Choice of Recall Period for Patient-Reported Outcome Measures: Criteria for Consideration

Josephine M. Norquist,[1] Cynthia Girman,[1] Sheri Fehnel,[2]
Carla DeMuro,[2] Nancy Santanello[1]

[1]Epidemiology, Merck Sharp & Dohme, Corp., North Wales, PA (USA); [2]RTI Health Solutions, Research Triangle Park, NC (USA)

Abstract

Aims: Understand the choice of recall period for PRO measures based on intended use, characteristics of the disease, treatment, and attributes of studies in which the measure will be used.

Methods: The FDA PRO Guidance suggests a preference for patients to describe their current or recent health state with as short a recall period as possible. Current practice and considerations were reviewed within a variety of disease areas (overactive bladder, menopausal hot flashes, niacin-induced flushing, osteoarthritis pain, irritable bowel symptoms, benign prostatic hyperplasia, and alopecia) where the choice of recall may depend on the rate of disease progression, frequency, fluctuation and burden of symptoms. Additionally, SEALD reviews were examined for feedback on recall periods.

Results: Across disease areas, rationales were identified for using different recall periods, including event-driven (immediate), daily, up to weekly, and longer than weekly recall periods. This work demonstrates that: 1) length of recall varies depending on what the PRO measure captures, its intended use, and attributes of the disease and study; 2) within the same disease area, recall can vary depending on the concept or phenomenon of interest (e.g. variability, frequency, or overall impact); 3) recalls must consider patient burden and ability of patients to easily and accurately recall the information requested; 4) recall must be consistent with the duration of the trial and the scheduled clinic visits.

Conclusions: The selection of the recall period is an important decision in the development of a PRO measure. Shorter recall periods are being encouraged at the US regulatory level for PROs supporting label claims. These may underestimate symptom burden when symptoms have diurnal or day-to-day fluctuation and may place undue burden on patients. On the other hand, recall intervals that are too long may either over- or under-estimate the health state. Therefore, a one size fits all approach is not effective, and a variety of factors should be considered to optimize data quality and completeness.

Introduction

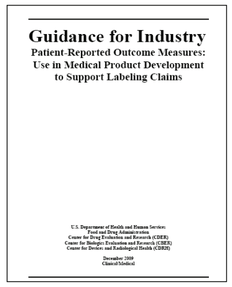
- Consistent criteria for selecting the most appropriate recall period when developing or choosing a patient-reported outcome (PRO) measure for use in clinical trials are needed[1]
- While some researchers and regulators have suggested that very short recall periods are better, choice of an optimal recall period may depend on a number of factors
- We describe a number of criteria and examples within various therapeutic areas where the choice of recall may depend on the rate of disease progression, frequency and fluctuation of symptoms, and burden of symptoms.

Background

FDA Guidance for Industry PRO Measures

- Final Guidance released Dec 2009[2]
  - Applies to all PROs intended for labeling or promotion in the U.S.
  - Specific requirements for PRO development, validation and interpretation
  - Increased need for rigor in study design, data collection and documentation

A PRO is defined as any report of the status of a patient's health condition that comes directly from the patient, without interpretation of the patient's response by a clinician or anyone else


Guidance for Industry

Patient-Reported Outcome Measures: Use in Medical Product Development to Support Labeling Claims

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)
Center for Devices and Radiological Health (CDRH)

December 2009
Clinical/Medical

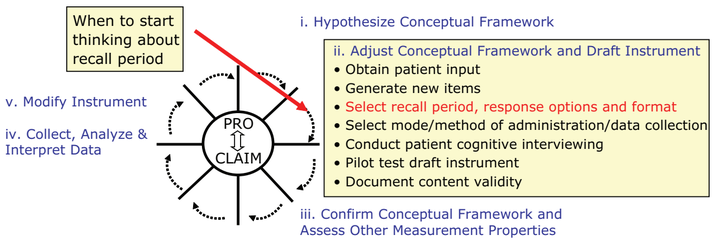
- **Recall period** — "The period of time patients are asked to consider in responding to a PRO item or question. Recall can be momentary (real time) or retrospective of varying lengths."

*" …. 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."*

*"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. …., **items with short recall periods or items that ask patients to describe their current or recent state are usually preferable.**"*

Results

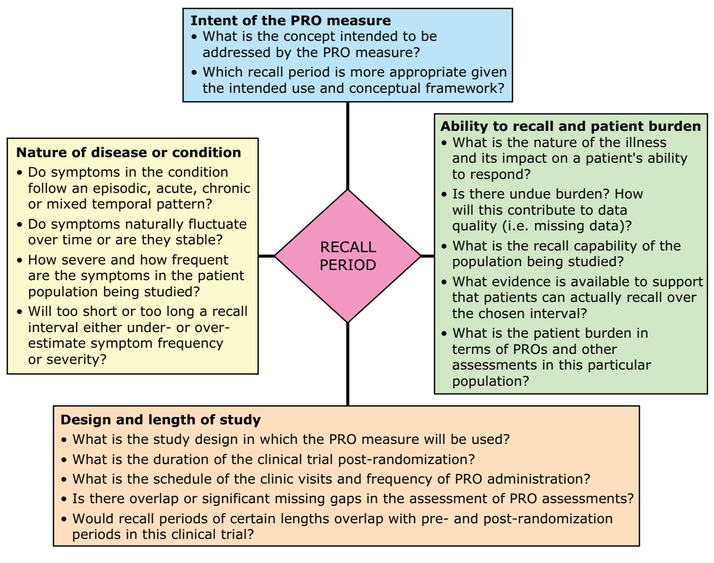
Development of a PRO Measure: An Iterative Process[2]



Researchers should

- Think about the recall period prior to obtaining patient input
- Use interviews to understand patient perspective of recall and what may be appropriate

Considerations in Selecting Length of Recall Period for PRO Measures



Examples of Daily and Event Driven Recall Periods and Rationale

Disease Condition	Symptoms	Recall	Rationale
Menopausal Hot Flashes [3]	Hot flashes	Event Log	Episodic in nature
Overactive Bladder (OAB)	Micturition, urinary incontinence and urgency	Event Log or Daily	Day-to-day variation. Symptoms typically recorded whenever they occur using a 3- to 7-day voiding diary
Irritable Bowel Syndrome - Diarrhea predominant [4-9]	Bowel and abdominal symptoms	Event Log or Daily	Bowel symptoms (e.g., stool consistency, straining) pertain to each bowel movement, whereas abdominal symptoms (e.g., pain, bloating) may be assessed daily
Niacin-Induced Flushing (NIF) [10]	Cutaneous flushing (redness, warmth, tingling and/or itching)	Daily	Day-to-day variation in severity and bother

Example of Weekly Recall Periods and Rationale

Disease Condition	Symptoms	Rationale
Facial Acne [11]	Whiteheads, blackheads, and/or inflamed red lesions (papules, pustules, and cysts)	Visibility and severity may not vary from day to day. More frequent assessments may be impractical based on patient burden and low compliance among of young adults
Intermittent Asthma	Cough, wheezing, and/or difficulty breathing	Patients may not experience symptoms every day, and symptoms may occur only during exacerbations
Osteoarthritis (OA)-related Functioning [12]	Joint pain	Symptoms not necessarily experienced daily

Example of Longer than Weekly Recall Periods and Rationale

Disease Condition	Symptoms	Rationale
Benign Prostatic Hyperplasia [13]	Urinary storage and voiding	Chronic condition that progresses quite gradually. Symptoms can wax and wane over time
Alopecia or Female Pattern Hair Loss (FPHL) [14]	Increased hair shedding, decreased hair density and finer hair	Disease progression and improvements are gradual. Subjects may need to look back over a period of time to determine how often or severely they were impacted by their hair loss
OAB Symptom Improvement	Changes in micturition, urinary incontinence and urgency	May be appropriate to have a long recall when assessing bother, disease-specific HRQL or improvement in overall symptoms

Summary of Different Recall Periods and Rationale

Type of Recall	Rationale for Selection
Event Driven (Immediate) or Daily Recall	• Nature of disease/condition is episodic (e.g. OAB, NIF) • Best assessed using event logs or diaries • Symptom frequency may be unpredictable (e.g. OAB) • Symptoms only noticeable in certain situation (e.g. vaginal dryness and dyspareunia during sexual intercourse) • Day-to-day variation in symptoms may be pronounced
Up to Weekly Recall	• Symptoms not necessarily experienced daily (e.g. intermittent asthma) • Interested in integrated assessments of severity or effects of medications (e.g. facial acne) over a period of time • Interested in impact on non-daily events or opportunities (e.g., social functioning, recreational activities)
Longer than Weekly Recall	• Changes in micturition, urinary incontinence and urgency

Examples of other FDA guidances that mention recall period for PROs

Chronic Obstructive Pulmonary Disease: Developing Drugs for Treatment[15]

- "… two efficacy endpoints may need to be declared as primary endpoints in phase 3 studies to support efficacy. An example of using two primary efficacy endpoints would be measurement of lung function, such as FEV1, plus a measure of a patient-reported outcome, such as a validated symptom score, activity scale, or disease-specific, health-related quality-of-life instrument"
  - "…scales that require patients to r**ecall prior symptoms (e.g., how do you feel now compared to baseline?) are problematic, because patients' memory may fade over time, particularly in studies lasting several months"**
  - Guidance recognizes important of St. George's Respiratory Questionnaire and Chronic Respiratory Questionnaire, yet no clear recommendations of appropriate recall period to use

Guidance for Industry Sinusitis: Designing Clinical Development Programs of Nonantimicrobial Drugs for Treatment[16]

- "The efficacy endpoint should include patient-reported symptoms and at least one objective measure declared as co-primary, meaning that both measures should statistically demonstrate the desired effect."
  - "The frequency of scoring should be driven by the dosing interval, but should be **at least twice daily**, once in the morning and once in the evening, with one or more scorings timed to precede dosing. The symptoms should be scored both as reflective score (evaluation of symptom severity over a predefined period, such as 12 hours) and as instantaneous score (evaluation of symptom severity immediately preceding the time of scoring). …"

Some recent SEALD reviews

Stelara® (Ustekinumab)

- Stelara™ is indicated for the treatment of adult patients (18 years or older) with moderate to severe plaque psoriasis who are candidates for phototherapy or systemic therapy.
- FDA approval September 2009
- SEALD feedback on Dermatology Life Quality Index (DLQI)
  - "…justification for the recall period for the DLQI was not included in the submission"
  - "The **one week recall period** used in the clinical trial may introduce bias…" [DLQI]
  - "…if there is variability in a symptom from day to day, it is unclear whether patients can adequately recall their experiences over a one week or two week period in an unbiased way"

SEALD, 2009. Study Endpoint Review.

Flibanserin[17]

- Flibanserin is a 5-hydroxytryptophan (HT)(1A) receptor agonist and 5-HT(2A) receptor antagonist for the treatment of hypoactive sexual desire disorder (HSDD)
- FDA advisory committee voted against approval in June 2010 due to lack of evidence of efficacy as compared to placebo
  - Also lack of confidence in the Female Sexual Function Index (FSFI) which was the primary endpoint
  - Concern was use of 4-week recall and potential recall bias

  "The Division never agreed that the FSFI sexual desire domain was an acceptable primary or key secondary endpoint. On the contrary, the Division provided early guidance in a 2005 meeting that the evaluation of change in sexual desire (a critical component for both the diagnosis of HSDD and treatment evaluation of flibanserin) **should not be based on a 28-day recall** (as is the case with the FSFI sexual desire domain), **but rather on a more frequent recording** (i.e., daily) of the endpoint. For this reason, the Division recommended that the eDiary Question 2 ("how would you rate your level of sexual desire?") be used as a key efficacy endpoint."

References

1. Stull DE, Leidy NK, Parasuraman B, Chassany O. Optimal recall periods for patient-reported outcomes: challenges and potential solutions. CMRO 2009; 24(4): 929-942.
2. Food and Drug Administration. Guidance for Industry: Patient-reported outcome measures: use in medical product development of support labeling claims. Federal Register 2009; http://www.fda.gov/downloads/Drugs/GuidanceComplianceRegulatoryInformation/Guidances/UCM193282.pdf
3. Williams RE, Kalilani L, DiBenedetti DB, Zhou X, Granger AL, Fehnel SE, et al. Frequency and severity of vasomotor symptoms among peri- and postmenopausal women in the United States. Climacteric 2008;11(1):32-43.
4. Camilleri M, Chey WY, Mayer EA, Northcutt AR, Heath A, Dukes GE, McSorley D, Mangel AM. A randomized controlled clinical trial of the serotonin type 3 receptor antagonist alosetron in women with diarrhea-predominant irritable bowel syndrome. Archives of Internal Medicine 2001;161(14):1733-40.
5. Leventer S, Raudibaugh K, Frissora C, Mangel A, Galbraith KB, Kucharik R, Ye ND, Phillips J, Keogh JC, Keim KL. The safety and efficacy of dextofisopam in patients with diarrhea-predominant or alternating irritable bowel syndrome. Gastroenterology 2005;128(4):A94.
6. Nyhlin H, Bang C, Elsborg L, Silvennoinen J, Holme I, Ruegg P, Jones J, Wagner A. A double-blind, placebo-controlled, randomized study to evaluate the efficacy, safety and tolerability of tegaserod in patients with irritable bowel syndrome. Scandinavian Journal of Gastroenterology 2004;39(2):119-26.
7. Miner P, Stanton DB, Carter F, Caras S, Krause G, Steinborn C. Cilansetron in irritable bowel syndrome with diarrhea predominance (IBS-D): Efficacy and safety in a 3 month US study. American Journal of Gastroenterology 2004;99(10):S277.
8. Fehnel S, Johnston J, Kurtz C, Mangel A. Assessing global change and symptom severity in subjects with IBS: qualitative item testing. 2006 Oct.
9. Irvine EJ, Whitehead WE, Chey WD, Matsueda K, Shaw M, Talley NJ, van Zanten SJOV. Design of treatment trials for functional gastrointestinal disorders. Gastroenterology 2006;130(5):1538-51.
10. Norquist, J.M., Watson, D.J., Yu, Q., Paolini, J.F., McQuarrie, K., Santanello, N.C. 'Development and Validation of a Questionnaire to Assess Niacin-Induced Cutaneous Flushing'. CMRO. Vol. 23, No. 7:1549–1560, 2007
11. Furlong WB, Feeny D, Yandow S. (2005). Patient-focused measures of functional health status and health-related quality of life in pediatric orthopedics: A case study in measurement selection. Health and Quality of Life Outcomes (3):1477-7525.
12. Bellamy N, Buchanan WW, Goldsmith CH, Campbell J, Stitt LW. Validation-Study of Womac - A Health-Status Instrument for Measuring Clinically Important Patient Relevant Outcomes to Antirheumatic Drug-Therapy in Patients with Osteo-Arthritis of the Hip Or Knee. Journal of Rheumatology 1988;15(12):1833-40.
13. Barry MJ, Fowler FJ, Oleary MP, Bruskewitz RC, Holtgrewe HL, Mebust WK, Cockett ATK, Blaivas JG, Wein AJ. The American-Urological-Association Symptom Index for Benign Prostatic Hyperplasia. Journal of Urology 1992;148(5):1549-57.
14. Olsen EA. Female pattern hair loss: Clinical features and potential hormonal factors. Journal of the American Academy of Dermatology 2001;45(3):S69
15. U.S. Department of Health and Human Services. Food and Drug Administration Center for Drug Evaluation and Research (CDER). (Draft) Guidance for Industry Chronic Obstructive Pulmonary Disease: Developing Drugs for Treatment. Nov 2007. Available at http://www.fda.gov/downloads/Drugs/GuidanceComplianceRegulatoryInformation/Guidances/ucm071575.pdf
16. U.S. Department of Health and Human Services. Food and Drug Administration Center for Drug Evaluation and Research (CDER).(Draft) Guidance for Industry Sinusitis: Designing Clinical Development Programs of Nonantimicrobial Drugs for Treatment. Nov 2006. Available at: http://www.fda.gov/downloads/Drugs/GuidanceComplianceRegulatoryInformation/Guidances/ucm072037.pdf
17. Division of Reproductive and Urologic Products Office of New Drugs Center for Drug Evaluation and Research Food and Drug Administration. Background Document for Meeting of Advisory Committee for Reproductive Health Drugs (June 18, 2010) NDA 22-526 Flibanserin http://www.fda.gov/downloads/AdvisoryCommittees/CommitteesMeetingMaterials/Drugs/ReproductiveHealthDrugsAdvisoryCommittee/UCM215437.pdf

Discussion

- Optimal selection of a recall period for a PRO measure depends on many factors
- It is clear that
  - Consistent criteria for selecting the most appropriate recall period are needed;
  - Defining the appropriate recall period may be challenging;
  - Even within the same disease area, appropriate recall can vary depending on the concept or phenomenon of interest (e.g. variability, frequency, or overall impact);
  - The appropriate recall period must take under consideration the patient burden and the ability of the patient to easily and accurately recall the information requested;
  - The chosen recall must be consistent with the duration of the trial and the scheduled clinic visits to avoid assessments that overlap within clinical trial periods.