. Example Undercoverage Sampling Frame



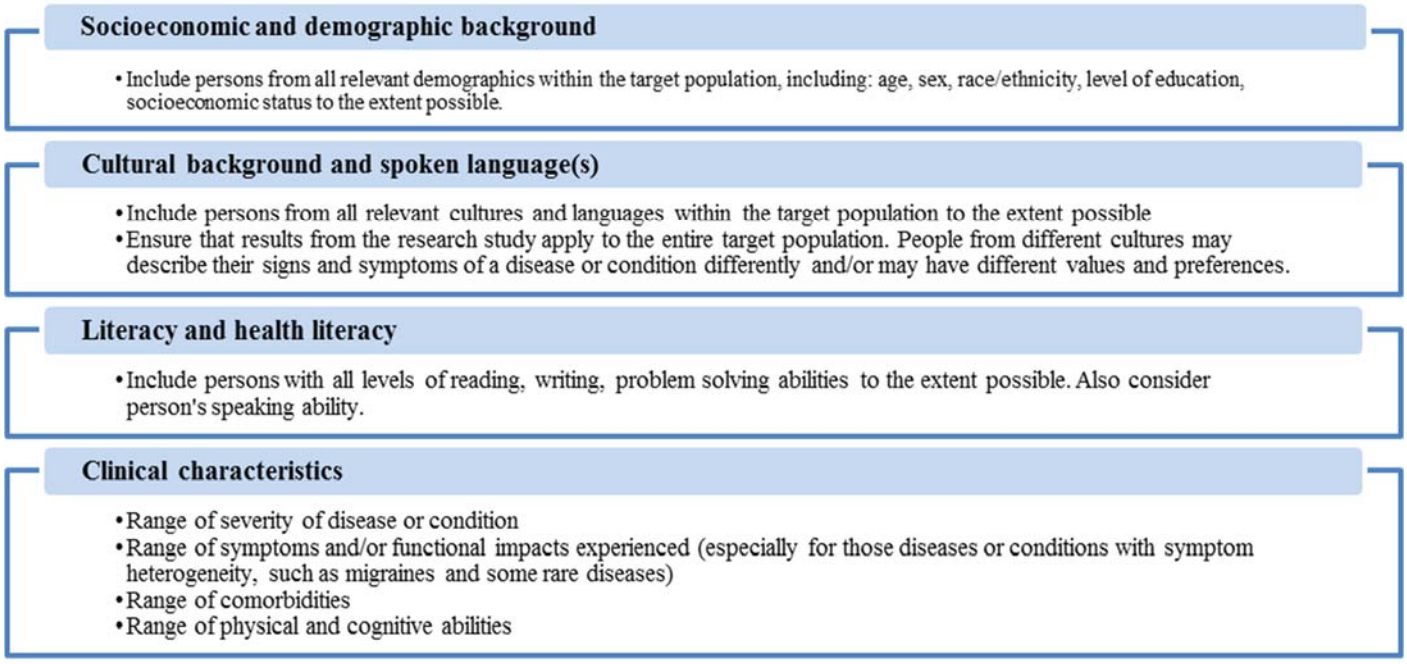
F. Additional Considerations

1. *Sufficient Representation*

For research questions in which probability sampling is appropriate, it might be possible to achieve sufficient representation through careful construction of a sampling frame, and selection and implementation of an appropriate probability sampling scheme.

However, there are scenarios in which probability sampling may not be feasible or required. Regardless of how individuals are selected into the study, it is important to ensure that patients in the study sample represent the target population, to the greatest extent possible, particularly with respect to the attributes that are associated with the endpoints of interest. For example, in a study of patients with a specified condition, to the extent that patients who have multiple comorbidity conditions may have patient experience distributions that are different than those with few or no comorbidity conditions, it is important that patients of varying levels of comorbidities are selected for your study. **Figure 3** provides some guidance regarding factors to consider to achieve sufficient representation.

Figure 3. Factors to Consider to Achieve Sufficient Representation



2. *Missing Data/Non-response*

Missing data is common in most types of studies. In randomized studies for example, some individuals may withdraw from the study prior to study completion. In observational epidemiological studies using electronic health records, some information that is important to the determination of a safety question was not collected by some health care providers or provided

by the individuals receiving care. In survey studies, individuals that were selected for the study refused to participate in the study (unit non-response) or refused to answer some questions (item non-response) after agreeing to participate in the study. In each of these cases, data that are determined to be useful for the assessment of the study questions are not available to the investigators. As missingness has the potential to introduce bias in unpredictable ways, particularly in the case where the reasons for missingness depend on the endpoint of interest, FDA recommends that investigators anticipate the occurrence of missing data and establish plans (in the study protocols) delineating strategies to minimize missing data, and where missingness cannot be avoided, to collect or determine the reasons for the missingness, where appropriate. Brick (2013), Calinescu, Schouten, and Bhulai (2012), Levy and Lemeshow (2013), and Schouten, Calinescu, and Luiten (2013) discuss design strategies for improving the response rate in the context of surveys. O'Neill and Temple (2012) and The National Research Council (2010) discuss design strategies for the prevention of missing data in the context of clinical trials.

3. *Leveraging Existing Data*

Sometimes engaging in primary data collection methods is not practical or feasible (e.g., collecting patient experience data from ultra-rare disease populations). Therefore, FDA encourages collaboration among multiple stakeholders and the use of methods to combine and reuse existing data (e.g., national registry data, archival databases) to fit the specific needs of the research question(s) and study goals. It is important to note that if you decide to explore the use of existing data, you should demonstrate the methodological rigor of the data collection method and data integrity as outlined in Section IV of this guidance.

III. METHODS FOR COLLECTING AND ANALYZING PATIENT EXPERIENCE DATA

This section provides an overview of potential methods for collecting patient experience data and considerations for analyzing patient experience data. However, FDA is open to the discussion of other methods.

Three main research approaches are commonly used to help guide the collection of patient experience data: qualitative research, quantitative research and mixed methods research (Johnson & Christensen, 2017). Additional discussion on these methods can be found in **Appendix 1**.

A. Qualitative Research Methods

What are qualitative research methods used for? Qualitative research is a method of inquiry used to gain insight into the patient experience and to better understand the meaning of research concepts (Johnson & Christensen, 2017; Neuman, 2014). Qualitative methods generally serve to generate in-depth information about the experiences, perspectives, and feelings of patients and other individuals (e.g., clinicians, caregivers), in their own words. Qualitative methods are used to elicit information related to research questions, whether it is to better understand burden of disease and/or treatment, or instrument design and feasibility.

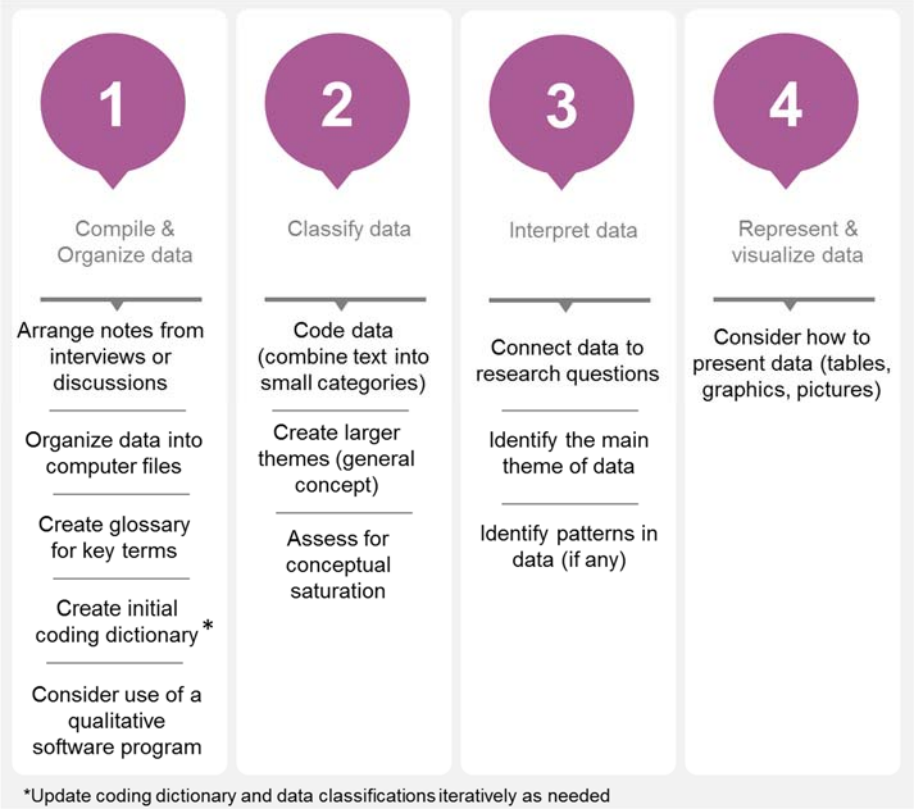
Qualitative research is a fluid, dynamic and evolving process. The key outcomes from this method include:

- Understanding patient experiences, perspectives, and feelings
- Discovering and confirming research concepts (e.g., parts of the patient experience that are important)
- Determining the meaning of and refining specific research concepts to measure in future clinical trials
- Evaluation of respondents' understanding of COA instruments

1. *Analyzing Qualitative Data*

How do you analyze data from studies using qualitative methods? FDA recommends stakeholders to consider the general steps outlined in **Figure 4** when analyzing qualitative data.

Figure 4. General Steps for Data Analysis in Qualitative Research



Source: Yin (2015), Qualitative Research From Start to Finish (2nd ed.), Guilford Press. Adapted with permission of Guilford Press.

Qualitative research can yield quantitative data at some level (e.g., proportion of patients who report a specific symptom). See **Section III.B** for additional details on different quantitative data types. See **Section IV.A.5** for additional details on organizing and recording information.

B. Quantitative Research Methods

What are quantitative research methods? Quantitative research methods are characterized by the collection of quantifiable data (e.g., numerical data) and the application of statistical methods to summarize the collected data. **Appendix 1** summarizes potential aims of quantitative research.

Example: A group of psoriasis patients are administered a survey to assess the symptom burden of psoriasis (e.g., patient's experience with psoriasis symptoms). The survey includes "closed-ended" questions with a fixed set of response options that generates a symptom score (e.g., numerical data).

1. *Analyzing Quantitative Data*

How do you analyze data from studies using quantitative methods? It is beyond the scope of this document to provide an exhaustive discussion of analytical approaches to analyze quantitative data. Generally, however, the analytic approach you take depends on the following:

- research objectives. This is partly related to the aims listed in **Appendix 1 (Table 13)**.
- study design. Potential designs include clinical trials, observational studies, surveys.
- types of data generated in your research study. Some examples include continuous, frequency, categorical, and longitudinal data.

Table 4 lists possible data types, descriptive approaches to summary statistics, distributional assumptions/methods for inference, and approaches to presentation of results.

Table 4. Possible approaches to quantitative data analyses and presentations

Data Types	Descriptive Statistics	Models/Methods	Data Presentation
Continuous (e.g., blood pressure level, pain score)	• Mean/median/mode • Standard deviation • Standard error • Confidence intervals • Range	• Normal distribution • Linear/non-linear regression • Analysis of variance • Analysis of covariance	• Tables • Graphs (e.g., scatter/density plots) • Stratification by age groups, gender, race/ethnicity, and other subgroups of interest

Data Types	Descriptive Statistics	Models/Methods	Data Presentation
Categorical, Count (e.g., number of hospital visits or adverse events per month; types or categories of adverse events)	• Frequencies • Proportions • Standard error • Confidence intervals	• Binomial/Multinomial/Poisson distributions • Generalized Linear Models (Agresti, 2002; Fleiss, Levin, & Paik, 2003; McCullagh & Nelder, 1989)	• Tables • Graphs • Stratification by age groups, gender, race/ethnicity, and other subgroups of interest
Longitudinal	Means, frequencies, or proportions at specific time points	See Diggle, Heagerty, Liang, and Zeger (2002) and Fitzmaurice, Laird, and Ware (2012).	• Tables • Graphs summarizing trend over time • Stratification by age groups, gender, race/ethnicity, and other subgroups of interest

2. *Additional Analytical Considerations for Data Obtained Under Probability Sampling*

For data obtained under probability sampling, it is important to incorporate the design feature into the analysis by weighting each sample unit by the reciprocal of the probability of selection, as this provides a mechanism for generalizing to the target population. This weight quantity is sometimes referred to as base weights and can be interpreted as the number of individuals that each patient in the sample represents in the population.

Example: Suppose there were 100,000 individuals in the sampling frame and 2,000 were sampled, then for simple random sampling with replacement, each individual has a probability of 0.02 of being selected. In the analysis, individuals are assigned a weight of 50 as determined by the reciprocal of 0.02. That is, each patient in the sample represents 50 individuals in the target population. For multistage designs, the sampling probabilities are obtained as the product of the sampling probabilities from each stage.

In cases where the design does not make use of stratification, gains in precision may be obtained by performing a post-stratification analysis via weighting class adjustment or raking. These algorithms are discussed by Korn and Graubard (1999) and Copeland and Ganesh (2015).

3. *Additional Analytical Considerations for Missing Data and Non-response*

To the extent that strategies designed to prevent or reduce missing data (see **Section II.F.2**) are unsuccessful, analyses using the observed data may or may not be valid, depending on the extent of missingness. FDA recommends the following:

- provide a table summary of missing data; useful information includes frequencies, percentages, stratification by important subgroups, reasons for missingness.

- for longitudinal data, summarize missingness stratified by assessment visits or time points.

- In addition, methods for handling missing data in analysis should be addressed in the protocol. See The National Research Council (2010) and ICH E9(R1) (International Council for Harmonization, 2017) for discussion of methods. Non-response missingness in surveys under probability sampling could be similarly addressed by missing data methods. Some of these approaches include weighting class-adjustment (Copeland & Ganesh, 2015; Korn & Graubard, 1999; Valliant *et al.*, 2013), calibration adjustment (Sarndal & Lundstrom, 2005), and propensity score modeling (Valliant et al., 2013).

C. Mixed Methods

What is mixed methods research? Qualitative data and quantitative data can complement each other. Mixed methods research is where both qualitative and quantitative methods are used. A mixed methods study addresses a set of research questions that require both qualitative and quantitative evidence and methods. Both the quantitative and qualitative data should be analyzed and interpreted together before reaching a conclusion.

Mixed methods studies can occur in different ways: mixing of data, of designs, and of analyses. These types of studies can be conducted within a single study or within a coordinated series of studies for integrated analyses. The simplest approach to a mixed method study involves the use of both qualitative and quantitative data.

A more complex approach to a mixed method study is mixing of designs. There are different types of mixed designs, which may include but not limited to the following:

Parallel (qualitative and quantitative in parallel)

- Using and analyzing open-ended (qualitative) and closed-ended (quantitative) items/questions as part of the same survey/questionnaire (Yin, 2016)
- Converting qualitative data into quantitative data through content analysis (Yin, 2016)

Sequential (qualitative first, then quantitative)

- Using qualitative data to define patient subgroups, based on observations of their experience with the disease/condition or treatment (qualitative) to identify variables or develop an instrument (Creswell & Clark, 2007; Yin, 2016)
- Using qualitative data from patients to characterize their disease experience (qualitative), and then comparing patients' responses to a survey/questionnaire (quantitative)

Sequential (quantitative first, then qualitative)

- Collecting quantitative data first and then using qualitative data to further understand the quantitative data. (Creswell & Clark, 2007; Yin, 2016).

Examples:

Scenario 1: A group of patients is administered a survey to assess the burden of Type 2 diabetes (e.g., patient's experience with Type 2 diabetes, including symptoms, doctor visits, medication use). The survey includes open-ended and closed-ended questions. With the use of these types of questions, the survey can produce both qualitative (textual) and quantitative (numeric or categorical) data.

Scenario 2: A ***patient-reported outcome*** (PRO) instrument is administered to patients in a clinical trial. Exit interviews where patients are asked whether they experienced a meaningful response are conducted shortly after the end of the clinical trial. Qualitative data from the interviews (patients' quotes) are used in comparing the PRO results from patients who reported a clinical benefit versus those who did not in order to interpret the PRO scores.

1. *Analyzing data from mixed methods*

How do you analyze data from mixed methods? Different types of analyses can be used to analyze data from a mixed method study, including combining the use of analyses described for qualitative (**Section III.A.1**) and quantitative (**Section III.B.1**) methods. FDA recommends that stakeholders choose the best analysis approach for their research objective.

IV. OPERATIONALIZING AND STANDARDIZING DATA COLLECTION AND DATA MANAGEMENT

A. Standard Approaches to Consider for Collecting and Managing Data

What activities occur during data collection? There are a series of inter-related activities in the process of collecting data, which include the following (Creswell, 2013):

- Locating patients/sites
- Gaining access
- Sampling strategy
- Collecting data
- Recording information
- Resolving site/field issues
- Managing and storing data

FDA encourages stakeholders to carefully plan these activities. Further, FDA recommends that stakeholders standardize data collection activities and methods to manage data quality to the extent possible.

1. *Locating Patients/Sites*

A critical step in the process of data collection is to identify the appropriate sample and/or sites to study. In order to have adequate generalization for multicenter clinical trials, patients should generally not be located from a single site. FDA generally recommends including patients from diverse sites to provide a complete picture of the topic of interest (see **Section II.D.2** on representativeness).

2. *Human Subjects Protection*

Research involving access to patient information or involves directly engaging with patients requires careful consideration of Federal, State, and local laws and institutional polices for the protection of human subjects. Because this guidance focuses on sampling methods for collecting patient experience data through a variety of research contexts (including, but limited to, clinical trials, observational studies, advisory boards, public meetings, etc.), a full discussion of which laws may apply to these collection methods is beyond the scope of this guidance. Research subject to FDA regulations must satisfy the requirements for informed consent at 21 CFR part 50 and the IRB requirements at 21 CFR part 56.[14,15] Research supported or conducted by the Department of Health and Human Services must satisfy the requirements at 45 CFR part 46.[16] FDA recommends that researchers work with their Institutional Review Boards (IRBs) and Health Insurance Portability and Accountability Act (HIPAA) Privacy Boards to determine what laws may apply.

FDA recommends research involving patient information be conducted in accordance with the principles of good clinical practice (GCP), including the International Conference on Harmonisation Guidelines (see **Appendix 2**).

[14] Details on 21 CFR part 50 can be found at https://www.ecfr.gov/cgi-bin/text-idx?SID=7d2cb8de0c8bebe70c93835ce013cdd3&mc=true&node=pt21.1.50&rgn=div5

[15] Details on 21 CFR part 56 can be found at https://www.ecfr.gov/cgi-bin/text-idx?SID=7d2cb8de0c8bebe70c93835ce013cdd3&mc=true&node=pt21.1.56&rgn=div5

[16] Details on 45 CFR part 46 can be found at https://www.ecfr.gov/cgi-bin/text-idx?SID=991d81fee482f9aa6ef549a0067a86e8&mc=true&node=pt45.1.46&rgn=div5

3. *Sampling Strategy*

Of similar importance within the data collection process is the determination of a strategy for the sampling of patients or sites. Refer to **Section II.D.1** on the different types of sampling.

4. *Collecting Data*

Stakeholders should consider which data collection approach is most appropriate for their research objective. Data collection methods can include but are not limited to the following:

- Observations
- Interviews
- Documents (including questionnaires)
- Audiovisual materials
- Digital Health Technology

Each of the four data collection methods generates different types of data (see **Table 5**), each of which has its own advantages and limitations. Some examples of data types are listed in **Table 5.** Additional detail, including potential advantages and limitations of each method, are discussed in subsections following the table.

Table 5. Data Collection Methods and Types of Data for Qualitative and Quantitative Research

Data Collection Method	Illustrative types of data	Specific examples of data
Interviews and Focus Groups	Language (verbal and body)	A person's description/explanation of some behavior or action; a memory; a belief or viewpoint (e.g., email, face-to-face, focus group, online focus group, telephone interviews; Delphi panel)
Observations	People's gestures; social interactions; actions; scenes and the physical environment	The communication/dynamics between two or more individuals; spatial arrangements of a person and a setting; efficiency of an intervention (e.g., time and motion studies)
Documents	Contents of: personal documents, other printed materials (e.g., literature), graphics, archival records, and physical artifacts	Information from public documents (e.g., official memos, minutes, records, archival material); medical records, chart audits; photo elicitation (participants take photographs or videotapes); scientific publications
Questionnaires	Set of questions (items) with a choice of answers (responses)	A person's response to a set of questions (e.g., surveys; clinical outcome assessments, such as patient-reported outcome instruments, observer-reported outcome instruments, clinician-reported outcome instruments)

Data Collection Method	Illustrative types of data	Specific examples of data
Audiovisual Materials	Sight and sound (recorded speech or actions)	Videotape or photographs of individuals or groups; sounds (laughter or other vocalized expressions); email or discussion board messages with audiovisual attachments; video chat/conferencing (e.g., Skype)
Social Media and Identifiable Patient Communities	Contents of: conversations or text (including elicited and non-elicited content) from online communities & social media interactions, online focus groups and interviews	Information on a disease or condition (e.g., symptoms and impacts) or individual patient experiences with a disease or condition as reported by patients in online patient groups (e.g., blogs, forums, message boards) and social media pages; phone (SMS/text messages), live video/chat conferencing or messaging
Digital Health Technology	Mobile health, health information technology, wearable devices, telehealth and telemedicine, and personalized medicine	Information from mobile health technology (e.g., accelerometers, heartrate trackers, etc.); information from certain mobile applications

Source: Adapted from Yin (2015), Qualitative Research From Start to Finish (2nd ed.), Guilford Press. Adapted with permission of Guilford Press.

a. Interviews and Focus Groups

Different interview types are used to collect patient experience data, including one-on-one interviews (semi-structured, structured or open-ended or group interviews (focus groups). The method of interviewing (e.g., in person, telephone or by video chat) may vary depending on the goals of the interview. For example, if visual cues are important for the context of the research objective, an appropriate data collection method may be face-to-face interviews either in-person or by video chat instead of telephone interviews. Further details and considerations regarding the different interview methods will be found in future guidances.

b. Observations

Observation can be a tool to collect patient experience data, and can include but is not limited to the observation of the interactions of a participant in particular setting, activity, or behavior (Creswell, 2013). Observations are helpful in situations for individuals who have barriers to communicating their thoughts orally or in writing. Additionally, observations of individuals or groups often can be done to supplement interviews (individual or group) by documenting cues from the environment and behaviors such as facial expressions, gestures, tone of voice, and other non-verbal indicators.

c. Documents

Various types of documents can be used to collect patient and/or caregiver input on burden of disease and treatment, some of which are listed in **Table 5**.

d. Questionnaires

Surveys or questionnaires can be used in observational studies to capture patient experience data. Questionnaires can also be used to collect patient-reported outcomes in clinical trials. This is a specific use to evaluate a patient's response to treatment and it necessitates certain measurement properties that will be described in detail in later guidances.

What are questionnaires? Questionnaires generally consist of a standard set of questions or items that are generally administered in the same order to each participant, but can be administered via computerized adaptive testing (Johnson & Christensen, 2017). Questionnaires can be administered in both observational studies and clinical trials. In these settings, data can be collected by questionnaires throughout the study or at the end of the study (e.g., exit surveys).

Exit surveys are a standardized method used to collect information about various experiences, including treatment satisfaction and study experience with minimal recall bias (Geldsetzer, Fink, Vaikath, & Barnighausen, 2018). Exit surveys are generally administered at the end of a participants' enrollment in a study. However, surveys also can be administered at multiple time points throughout the study (Hrisos *et al.*, 2009; Turner, Angeles, Tsui, Wilkinson, & Magnani, 2001).

Questionnaires can be administered in different modes:

- In-person paper administration: paper questionnaires filled out in person by the participant
- Interviewer administration: questionnaire administered by an interviewer following a structured protocol
- Telephone questionnaire administration: questions administered over the phone
- Electronic administration: participants can complete questions via email, web interface, or electronic device
- Interactive voice response systems: questions administered over an automated telephone system

What are some key considerations when using questionnaires to collect patient experience data? When using questionnaires to collect patient experience data, FDA generally recommends the following:

- Each participant in a sample is asked the same set of questions to the extent possible
- Design questions that are interpreted and understood well by participants
- Pre-test/pilot-test questions
- Avoid using incomplete questions (e.g., Age? Reason last saw doctor?)

- Avoid using questions that ask two or more concepts at once (e.g., How embarrassed or self-conscious have you been because of your condition?)
- Create distinct and non-overlapping response options for each question

If questionnaires are intended to be used in observational survey studies (paper or electronic-based), FDA encourages the following steps (Cooper *et al.*, 2006):

- Select pool of participants or panelists (e.g., health panels) to be observed. Obtain the required permissions needed to gain access to the participants and/or panelists.
- Create a system in which questions can be entered, as well as possible responses, into a database table.
- Generate tables to record the data entered through the questionnaire from the database table of questions and possible responses.
- Develop a simple, user-friendly paper-based or electronic-based questionnaire.
- Provide data validation during the entry process.
- Develop a coding manual that could be used as a reference document.
- For web-based surveys, generate descriptive statistics that could be observed through the web during the entry phase of the questionnaire.
- Develop program files that allow opportunity to do more advanced statistics once the questionnaire is completed.
- Maintain a database to access the questionnaire table and data entered into the questionnaire. This database should have built-in features or capacity to interface with software that has features such as forms, queries, and reports to further work with the data.

If questionnaires are intended to be a study endpoint in a clinical trial, FDA recommends that stakeholders adopt good measurement principles. Refer to the FDA PRO Guidance (FDA, 2009) on factors to consider when administering questionnaires in clinical trials. Refer to **Appendix 2** regarding standards and requirements pertaining data submission to FDA.

e. Audiovisual Materials

Audiovisual materials (e.g., audiotape, videotape, photographs, social media) also can be used to collect data in characterizing the patient experience (see **Table 5**).

Steps to consider when using audiovisual materials in the data collection process include:

- Obtain the required permissions needed to use materials, including informed consent.
- Obtain permission to extract information from web content, if necessary (e.g., request permission to join online forums and inquire whether there are restrictions on use of information for research purposes).

f. Social Media and Identifiable Patient Communities

FDA encourages external stakeholders to explore the use of ***social media*** tools (e.g., medical community blogs; crowdsourcing; and social media pages) to shed light on patients' perspectives regarding symptoms and impacts of a disease or condition. Targeted social media searches may be useful during the preliminary stages of a study to complement literature review findings, inform the development of research tools (e.g., qualitative study discussion guides) or as a supplement to traditional research approaches (e.g., literature, one-on-one interviews, focus groups or expert opinion).

Common methods for generating patient input through social media, along with some potential strengths and limitations of these methods, are detailed in **Table 6.**

Table 6. Common Methods for Gathering Patient Input Using Social Media

Social Media Qualitative Research Methodology	Strengths	Limitations
All social methodologies	• May allow access to hard to reach populations • Cost & time saving for researchers • Relatively easy to implement • Accurate & automatic capture of data • Participant convenience & comfort • Greater self-disclosure	• Self-selection bias (social media participants may include a narrow band of patients with regard to clinical or demographic characteristics)
Asynchronous online focus groups or interviews (occur at different places, different times) (Tates *et al*., 2009; Wilkerson, Iantaffi, Grey, Bockting, & Rosser, 2014)	• Can be conducted using email, discussion forums, and other forms of social media • Provide flexibility and convenience of logging in at own place and time • Lack of time pressure & greater reflection	• Lack of visual cues • Underlying selection process might be difficult, if not impossible, to quantify • Representativeness might be questionable without strong assumptions
Synchronous online focus groups or interviews (among younger participants; occur at different places, same time [for focus groups]) (Fox, Morris, & Rumsey, 2007; Wilkerson *et al.*, 2014)	• Data captured in real-time (synchronous) • Can be conducted using the phone (SMS/text messages), chat methods, video messaging • Interaction is often dynamic, immediate, conversational (similar to everyday interactions) • Assessment of visual cues (through video or emotions	• Scheduling can be difficult; must find a common meeting time (for focus groups) • Requires a fast internet connection, webcam/ audio/video capabilities which some participants may not have readily available • Technology rich interface can present more technical difficulties

Social Media Qualitative Research Methodology	Strengths	Limitations
	conveyed through emoticon use) • Less threatening methodology for younger participants	• Moderation can be difficult with too many participants; sometimes participants have trouble taking turns (for focus groups) • Faster typing speed gives participants an advantage and these participants can dominate the conversation (for focus groups) • Increased likelihood of passive participation (e.g., a participant logging on and observing but not participating) • Groups with more than 5 participants require 2 moderators (for focus groups) • Self-selection bias (social media participants may include a narrow band of patients with regard to clinical or demographic characteristics)
Designed online communities and social media data collection (Grajales *et al.*, 2014; Paulus & Lester, 2013; Varga & Paulus, 2014)	• Generated through platforms like online support groups and online educational groups • Groups include identifiable patients and identifiable reporters • Helpful for: - gathering information on health conditions (Prieto *et al*., 2014) - sharing treatments and experiences of care (McGregor *et al*., 2014) - recruiting research participants (O'Connor, Jackson, Goldsmith, & Skirton, 2014)	• Must have authorization to obtain identifiable information (e.g., Personal Health Information (PHI))
Spontaneous online communities and social	• Generated through easily accessible platforms	• Participants are unknown; Respondent identification not verifiable

Social Media Qualitative Research Methodology	Strengths	Limitations
media data collection (unelicited data) (Grajales, Sheps, Ho, Novak-Lauscher, & Eysenbach, 2014)	• Low burden for people providing data • Helpful for: - gathering information on health conditions (Prieto, Matos, Alvarez, Cacheda, & Oliveira, 2014)	• PHI not verifiable • Underlying selection process is difficult if not impossible to quantify • Representativeness is highly questionable without strong assumptions

While social media tools can provide useful data, limitations related to sampling need to be considered. With most social media sources, there is no mechanism for verifying patient identity, or clinical and demographic characteristics; you must rely on patient self-identification and diagnosis, which can be inaccurate. Additionally, different demographic groups tend to use different types of social media (e.g., Pinterest is often dominated by female users, Instagram is dominated by young adults). Based on this variability, you may need to use different social media tools to gather information from the demographic group(s) you are targeting. Likewise, when submitting information for regulatory review, you would have to demonstrate how the data collection methods used to generate data addresses these limitations and to ensure rigor in methodology and data integrity.

It may be possible to mitigate concerns around the lack of ability to confirm patient characteristics (e.g. diagnosis) in various ways. For example, to have an identifiable patient, there should be enough information to indicate the existence of a specific patient, including but not limited to age (or age category, e.g., adolescent, adult, elderly), gender, initials, date of birth, name, or patient identification number. Clinical information may also be available, by permission, through a central database (e.g., for patients who are members of patient advocacy group message boards, social networking groups, or medical community blogs). An identifiable reporter can be a family member, doctor, other health care practitioner, or other individual who has sufficient information to indicate that they are an identifiable person who has knowledge about the patient.

g. Digital Health Technology

Digital health technology can be one approach to mobile data collection and can include devices that allow participants to track some aspect of their health data. FDA recommends stakeholders who are collecting patient experience data with digital health technology to discuss the planned method early with FDA and obtain feedback from the relevant FDA review division.

5. *Recording Information*

FDA recommends that stakeholders develop written forms or protocols to collect patient experience data, such as a discussion guide or observational protocol. A discussion guide or observational protocol is a pre-designed form used to record information collected during an interview or observation (e.g., interviewer may take notes on the discussion guide or

observational protocol). Patient experience data can also be recorded through various forms, such as interview summaries and audio-and video-recordings.

6. *Resolving Site/Field Issues*

FDA recommends that standardized training be provided to the members of the research team to improve consistency of research. The roles and responsibilities of the team should be outlined in the research protocol. This will help to prevent many site issues. FDA encourages stakeholders to also have a troubleshooting guide. Researcher(s) should anticipate and address site/field issues that might arise during data collection. Some issues to consider are listed below (Creswell, 2013):

Access to patients/sites

- Patients' willingness to participate in research
- Patient responsiveness
- Appropriateness of a site
- Building of trust and credibility at the field site
- IRB unfamiliar with certain methodologies

Interviews

- Mechanics of conducting interviews (unexpected participant behaviors, sensitive issues, inexperienced researchers)

Paper Questionnaire Administration

- Quality control at the visit (e.g., administering correct version of the questionnaire, looking for non-response patterns, such as not completing a particular section)

Electronic Questionnaire Administration

- Consistency in data monitoring procedures and follow-up (e.g., monitoring for timely completion and attrition)

Observations

- Consistency in the role of observer
- Mechanics of observing (remembering to take site notes)
- Recording accurate quotes/notes
- Managing information sufficiently at site
- Funneling information from the observations appropriately

Documents, Audiovisual materials, Digital Health Technology

- Locating materials
- Obtaining permission from the participant to use materials (e.g., audio-/video-recorder)
- Minimal noise disturbance
- Best location for video recorder/camera

Ethical issues

- Informed consent, if required
- IRB oversight, if required
- Conflicts of interest
- Dishonest or hidden (secret) activities
- Confidentiality and privacy considerations
- Benefits of research to participants, and risks

7. *Data Management*

FDA recommends that data management be addressed in the early stages of a research study. Before initiating data collection, you should formulate a data management plan (DMP)—a written document that describes the data you expect to acquire or generate during your research study; how you intend to manage, describe, analyze, and store said data; and what mechanisms you will use at the end of your study to preserve and share your data (Stanford University Libraries, n.d.-a); creating a written DMP helps formalize the data management process, identify potential weaknesses in the DMP, and provides a record of what you intend(ed) to do. See **Appendix 3** for resources to consider when developing a data management plan, as well as components of a good data management plan.

8. *Data Standards*

External stakeholders should use appropriate data standards when collecting, managing, and reporting patient experience data. When planning a study (including the design of case report forms, data management systems, and statistical analysis plans), you should determine which FDA-supported standards to use or request a waiver of those requirements. There may be versions of a standard available that are not yet supported by FDA (e.g., specific SDTM or ADaM versions) or there may be FDA-supported standards that, currently, have only specific components developed (e.g., SEND study types).[17] See **Appendix 2** for some data standards resources.

While compliance with these standards may not be required for studies[18] other than those conducted to support a regulatory medical product application (e.g., an IND, NDA, or BLA) or medical product labeling language, we encourage researchers to, at a minimum, bear these standards in mind, because patient experience data that are ultimately intended for use in clinical studies *would* be subject to the applicable standards.

9. *Monitoring and Quality Assurance*

FDA expects that external stakeholders will be responsible for monitoring the study, ensuring data integrity, and performing the data analysis.

[17] Study Data Tabulation Model (SDTM), Analysis Data Model (ADaM), Standard for Exchange of Nonclinical Data (SEND), available at www.cdisc.org.

[18] Such as stand-alone psychometric validation studies submitted to the COA Drug Development Tool (DDT) Qualification Program

10. *Storing Data*

FDA recommends that external stakeholders plan how to store their data in advance of starting their study. Researchers should decide how data will be best stored so that it can be easily retrieved and protected from any type of damage or loss. The approach to data storage should reflect the type of data collected. Regarding the length of time to keep records of data, researchers should comply with their IRB and applicable regulations.

Principles to consider about data storage and handling data include the following (Creswell, 2013):

- Create back-up copies of computer files
- Use high-quality equipment for audio-recording information during interviews
- Protect the anonymity of participants by de-identification
- Create a data collection table or database to track and identify data
- Maintain a list of types of data collected

11. *Confidentiality*

All personal participant data collected and processed for research should be managed by the research team with adequate precautions to ensure confidentiality of the data in accordance with applicable national and/or local laws and regulations on personal data protection.

V. CONCLUSIONS

This document has provided an overview of methods to collect robust, meaningful, and sufficiently representative patient input to inform medical product development and regulatory decision making. The proposed methods presented serve only as a basis for dialogue in the evolving and growing area of the science of patient input. If you are considering collecting patient experience data, FDA encourages you to have early interactions with FDA and obtain feedback from the relevant FDA review division on appropriate research design and any applicable regulatory requirements.

VI. REFERENCES

Agresti, A. (2002). *Categorical Data Analysis* (2nd ed.). Hoboken, NJ: John Wiley & Sons.

Bower, P., Brueton, V., Gamble, C., Treweek, S., Smith, C. T., Young, B., & Williamson, P. (2014). Interventions to improve recruitment and retention in clinical trials: a survey and workshop to assess current practice and future priorities. *Trials*, *15*(1), 399.

Brick, J. M. (2013). Unit nonresponse and weighting adjustments: A critical review. *Journal of Official Statistics*, *29*(3), 329-353. doi: 10.2478/jos-2013-0026

1178 Calinescu, M., Schouten, B., & Bhulai, S. (2012). Adaptive survey designs that minimize
1179 nonresponse and measurement risk. *Statistics Netherlands*.

1180 Chow, S. C., Wang, H., & Shao, J. (2008). *Sample Size Calculations in Clinical Research*. Boca
1181 Raton, FL: CRC Press.

1182 Cooper, C. J., Cooper, S. P., Del Junco, D. J., Shipp, E. M., Whitworth, R., & Cooper, S. R.
1183 (2006). Web-based data collection: detailed methods of a questionnaire and data gathering
1184 tool. *Epidemiologic Perspectives & Innovations, 3*(1), 1.

1185 Copeland, K. R., & Ganesh, N. (2015). Sample Weighting for Health Surveys. In T. P. Johnson
1186 (Ed.), *Handbook of Health Survey Methods* (pp. 669-694). Hoboken, NJ: John Wiley & Sons.

1187 Creswell, J. W. (2013). *Qualitative Inquiry and Research Design: Choosing Among Five*
1188 *Approaches*. Los Angeles, CA: SAGE Publications.

1189 Creswell, J. W., & Clark, V. L. P. (2007). *Designing and Conducting Mixed Methods Research*.
1190 Los Angeles, CA: SAGE Publications, Inc.

1191 Diggle, P., Heagerty, P., Liang, K.-Y., & Zeger, S. (2002). *Analysis of Longitudinal Data* (2nd
1192 ed.). New York, NY: Oxford University Press.

1193 Dworkin, S. L. (2012). Sample size policy for qualitative studies using in-depth interviews.
1194 *Archives of Sexual Behavior, 41,* 1319-1320. doi: 10.1007/s10508-012-0016-6

1195 FDA. (2009). Patient-Reported Outcome Measures: Use in Medical Product Development to
1196 Support Labeling Claims. *Guidance for Industry*. Retrieved March 15, 2018, from
1197 https://www.fda.gov/downloads/Drugs/GuidanceComplianceRegulatoryInformation/Guidanc
1198 es/UCM193282.pdf

1199 FDA. (2013). Electronic Source Data in Clinical Investigations. *Guidance for Industry*.
1200 Retrieved March 11, 2018, from
1201 https://www.fda.gov/Drugs/GuidanceComplianceRegulatoryInformation/Guidances/UCM32
1202 8691

1203 FDA. (2014a). Providing Regulatory Submissions in Electronic Format -- Submissions Under
1204 Section 745A(a) of the Federal Food, Drug, and Cosmetic Act. *Guidance for Industry*.
1205 Retrieved March 11, 2018, from
1206 https://www.fda.gov/downloads/drugs/guidances/ucm384686.pdf

1207 FDA. (2014b). Providing Regulatory Submissions In Electronic Format — Standardized Study
1208 Data. *Guidance for Industry*. Retrieved March 11, 2018, from
1209 https://www.fda.gov/Drugs/GuidanceComplianceRegulatoryInformation/Guidances/UCM29
1210 2334

1211 FDA. (2015). Clinical Outcome Assessment (COA): Glossary of Terms. Retrieved March 11,
1212 2018, from

https://www.fda.gov/drugs/developmentapprovalprocess/drugdevelopmenttoolsqualificationprogram/ucm370262.htm

FDA. (2017). Providing Regulatory Submissions in Electronic Format — Certain Human Pharmaceutical Product Applications and Related Submissions Using the eCTD Specifications. *Guidance for Indsutry*. Retrieved March 11, 2018, from https://www.fda.gov/downloads/Drugs/GuidanceComplianceRegulatoryInformation/Guidances/UCM333969.pdf

Fitzmaurice, G. M., Laird, N. M., & Ware, J. H. (2012). *Applied Longitudinal Analysis* (2nd ed.). Hoboken, NJ: John Wiley & Sons.

Fleiss, J. L., Levin, B., & Paik, M. C. (2003). *Statistical Methods for Rates and Proportions* (3rd ed.): John Wiley & Sons.

Fox, F. E., Morris, M., & Rumsey, N. (2007). Doing synchronous online focus groups with young people: Methodological reflections. *Qual Health Res, 17*(4), 539-547. doi: 10.1177/1049732306298754

Francis, J. J., Johnston, M., Robertson, C., Glidewell, L., Entwistle, V., Eccles, M. P., & Grimshaw, J. M. (2010). What is an adequate sample size? Operationalising data saturation for theory-based interview studies. *Psychology & Health, 25*, 1229-1245.

Fricker, R. (2008). Sampling Methods for Web and E-mail Surveys. In N. Fielding, R. M. Lee & G. Blank (Eds.), *The SAGE Handbook of Online Research Methods* (pp. 195-216). Thousand Oaks, CA: SAGE Publications, Inc.

Geldsetzer, P., Fink, G., Vaikath, M., & Barnighausen, T. (2018). Sampling for patient exit interviews: Assessment of methods using mathematical derivation and computer simulations. *Health Services Research, 53*(1), 256-272. doi: 10.1111/1475-6773.12611

Grajales, F. J., Sheps, S., Ho, K., Novak-Lauscher, H., & Eysenbach, G. (2014). Social media: A review and tutorial of applications in medicine and health care. *J Med Internet Res, 16*(2), e13. doi: 10.2196/jmir.2912

Groves, R. M., Fowler Jr, F. J., Couper, M. P., Lepkowski, J. M., Singer, E., & Tourangeau, R. (2009). *Survey Methodology*. Hoboken, NJ: John Wiley & Sons.

Heckathorn, D. D. (1997). Respondent-driven sampling: A new approach to the study of hidden populations. *Social Problems, 44*(2), 174-199.

Hrisos, S., Eccles, M. P., Francis, J. J., Dickinson, H. O., Kaner, E. F. S., Beyer, F., & Johnston, M. (2009). Are there valid proxy measures of clinical behaviour? A systematic review. *Implementation Science, 4*.

International Council for Harmonization. (2017). ICH E9(R1) Statistical Principles for Clinical Trials - Addendum: Estimands and Sensitivity Analysis in Clinical Trials. *Draft Guidance.*

Retrieved April 12, 2018, from https://www.fda.gov/downloads/Drugs/GuidanceComplianceRegulatoryInformation/Guidances/UCM582738.pdf

Jensen, E. B. (2013). A review of methods for estimating emigration. *Washington DC: US Census Bureau (Population Division Working Paper 101).*

Johnson, R. B., & Christensen, L. (2014). *Educational Research*. Thousand Oaks, CA: SAGE Publications, Inc.

Johnson, R. B., & Christensen, L. (2017). *Educational Research: Quantitative, Qualitative, and Mixed Approaches*. Thousand Oaks, CA: SAGE Publications, Inc.

Johnson, T. P. (2015). *Handbook of Health Survey Methods* Hoboken, NJ: John Wiley & Sons.

Korn, E. L., & Graubard, B. I. (1999). *Analysis of Health Surveys*. Hoboken, NJ: John Wiley & Sons.

Levy, P. S., & Lemeshow, S. (2013). *Sampling of Populations: Methods and Applications.* Hoboken, NJ: John Wiley & Sons.

McCarthy, S., O'Raghallaigh, P., Woodworth, S., Lim, Y. L., Kenny, L. C., & Adam, F. (2016). An integrated patient journey mapping tool for embedding quality in healthcare service reform. *Journal of Decision Systems, 25*(Suppl 1), 354-368.

McCullagh, P., & Nelder, J. A. (1989). *Generalized Linear Models* (2nd ed.). Boca Raton, FL: Chapman & Hall.

McGregor, F., Somner, J. E., Bourne, R. R., Munn-Giddings, C., Shah, P., & Cross, V. (2014). Social media use by patients with glaucoma: What can we learn? *Ophthalmic Physiol Opt, 34*(1), 46-52. doi: 10.1111/opo.12093

National Research Council. (2010). *The Prevention and Treatment of Missing Data in Clinical Trials*. Washington, DC: The National Academies Press. https://doi.org/10.17226/12955.

Neuman, W. L. (2014). *Social Research Methods: Qualitative and Quantitative Approaches*. Essex, UK: Pearson Education Ltd.

NSF. (2014). Grant Proposal Guide. Retrieved March 11, 2018, from https://www.nsf.gov/pubs/policydocs/pappguide/nsf15001/gpg_2.jsp#IIC2j

O'Connor, A., Jackson, L., Goldsmith, L., & Skirton, H. (2014). Can I get a retweet please? Health research recruitment and the Twittersphere. *J Adv Nurs, 70*(3), 599-609. doi: 10.1111/jan.12222

O'Neill, R. T., & Temple, R. (2012). The prevention and treatment of missing data in clinical trials: An FDA perspective on the importance of dealing with it. *Clinical Pharmacology & Therapeutics, 91*, 550-554.

Paulus, T. M., & Lester, J. N. (2013). Making learning ordinary: Ways undergraduates display learning in a CMC task. *Text & Talk, 33*(1), 53-70.

Prieto, V. M., Matos, S., Alvarez, M., Cacheda, F., & Oliveira, J. L. (2014). Twitter: A good place to detect health conditions. *Plos One, 9*(1), e86191. doi: 10.1371/journal.pone.0086191

Rothenberg, R. T. (1995). Sampling in social networks. *Connections, 18*, 104-110.

Sandelowski, M. (1995). Sample-size in qualitative research. *Research in Nursing & Health, 18*, 179-183.

Sarndal, C. E., & Lundstrom, S. (2005). *Estimation in Surveys with Nonresponse*. Hoboken, NJ: John Wiley & Sons.

Schouten, B., Calinescu, M., & Luiten, A. (2013). Optimizing quality of response through adaptive survey designs. *Survey Methodology, 39*, 29-58.

Society for Clinical Data Management. (2013). Good Clinical Data Management Practices. Retrieved March 11, 2018, from http://scdm.org/publications/gcdmp/

Stanford University Libraries. (n.d.-a). Data best practices. Retrieved March 14, 2018, from http://library.stanford.edu/research/data-management-services/data-best-practices

Stanford University Libraries. (n.d.-b). Data management plans. Retrieved March 14, 2018, from https://library.stanford.edu/research/data-management-services/data-management-plans

Tates, K., Zwaanswijk, M., Otten, R., van Dulmen, S., Hoogerbrugge, P. M., Kamps, W. A., & Bensing, J. M. (2009). Online focus groups as a tool to collect data in hard-to-include populations: Examples from paediatric oncology. *BMC Med Res Methodol, 9*, 15. doi: 10.1186/1471-2288-9-15

Teddlie, C., & Tashakkori, A. (2009). *Foundations of Mixed Methods Research: Integrating Quantitative and Qualitative Approaches in the Social and Behavioral Sciences*: SAGE Publications, Inc.

Teherani, A., Martimianakis, T., Stenfors-Hayes, T., Wadhwa, A., & Varpio, L. (2015). Choosing a qualitative research approach. *J Grad Med Educ, 7*, 669-670.

Thompson, S. K. (1987). Sample-Size for Estimating Multinomial Proportions. *American Statistician, 41*, 42-46.

Turner, A. G., Angeles, G., Tsui, A. O., Wilkinson, M., & Magnani, R. (2001). Sampling Manual for Facility Surveys for Population, Maternal Health, Child Health and STD Programs in

Developing Countries *MEASURE Evaluation Manual Series, No. 3*: MEASURE Evaluation. Carolina Population Center, University of North Carolina at Chapel Hill.

Valliant, R., Dever, J. A., & Kreuter, F. (2013). *Practical Tools for Designing and Weighting Survey Samples*. New York, NY: Springer.

Varga, M. A., & Paulus, T. M. (2014). Grieving online: Newcomers' constructions of grief in an online support group. *Death Stud, 38*(7), 443-449. doi: 10.1080/07481187.2013.780112

Wilkerson, J. M., Iantaffi, A., Grey, J. A., Bockting, W. O., & Rosser, B. R. (2014). Recommendations for internet-based qualitative health research with hard-to-reach populations. *Qual Health Res, 24*(4), 561-574. doi: 10.1177/1049732314524635

Wilson, H., Dashiell-Aje, E., Anatchkova, M., Coyne, K., Hareendran, A., Leidy, N. K., . . . Wyrwich, K. (2018). Beyond study participants: A framework for engaging patients in the selection or development of clinical outcome assessments for evaluating the benefits of treatment in medical product development. *Quality of Life Research, 27*, 5-16.

Wolf, J. A., Niederhauser, V., Marshburn, D., & LaVela, S. L. (2014). Defining patient experience. Patient Experience Journal, 1(1):7-19. *Patient Experience Journal, 1*, 7-19.

Yin, R. K. (2016). *Qualitative Research From Start to Finish* (2nd ed.). New York, NY: Guilford Press.

GLOSSARY

This glossary defines terms that will be used in the series of methodological Patient-Focused Drug Development (PFDD) FDA guidance documents that are required by the 21st Century Cures Act, and part of commitments made by FDA under the sixth authorization of the Prescription Drug User Fee Act (PDUFA VI). The goal of this glossary is to provide standardized nomenclature and terminologies related to patient-focused medical product development. As appropriate, definitions from existing federal resources (e.g., BEST (Biomarkers, Endpoints, and Other Tools) Resource) have been incorporated into this glossary. External resources were also utilized to define terms and have been cited.

Attribute: An attribute is a feature or characteristic of a medical product—such as efficacy or effectiveness, safety, means of administration, duration of effect, or duration of use—that may affect benefit-risk considerations.

Benefit: Benefits are the favorable effects of a medical product. Types of benefit include clinical benefit *(see definition below)*. Benefits may also include important characteristics of the medical product, such as convenience (e.g., a more convenient dosing regimen or route of administration) that may lead to improved patient compliance, or benefits that affect those other than the patient. (*Source: International Conference on Harmonisation (ICH) Guidelines – Efficacy M4E(R2)*)

Benefit-risk assessment: Evaluation of the demonstrated benefits and risks of a medical product and making a judgment as to whether the expected benefits outweigh the potential risks associated with its expected use.

Biomarker: A defined characteristic that is measured as an indicator of normal biological processes, pathogenic processes, or responses to an exposure or intervention, including therapeutic interventions. Molecular, histologic, radiographic, or physiologic characteristics are types of biomarkers. A biomarker is not an assessment of how an individual feels, functions, or survives. (*Source: BEST (Biomarkers, Endpoints and Other Tools) Resource*)

Caregiver: A person who helps a patient with daily activities, health care, or any other activities that the patient is unable to perform himself/herself due to illness or disability, and who understands the patient's health-related needs. This person may or may not have decision-making authority for the patient and is not the patient's healthcare provider.

Caregiver preference: A statement of the relative desirability or acceptability to caregivers of attributes by which alternative health interventions may differ.

Clinical benefit: A positive clinically meaningful effect of an intervention, e.g., a positive effect on how an individual feels, functions, or survives. (*Source: BEST (Biomarkers, Endpoints and Other Tools) Resource*)

Clinical outcome: An outcome that describes or reflects how an individual feels, functions or survives. (*Source: BEST (Biomarkers, Endpoints and Other Tools) Resource*)

Clinical outcome assessment: Assessment of a clinical outcome can be made through report by a clinician, a patient, a non-clinician observer or through a performance-based assessment. There are four types of COAs: patient-reported outcome (PRO), clinician-reported outcome (ClinRO) measures, observer-reported outcome (ObsRO), and performance outcome (PerfO). (*Source: BEST (Biomarkers, Endpoints and Other Tools) Resource*)

Clinical relevance: The extent to which an endpoint can capture and measure an aspect of a potential clinical benefit (improvement in how the patient feels, functions, and/or survives) that is important from a clinical perspective and from the patient's perspective.

Clinician-reported outcome (ClinRO): A measurement based on a report that comes from a trained health-care professional after observation of a patient's health condition. Most ClinRO measures involve a clinical judgment or interpretation of the observable signs, behaviors, or other manifestations related to a disease or condition. ClinRO measures cannot directly assess symptoms that are known only to the patient (e.g., pain intensity). (*Source: BEST (Biomarkers, Endpoints and Other Tools) Resource*)

Concept (also referred to as concept of interest): In a regulatory context, the concept is the aspect of an individual's clinical, biological, physical, or functional state, or experience that the assessment is intended to capture (or reflect). (*Source: BEST (Biomarkers, Endpoints and Other Tools) Resource*)

Data analysis plan: A roadmap for how the data will be organized and analyzed and how results will be presented. A data analysis plan should be established when planning a research study (i.e., before data collection begins). Among other things, the data analysis plan should describe: (a) the data to be collected; (b) the analyses to be conducted to address the research objectives, including assumptions required by said analyses; (c) data cleaning and management procedures; (d) data transformations, if applicable; and (e) how the study results will be presented (e.g., graphs, tables).

Data management plan (DMP): A written document that describes the data you expect to acquire or generate during the course of your research study; how you intend to manage, describe, analyze, and store said data; and what mechanisms you will use at the end of your study to preserve and share your data. (*Source: Stanford University Libraries n.d.(b)*)

Disease burden: The impacts, direct and indirect, of the patient's health condition that has a negative effect on his or her health, functioning, and overall well-being. Disease burden includes (but is not limited to): the physical and physiologic impacts of the disease and its symptoms; co-morbidities; emotional and psychological effects of the disease, its management, or prognosis; social impacts; effects on relationships; impacts on the patient's ability to care for self and others; time and financial impacts of the disease and its management; and considerations on the impacts on the patient's family.

Endpoint: A precisely defined variable intended to reflect an outcome of interest that is statistically analyzed to address a particular research question. A precise definition of an endpoint typically specifies the type of assessments made, the timing of those assessments, the assessment tools used, and possibly other details, as applicable, such as how multiple assessments within an individual are to be combined. (*Source: BEST (Biomarkers, Endpoints and Other Tools) Resource*)

Fit-for-purpose: A conclusion that the level of validation associated with a medical product development tool is sufficient to support its context of use. (*Source: BEST (Biomarkers, Endpoints and Other Tools) Resource*)

Generalizability: The extent to which study findings can be reliably extended to the target population of interest.

Health literacy: The degree to which individuals have the capacity to obtain, process, and understand basic health information and services needed to make appropriate health decisions. (*Source: U.S. Department of Health and Human Services Quick Guide to Health Literacy)* Health literacy also includes numeracy skills—such as calculating cholesterol and blood sugar levels, measuring medication doses, and understanding nutrition labels—and knowledge of health topics.

Literacy: A person's ability to read, write, speak, and compute and solve problems at levels necessary to: (a) function on the job and in society; (b) achieve one's goals; and (c) develop one's knowledge and potential. (*Source: U.S. Department of Health and Human Services Quick Guide to Health Literacy)*

Methodologically sound: Assurance that the methods and processes used to obtain and analyze patient experience data are rigorous, robust, and adhere to scientifically established principles and best practices for method development or implementation. Evidence generated by methodologically sound methods and processes increases confidence that the results can be trusted, interpreted, and support the intended regulatory uses.

Mixed methods research: Research that uses both qualitative and quantitative research methods. See definitions for qualitative and quantitative research methods.

Observer-reported outcome (ObsRO): A measurement based on a report of observable signs, events or behaviors related to a patient's health condition by someone other than that patient or a health professional. Generally, ObsROs are reported by a parent, caregiver, or someone who observes the patient in daily life and are particularly useful for patients who cannot report for themselves (e.g., infants or individuals who are cognitively impaired). An ObsRO measure does not include medical judgement or interpretation. (*Source: BEST (Biomarkers, Endpoints and Other Tools) Resource*). Examples of ObsROs include a parent report of a child's vomiting episodes or a report of wincing thought to be the result of pain in patients who are unable to report for themselves.

Patient: Any individual with or at risk of a specific health condition, whether or not he or she currently receives any therapy to prevent or treat that condition. Patients are the individuals who directly experience the benefits and harms associated with medical products.

Patient advocate: An individual or group of individuals, who may or may not be part of the target patient population, who has a role in promoting an interest or cause to influence policy with respect to patients' health or healthcare.

Patient-centered: See ***patient-focused***

Patient-centered outcome: An outcome that is important to patients' survival, functioning, or feelings as identified or affirmed by patients themselves, or judged to be in patients' best interest by providers and/or caregivers when patients cannot report for themselves. (*Source: ISPOR Plenary,* Patrick 2013)

Patient engagement: Activities that involve patient stakeholders sharing their experiences, perspectives, needs, and priorities that help inform FDA's public health mission. Such activities may include (but are not limited to): testimony at Advisory Committee meetings, submission to regulations.gov public docket; meetings attended by patients, FDA, and other stakeholders; other correspondence with FDA; interactions through social media; and interactions with or information from patient representatives or patient advocates.

Patient experience data: Defined in Title III, section 3001 of the 21st Century Cures Act, as amended by section 605 of the FDA Reauthorization Act of 2017 (FDARA),[19] and includes data that are collected by any persons and are intended to provide information about patients' experiences with a disease or condition. Patient experience data can be interpreted as information that captures patients' experiences, perspectives, needs, and priorities related to (but not limited to): 1) the symptoms of their condition and its natural history; 2) the impact of the conditions on their functioning and quality of life; 3) their experience with treatments; 4) input on which outcomes are important to them; 5) patient preferences for outcomes and treatments; and 6) the relative importance of any issue as defined by patients.

Patient-focused (also referred to as ***patient-centered*****):** Ensuring that patients' experiences, perspectives, needs, and priorities are meaningfully incorporated into decisions and activities related to their health and well-being.

Patient-focused drug development (PFDD) (also referred to as ***patient-focused medical product development*****:** A systematic approach to help ensure that patients' experiences,

[19] "PATIENT EXPERIENCE DATA.—For purposes of this section, the term 'patient experience data' includes data that (1) are collected by any persons (including patients, family members and caregivers of patients, patient advocacy organizations, disease research foundations, researchers, and drug manufacturers); and (2) are intended to provide information about patients' experiences with a disease or condition, including (A) the impact (including physical and psychosocial impacts) of such disease or condition, or a related therapy, on patients' lives; and (B) patient preferences with respect to treatment of such disease or condition." The definition is codified at section 569C(c)(4) of the FD&C Act, and applies to section 3002 of the 21st Century Cures Act, which directed FDA to issue certain guidance documents regarding the collection of patient experience data, see section 3002(b).

perspectives, needs, and priorities are captured and meaningfully incorporated into the development and evaluation of medical products throughout the medical product life cycle.

Patient input: Information that captures patients' experiences, perspectives, needs, and priorities. See *Patient Experience Data*.

Patient partner: An individual patient, caregiver or patient advocacy group that engages other stakeholders to ensure the patients' wants, needs and preferences are represented in activities related to medical product development and evaluation. (*Source:* Wilson *et al*, 2018)

Patient perspective: A type of patient experience data that specifically relates to patients' attitudes or points of view about their condition or its management. Patient perspectives may include (but are not limited to): perceptions, goals, priorities, concerns, opinions, and preferences.

Patient preference: A statement of the relative desirability or acceptability to patients of specified alternatives or choice among outcomes or other attributes that differ among alternative health interventions. *(Source: FDA Guidance on PPI for medical devices)*

Patient preference information (PPI): Assessments of the relative desirability or acceptability to patients of specified alternatives or choices among outcomes or other attributes that differ among alternative health interventions. The methods for generating PPI may be qualitative, quantitative, or mixed methods. *(Source: FDA Guidance on PPI for medical devices)*

Patient-provided input: *Patient experience data* or other information that comes directly from patients.

Patient-reported outcome (PRO): A measurement based on a report that comes directly from the patient (i.e., study subject) about the status of a patient's health condition without interpretation of the patient's response by a clinician or anyone else. A PRO can be measured by self-report or by interview, provided that the interviewer records only the patient's response. Symptoms or other unobservable concepts known only to the patient (e.g., pain severity or nausea) can only be measured by PRO measures. PROs can also assess the patient perspective on functioning or activities that may also be observable by others. (*Source: BEST (Biomarkers, Endpoints and Other Tools) Resource*)

Patient representative: An individual, who may or may not be part of the target population, who has direct experience with a disease or condition (e.g., a patient or caregiver) and can provide information about a patient's experience with the disease or condition.

Performance outcome (PerfO): A measurement based on a standardized task(s) performed by a patient that is administered and evaluated by an appropriately trained individual or is independently completed. PerfOs require patient cooperation and motivation. These include measures of gait speed (e.g., timed 25 foot walk test), memory recall (e.g., word recall test), or

other cognitive testing (e.g., digit symbol substitution test). (*Source: BEST (Biomarkers, Endpoints and Other Tools) Resource*)

Qualitative research methods: Methods associated with the gathering, analysis, interpretation, and presentation of narrative information (e.g., spoken or written accounts of experiences, observations, and events). Qualitative research methods may also include direct observations (e.g., non-verbal communication and behaviors).

Quantitative research methods: Methods associated with the gathering, analysis, interpretation, and presentation of numerical information.

Real-World Data (RWD): Data relating to patient health status and/or the delivery of health care routinely collected from a variety of sources. (*Source: FDA Guidance on Use of Real-World Evidence to Support Regulatory Decision-Making for Medical Devices*)

Real world evidence (RWE): The clinical evidence regarding the usage and potential benefits or risks of a medical product derived from analysis of real-world data. (*Source: FDA Guidance on Use of Real-World Evidence to Support Regulatory Decision-Making for Medical Devices*)

Reporter: In research studies designed to collect patient experience data, the reporter is the individual, group of individuals, or entity providing patient experience data. Reporters may be patients, parents, sexual/romantic partners, caregivers, physicians, or other healthcare professionals. Selection of an appropriate reporter in a given research study will depend on the definition of the target patient population of interest. If a patient in the target population can be reasonably expected to reliably self-report, then one would expect the patient herself/himself to be the reporter in that research study.

Representativeness: Confidence that a sample from which evidence is generated is sufficiently similar to the intended population. In the context of patient experience data, representativeness includes the extent to which the elicited experiences, perspectives, needs, and priorities of the sample are sufficiently similar to those of the intended patient population.

Research protocol: A document that describes the background, rationale, objectives, design, methodology, statistical considerations, and organization of a clinical research project. (*Source: UCSF Clinical Research Resource HUB*) A research protocol guides the study and associated data collection and analysis in a productive and standardized manner.

Risk: Risks are adverse events and other unfavorable effects associated with a medical product. Risks include drug interactions, risks identified in the non-clinical data, risks to those other than the patient (e.g., fetus, those preparing and administering the medical product), and risks based on pharmacologic class or current knowledge of the product. Factors such as potential misuse, abuse, or diversion of the product may also be considered. (*Source: International Conference on Harmonisation (ICH) Guidelines – Efficacy M4E(R2)*)

Risk tolerance: The degree to which a patient would accept increased probability or severity of a harm in exchange for a specific expected benefit. *(Source: Medical Device Innovation Consortium (MDIC) Patient Centered Benefit-Risk Project Report)*

Science of patient input: Methods and approaches of systematically obtaining, analyzing, and using information that captures patients' experiences, perspectives, needs, and priorities in support of the development and evaluation of medical products.

Social Media: Web-based tools that are used for computer-mediated communication. Social media may include but is not limited to: (1) blogs, (2) microblogs, (3) social networking sites, (4) professional networking sites, (5) thematic networking sites, (6) wikis, (7) mashups, (8) collaborative filtering sites, (9) media sharing sites, and others. (*Source: Grajales III et al. 2014*)

Subgroup: A subset of the study population or study sample defined by specific baseline characteristics. For example, demographic subgroups are commonly defined by subject sex, race, and age.

Surrogate endpoint: A type of endpoint used in clinical trials as a substitute for a direct measure of how a patient feels, functions, or survives. A surrogate endpoint does not measure the clinical benefit of primary interest in and of itself but rather is expected to predict that clinical benefit or harm based on epidemiologic, therapeutic, pathophysiologic, or other scientific evidence. From a U.S. regulatory standpoint, surrogate endpoints and potential surrogate endpoints can be characterized by the level of clinical validation: (a) validated surrogate endpoints; (b) reasonably likely surrogate endpoints; and (c) candidate surrogate endpoints. (*Source: BEST (Biomarkers, Endpoints and Other Tools) Resource*)

Target population (also referred to as the *target patient population*, the *underlying population*, or *intended population*)**:** The group of individuals (patients) about whom one wishes to make an inference.

Trade-off: The extent to which a change in the level of one or more attributes of a medical product that is offset by a change in one or more other attributes of that product. *(Source: Medical Device Innovation Consortium (MDIC) Patient Centered Benefit-Risk Project Report)*

Treatment burden: The impacts of a specific treatment or treatment regimen that have a negative effect on the patient's health, functioning, or overall well-being. Treatment burden includes (but is not limited to): side effects, discomfort, uncertainty about treatment outcomes, dosing and route of administration, requirements, and financial impacts.

Treatment effect: The amount of change in a disease/condition, symptom, or function that results from a medical intervention (as compared to not receiving the intervention or receiving a different intervention).

Treatment outcome: The benefits or harms to a patient who receives an intervention; the impact on a patient's health, function, or well-being—or on a clinical indicator thereof—that is assumed

to result from an intervention. *(Source: Patient-Centered Outcomes Research Institute (PCORI) Methodology Report)*

Unmet medical need: An unmet medical need is a condition whose treatment or diagnosis is not addressed adequately by available therapy. An unmet medical need includes an immediate need for a defined population (e.g., to treat a serious condition with no or limited treatment) or a longer-term need for society (e.g., to address the development of resistance to antibacterial drugs). *(Source: FDA Guidance on Expedited Programs for Serious Conditions)*

APPENDICES

Appendix 1. Methods for Collecting Patient Experience Data

1. *Qualitative Research Methods*

Some aims of qualitative research are shown in **Table 7**.

Table 7. Qualitative Research Aims

Qualitative Research Aims[20]	Examples of potential research questions	Illustrative examples of qualitative data generation by question type
To Understand 'What'	• What symptoms do heart failure patients experience?	• Patient #101: "I have trouble going up long flights of stairs."
	• What signs do caregivers observe that tell them their loved one is having asthma symptoms?	• Caregiver #201: "I know my daughter is having a hard time with her asthma when she is wheezing."
	• Based on your experience with COPD patients, what would you consider to be signs of severe COPD?	• Clinician #301: "When a patient presents with typical symptoms and has had more than one COPD flare per year or if they have been hospitalized due to your COPD, I would consider them severe."
To Explore 'Why'	• Why are asthma symptoms bothersome to you?	• Patient #101: "My asthma prevents me from being able to exercise without an inhaler."
	• Why do you prefer the auto-injector to intravenous (IV) injection?	• Patient #201: "The auto-injector is more convenient because I can administer it at home and it takes less time. My IV injections require a clinic visit and take hours."
To Examine 'How'	• How have arthritis symptoms impacted a patient's mobility?	• Patient #101: "My knees are stiff because of my arthritis. I find it hard to go up and down stairs."
	• How has a patient's dementia impacted the relationship dynamics in their family?	• Caregiver #201: "My mom now requires 24-hour care. I'm often stressed about this and it's putting a strain on my marriage."
	• How have symptoms improved with treatment?	• Patient #101: "Since receiving my lupus treatments, I've not

[20]Note: any qualitative study could address one or more of these aims

Qualitative Research Aims[20]	Examples of potential research questions	Illustrative examples of qualitative data generation by question type
		been in the hospital as much as before."

2. *Quantitative Research Methods*

Table 8 summarizes some potential aims of quantitative research.

Table 8. Quantitative Research Aims

Quantitative Research Aims	Examples of potential research questions	Examples of potential quantities of interest
To Describe	• How many (proportion of) patients experience stomach pain symptoms? • How frequently do epileptic patients experience seizures in a week? • How severe are patients' heartburn symptoms? • Please rank your 3 most bothersome symptoms.	Frequencies, proportions, means, medians, distributions
To Compare	• What is the difference in daily exacerbations among mild, moderate, and severe COPD patients? • To what extent do questionnaire responses differ among members of separate subgroups?	Differences in frequencies, proportions, means, medians
To Relate	• What is the correlation between patient-reported sleep disturbance and actigraphy ratings?	• Measures of association, trend, or interaction

3. *Software for Analyzing Quantitative Patient Experience Data*

Statistical software is available for analyzing quantitative patient experience data. Some commonly used statistical software includes R, SAS, SPSS, and SUDAAN. SUDAAN, SAS, STATA, and R are also commonly used to analyze survey data obtained from probability sampling as each permits the specification of relevant design features such as clustering, stratification, weights, and methods of variance estimation.

Regardless of which software you use, we recommend checking the current defaults and computational algorithms utilized as they vary both across software and across versions of the same software. Different estimation procedures and defaults may generate different results. In addition, be sure to note which software *version* was used as part of the study analysis documentation.

Appendix 2. Standards and Requirements Pertaining to Submission of Data

Regulations, guidances, standards, and requirements pertaining to capture/collection, transmission, processing, storage, archiving, retention, and submission of data from clinical studies include (but are not limited to):

- FDA forms and submission requirements (https://www.fda.gov/drugs/developmentapprovalprocess/formssubmissionrequirements/)
- The International Council on Harmonisation (ICH) Guidelines[21], such as *ICH Harmonised Guideline for Good Clinical Practice: E6(R2)* and the *Electronic Common Technical Document (eCTD)*
- 21 eCFR, Volumes 1 – 8[22]
- Guidance for Industry on Providing Regulatory Submissions In Electronic Format—Standardized Study Data (FDA, 2014b)
- Guidance for Industry on Providing Regulatory Submissions in Electronic Format—Submissions Under Section 745A(a) of the Federal Food, Drug, and Cosmetic Act (FDA, 2014a)
- Guidance for Industry on Providing Regulatory Submissions in Electronic Format—Certain Human Pharmaceutical Product Applications and Related Submissions Using the eCTD Specifications (FDA, 2017)
- Guidance for Industry on Electronic Source Data in Clinical Investigations (FDA, 2013)
- The FDA Data Standards Catalog.

Current and more detailed information on study data standards resources, please see: https://www.fda.gov/ForIndustry/DataStandards/StudyDataStandards/default.htm.

[21] https://www.fda.gov/RegulatoryInformation/Guidances/default.htm

[22] https://www.ecfr.gov/cgi-bin/text-idx?tpl=/ecfrbrowse/Title21/21tab_02.tpl

Appendix 3. Considerations for Data Management

Resources you may wish to consider when developing a data management plan include:

- Stanford University Libraries' guide to DMPs (Stanford University Libraries, n.d.-b);
- The Society for Clinical Data Management's (SCDM) standard for Good Clinical Data Management Practices (Society for Clinical Data Management, 2013); and
- Data management considerations laid forth in the National Science Foundation (NSF) Grant Proposal Guide Chapter II.C.2.j (NSF, 2014).

Components of good data management plans and practices include (NSF, 2014; Society for Clinical Data Management, 2013; Stanford University Libraries, n.d.-a):

- Having a complete draft of the DMP "prior to enrollment of the first study subject" and ensuring that "an approved, signed version of the DMP is completed prior to starting on the work it describes" (Society for Clinical Data Management, 2013);
- Ensuring "compliance with applicable regulations and oversight agencies" (Society for Clinical Data Management, 2013);
- Identifying and defining the "personnel and roles involved with decision making, data collection, data handling, and data quality control" (Society for Clinical Data Management, 2013);
- Ensuring "data management processes are described and defined from study initiation until database closeout" (Society for Clinical Data Management, 2013);
- Developing the DMP "in collaboration with all stakeholders to ensure that all responsible parties understand and will follow the processes and guidelines put forth in the DMP from study initiation to database closeout" (Society for Clinical Data Management, 2013);
- Developing and maintaining a DMP template that "ensures consistency and standardization across all projects" (Society for Clinical Data Management, 2013);
- Ensuring the DMP for each study is "kept current, including proper versioning, and that all responsible parties are aware of and agree to the current content" (Society for Clinical Data Management, 2013);
- Pre-specifying the types of data to be collected over the course of the study (NSF, 2014);
- Using standard, predetermined structure(s) for collecting patient experience data (e.g., interview scripts, questionnaire layouts, electronic devices, telephone prompts, etc.);

- Specifying "standards to be used for data and metadata format and content" (NSF, 2014);

- Using "descriptive and informative file names" (Stanford University Libraries, n.d.-a);

- Choosing "file formats that will ensure long-term access" to the data (Stanford University Libraries, n.d.-a);

- Having a systematic method for tracking different versions of datasets and documents (e.g., data and metadata) (Stanford University Libraries, n.d.-a);

- Creating metadata for each analysis performed (Stanford University Libraries, n.d.-a);

- Having processes in place to ensure compliance with regulatory requirements f the protection and ownership of source data (Society for Clinical Data Management, 2013);

- Having policies in place for accessing and sharing data, including:

    - Provisions for appropriate protection of privacy, confidentiality, security, intellectual property, or other rights or requirements (NSF, 2014);

    - "Policies and provisions for re-use, re-distribution, and the production of derivatives" (NSF, 2014);

- Handling sensitive, confidential, and personally identifiable information and data in an appropriate manner, including ensuring an appropriate level of network and infrastructure security (Society for Clinical Data Management, 2013; Stanford University Libraries, n.d.-a); and

- Planning how data, samples, and other research products will be archived, and how access to these materials will be preserved for future access (NSF, 2014).

Other considerations and recommendations include:

- Data validation rules and electronic edit checks should be programmed to enhance data quality at the point of data entry. Prior to database lock, appropriate quality control measures should be taken to ensure that records with inconsistent values of variables (e.g., age or gender) are identified, examined, and addressed.

- For observational studies, ensure proper logistics are in place to collect and manage data generated by follow-up queries, if needed. Variables should be cross-checked to verify subgroup assignment, subject disposition, reason for exclusion (where applicable), and type of error(s) detected in the record, if any.

- Researcher(s) should design data management features to enhance data quality, minimize missing or erroneous data, and minimize data cleaning. In addition, use of customized error messages and automated data validations may facilitate survey completion.

If a research subject is excluded from an analysis, the reason for excluding the data collected from said subject should be thoroughly documented (and included in your submission to FDA).