collection methods can include paper-based, computer-assisted, and telephone-based assessments. We intend to review the comparability of data obtained when using multiple data collection methods or administration modes within a single clinical trial to determine whether the treatment effect varies by method or mode. If a patient diary or some other form of unsupervised data entry is used, we plan to review the clinical trial protocol to determine what steps are taken to ensure that patients make entries according to the clinical trial design and not, for example, just before a clinic visit when their reports will be collected.

3. Recall Period

Sponsors should also evaluate the rationale and the appropriateness of the recall period for a PRO instrument. To this end, it is important to consider patient ability to validly recall the information requested. The choice of recall period that is most suitable depends on the instrument's purpose and intended use; the variability, duration, frequency, and intensity of the concept measured; the disease or condition's characteristics; and the tested treatment. When evaluating PRO-based claims, we intend to review the clinical trial protocol to determine what steps were taken to ensure that patients understood the instrument recall period. In many cases, what is of real interest is not the integrated effect over a short time period (e.g., 2-week period), but the effect at regular intervals (e.g., 2, 4, and 6 weeks), similar to how measurements might be made every 2 weeks in a blood pressure trial. In that case, patients can be asked to report on recent status. Note also that any problems created by differential recall are likely to add noise and obscure treatment effects.

PRO instruments that call for patients to rely on memory, especially if they must recall over a long period of time, compare their current state with an earlier period, or average their response over a period of time, are likely to undermine content validity. Response is likely to be influenced by the patient's state at the time of recall. For these reasons, items with short recall periods or items that ask patients to describe their current or recent state are usually preferable. If detailed recall of experience over a period of time is necessary, we recommend the instrument use appropriate methods and techniques for enhancing the validity and reliability of retrospectively reported data (e.g., ask patients to respond based on their worst (or best) experience over the recall period or make use of a diary for data collection).

4. Response Options

It is also important to consider whether the response options for each item are consistent with its purpose and intended use. Table 3 describes some of the various types of item response options that are typically seen in PRO instruments.

Table 3. Response Option Types

Type	Description
Visual analog scale (VAS)	A line of fixed length (usually 100 mm) with words that anchor the scale at the extreme ends and no words describing intermediate positions. Patients are instructed to indicate the place on the line corresponding to their perceived state. The mark's position is measured as the score.
Anchored or categorized VAS	A VAS that has the addition of one or more intermediate marks positioned along the line with reference terms assigned to each mark to help patients identify the locations between the scale's ends (e.g., half-way).
Likert scale	An ordered set of discrete terms or statements from which patients are asked to choose the response that best describes their state or experience.
Rating scale	A set of numerical categories from which patients are asked to choose the category that best describes their state or experience. The ends of rating scales are anchored with words but the categories are numbered rather than labeled with words.
Recording of events as they occur	Specific events are recorded as they occur using an event log that can be included in a patient diary or other reporting system (e.g., interactive voice response system).
Pictorial scale	A set of pictures applied to any of the other response option types. Pictorial scales are often used in pediatric questionnaires but also have been used for patients with cognitive impairments and for patients who are otherwise unable to speak or write.
Checklist	Checklists provide a simple choice between a limited set of options, such as *Yes*, *No*, and *Don't know*. Some checklists ask patients to place a mark in a space if the statement in the item is true. Checklists are reviewed for completeness and nonredundancy.

Item response options generally are considered appropriate when:

- Wording used in responses is clear and appropriate (e.g., anchoring a ***scale*** using the term *normal* assumes that patients understand what is normal for the general population).

- The item response options are appropriate for the intended population. For example, patients with visual impairment may find a VAS difficult to complete.

- Responses offer a clear distinction between choices (e.g., patients may not distinguish between *intense* and *severe* if both are offered as response choices to describe their pain).

- Instructions to patients for completing items and selecting responses for the items are adequate.

- The number of response options is justified empirically (e.g., using qualitative research, initial instrument testing, or existing literature).

- Responses for an item are appropriately ordered and represent similar intervals.

- Responses for items avoid potential ceiling or floor effects (e.g., it may be necessary to introduce more responses to capture worsening or improvement so that fewer patients respond at the response continuum top or bottom).

- Responses do not bias the direction of responses (e.g., bias exists if possible responses are weighted toward the severity spectrum's mild end with two severity options for *mild* and only one each for *moderate* and *severe*).

5. Instrument Format, Instructions, and Training

Results obtained using a PRO instrument can vary according to the instructions given to patients or the training given to the interviewer or persons supervising PRO data collection during a clinical trial. Sponsors should consider all PRO instrument instructions and procedures contained in publications and user manuals provided by developers, including procedures for reviewing completed questionnaires and procedures used to avoid missing data or clarify responses.

It is important that the PRO instrument format used in the clinical trial be consistent with the format that is used during the instrument development process. *Format* refers to the exact questionnaire, diary, or interview script appearance used to collect the PRO data. Format is specific to the administration mode and the data collection method. We plan to review the specific format used in the clinical trial including the order and numbering of items, the presentation of response options in single response or grid formats, the grouping of items, patterns for skipping questions, and all instructions to interviewers or patients.

We recommend that the user manual provided by a developer during the PRO instrument development process specify how to incorporate the instrument into a clinical trial in a way that minimizes administrator burden, patient burden, missing data, and poor data quality. The user manual should explain to investigators and interviewers critical principles of PRO administration.

6. Patient Understanding

When the initial and subsequent drafts of an instrument are prepared, sponsors are encouraged to examine all items and procedures in a pilot test of whether patients understand the items and instructions included in the PRO instrument. This examination should include documentation that the concepts represented in the PRO instrument's conceptual framework are confirmed, that the response options and recall period are appropriately comprehended, and that the instrument's readability is adequate for the intended population. The FDA's evaluation of these procedures is likely to include a review of a cognitive interviewing report containing the script used in patient cognitive interviews, the interview transcripts, the readability test used (if applicable), the ***usability testing*** process description (if applicable), the cognitive interviews analysis, and the actions taken to delete or modify items, response scales, or patient instructions in response to the cognitive interview or pilot test results. Evidence from the patient cognitive interview studies (i.e., the interview schedule, transcript, and listing of all concepts elicited by a single item) can be used to determine when a concept is adequately captured. Repeating cognitive interviews can help confirm content validity.

7. Scoring of Items and Domains

For each item, numerical scores generally should be assigned to each answer category based on the most appropriate scale of measurement for the item (e.g., nominal, ordinal, interval, or ratio