ISPOR TASK FORCE REPORTS

Content Validity—Establishing and Reporting the Evidence in Newly Developed Patient-Reported Outcomes (PRO) Instruments for Medical Product Evaluation: ISPOR PRO Good Research Practices Task Force Report: Part 1—Eliciting Concepts for a New PRO Instrument

Donald L. Patrick, PhD, MSPH[1,*], *Laurie B. Burke, RPh, MPH*[2], *Chad J. Gwaltney, PhD*[3], *Nancy Kline Leidy, PhD*[4], *Mona L. Martin, RN, MPA*[5], *Elizabeth Molsen, RN*[6], *Lena Ring, PhD*[7]

[1]Department of Health Services, University of Washington, Seattle, WA, USA; [2]Office of New Drugs, Center for Drug Evaluation Research, Food and Drug Administration, Silver Spring, MD, USA; [3]Department of Community Health, Brown University, Providence, RI, USA, and PRO Consulting, Pittsburgh, PA, USA; [4]United BioSource Corporation, Bethesda, MD, USA; [5]Health Research Associates, Inc., Seattle, WA, USA ; [6]International Society for Pharmacoeconomics and Outcomes Research, Lawrenceville, NJ, USA; [7]Health Economics & Outcomes Research Division, AstraZeneca, Södertälje, Sweden, and Pharmaceutical Outcomes Research, Department of Pharmacy, Uppsala University, Uppsala, Sweden

A B S T R A C T

The importance of content validity in developing patient reported outcomes (PRO) instruments is stressed by both the US Food and Drug Administration and the European Medicines Agency. Content validity is the extent to which an instrument measures the important aspects of concepts that developers or users purport it to assess. A PRO instrument measures the concepts most significant and relevant to a patient's condition and its treatment. For PRO instruments, items and domains as reflected in the scores of an instrument should be important to the target population and comprehensive with respect to patient concerns. Documentation of target population input in item generation, as well as evaluation of patient understanding through cognitive interviewing, can provide the evidence for content validity. Developing content for, and assessing respondent understanding of, newly developed PRO instruments for medical product evaluation will be discussed in this two-part ISPOR PRO Good Research Practices Task Force Report. Topics include the methods for generating items, documenting item development, coding of qualitative data from item generation, cognitive interviewing, and tracking item development through the various stages of research and preparing this tracking for submission to regulatory agencies. Part 1 covers elicitation of key concepts using qualitative focus groups and/or interviews to inform content and structure of a new PRO instrument. Part 2 covers the instrument development process, the assessment of patient understanding of the draft instrument using cognitive interviews and steps for instrument revision. The two parts are meant to be read together. They are intended to offer suggestions for good practices in planning, executing, and documenting qualitative studies that are used to support the content validity of PRO instruments to be used in medical product evaluation.

Keywords: content validity, European Medicines Agency, Food and Drug Administration, patient reported outcomes, quality of life.

Copyright © 2011, International Society for Pharmacoeconomics and Outcomes Research (ISPOR). Published by Elsevier Inc.

Background to the Task Force

In March 2009, the International Society for Pharmacoeconomics and Outcomes Research (ISPOR) Board of Directors approved the formation of the Patient Reported Outcomes (PRO) Content Validity Good Research Practices Task Force to develop a good research practices report to address methods for ensuring and documenting the content validity of newly developed PRO instruments to support medical product indications and labeling claims. This task force report extends the work of a previously published ISPOR PRO task force report on the use of existing or modified PRO instruments [1], which did not address how to establish and document content validity; that is, the specific methodologic practices involved in designing studies to gather evidence of content validity and the methods for evaluating and documenting content validity.

Researchers experienced in psychometrics and PRO instrument development working in academia, government, research organizations, and industry from North America and Europe were invited to join the task force leadership group. The task force met bimonthly to develop the topics, outline, and prepare the first draft report. Due to the large volume of information, the task force report was split into two parts. Part 1 covers elicitation of key concepts using qualitative focus groups and/or interviews to inform content and structure of a new PRO instrument. Part 2 [2] covers the instrument development process, the assessment of patient understanding of the draft instrument using cognitive interviews and steps for instrument revision.

The task force authors presented their work to date at the ISPOR 15th Annual International Meeting in May 2010 in Orlando, Florida. During July 2010 the draft report papers (Part 1 and Part 2) were sent for review to the nearly 400 ISPOR PRO Review Group members. The task force received many comments that were considered and addressed as appropriate. The task force authors presented their revised draft report for final verbal comments at the ISPOR 16th Annual International Meeting in Baltimore, Maryland, during May 2011. The revised report was sent for a final review to all ISPOR members during June 2011.

Collectively, the task force received 41 written reviews by 52 ISPOR members submitted individually or representing an organization. All written comments are published at the ISPOR Web site. A list of those members who commented is available. For these comments, please go to the 'Evaluating and Documenting Content Validity for PRO Instruments' link at the ISPOR Good Outcomes Research Practices index under the Patient Reported Outcomes heading at: http://www.ispor.org/workpaper/practices_index.asp or via the purple Research Tools menu at the top of the ISPOR homepage (www.ispor.org). All comments, many of which were substantive and constructive, were considered. Once consensus was reached by all authors, the final report was submitted to *Value in Health* in July 2011.

Introduction

According to the US Food & Drug Administration (FDA) [3], a PRO is "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." It can be measured in absolute terms (e.g., severity of a sign, symptom, or state of a disease) or as a change from a previous measure [3].

The European Medicines Agency's (EMA) Reflection Paper on the Regulatory Guidance for the Use of Health Related Quality Of Life (HRQL) Measures in the Evaluation of Medicinal Products [4] defines a PRO similarly as "any outcome directly evaluated by the patient and based on patient's perception of a disease and its treatment(s)" [4]. EMA uses PRO as an umbrella term encompassing both single and multidimension domains; that is, measures of symptoms, health status, and satisfaction with treatment [5]. HRQL is one of these multidimensional assessments under the PRO heading, broadly defined as a patient's subjective perception of the effects of the disease and treatment(s) on daily life; well-being; and psychological, physical, and social functioning. In the drug approval context, HRQL is considered a specific type of PRO [4].

Because the term PRO is often used interchangeably to refer to a PRO concept, questionnaire, instrument, score, or claim, it is useful to define these terms. PRO is the general reference to the concept (outcome) of interest. The PRO field is the general area of study. Elements of the field include PRO research; for example, burden of illness studies, qualitative theory-development studies, clinical trials, instrument development research, and PRO instrument development. Instrument development comprises the qualitative and quantitative studies that identify and measure outcomes reported by patients themselves.

A PRO instrument (i.e., a questionnaire plus the information and documentation that support its use) is a means to collect data about a PRO concept. A PRO instrument extends patient outcome assessment beyond survival, traditional clinical efficacy, and adverse effects. It assesses the concepts most relevant and important to a patient's condition and treatment. A PRO measure refers to a specific questionnaire used to collect data that produces a score representing the PRO concept of interest.

In medical product development, PRO instruments may be used in clinical trials to capture and quantify treatment benefit or risk [3,6]. This information potentially may be used to support a claim in medical product labeling or advertising. Within this context, it is useful to distinguish the PRO concept, claim, instrument, and score [6]. For example, pain intensity is a PRO (i.e., the concept), whereas a decrease in pain intensity might be a PRO claim based on a prespecified endpoint in a clinical trial. A 10-centimeter visual analog scale that assesses pain intensity—including the anchors, instructions, and recall period—is a PRO instrument. Finally, the value a subject assigns to pain intensity on the visual analog scale is a PRO score.

PRO instruments are designed to capture concepts related to the health experiences of individuals—how patients feel or function in relationship to their disease, condition, or treatment. Thus, the instruments must possess content validity. In the FDA Guidance on PRO measurement, content validity is defined by the empiric evidence that demonstrates the items and domains of an instrument are appropriate and comprehensive relative to its intended measurement concept, population, and use [3]. In practical terms, content validity is determined by documenting that the structure and content (items) capture the connection between the intended measurement concept and the way patients from the target population understand and discuss that concept. A full description of these methods and their results provides evidence that scores produced by the instrument represent the intended concept; that is, are content valid.

Qualitative data are essential for establishing the content validity of a PRO instrument. Quantitative data, including factor analysis, Rasch analyses, or item response theory analyses may be supportive, but such data are insufficient on their own to document content validity of the measure in the context of medical product development. Content validity must be based on direct input from an adequate sample of patients from the targeted clinical study population. Involving a diverse sample helps ensure that the final instrument measures the intended concept despite important variations in demographic and clinical characteristics and experiences within the target population.

Part 2 of this report describes the second phase of establishing and reporting evidence of content validity for a new PRO instrument—its development and the methods for gathering evidence that persons in the target population understand the instrument's structure and content [2].

Good Practices in Eliciting Concepts for a New Pro Instrument

Table 1 lists five steps to elicit concepts for establishing and documenting content validity of a new PRO instrument, consistent with the wheel and spokes diagram presented by the FDA [3]. These five steps represent the initial stages of instrument development. The development process in general is an iterative, rather than linear process, often requiring researchers to revisit previous steps to ensure adequate and accurate information related to instrument content or structure and to fully document content validity relative to the context of use. Each of these steps is described below.

Good practice 1: Determine the context of use

The development of an instrument, whether simple or complex, starts with the identification of the concept and the target medical product labeling claims, so those targets can be considered throughout the instrument development process. The purpose of Step 1 in Table 1 is to ensure that the context of use in medical product labeling is clearly defined, and the approach for concept measurement is appropriate for the intended context. This includes an understanding of the disease or condition in the target population, development of an endpoint model for the context of use, considerations related to specific aspects of the target population, the possible range of instrument content and structure, the theoretical and qualitative methodologic approach, and the development of a hypothesized conceptual framework. Each of these are discussed below.

Table 1 – Five steps to elicit concepts for new patient-reported outcome instruments and document content validity consistent with good research practices. *Steps to develop an instrument, evaluate the new measure through cognitive interviewing, and document that aspect of content validity are addressed in Part 2 of the Task Force report [2].

1. Determine the context of use (medical product labeling)
   - Understand the disease or condition in the target population
   - Develop an endpoint model for the context of use
   - Consider the target population – cultural/language groups
   - Consider preliminary issues related to instrument content and structure
   - Consider the theoretical and qualitative methodologic approach
   - Develop an hypothesized conceptual framework
2. Develop the research protocol for qualitative concept elicitation and analysis
   - Define the target sample characteristics
   - Select the data collection method - focus groups, individual interviews, both
   - Determine the setting and location for data collection
   - Develop the interview guide—draft, pilot, revise
3. Conduct the concept elicitation interviews and focus groups
   - Obtain institutional review board approval
   - Recruit and train sites
   - Recruit participants; monitor sample characteristics to assure diversity of participation from the target population
   - Select and train interviewers
   - Conduct interviews—implement quality control measures
   - Record or videotape interviews
   - Transcribe and clean transcripts
4. Analyze the qualitative data
   - Analyze qualitative data according to theoretical approach used
   - Establish preliminary coding framework; update as data are coded
   - Establish coding procedures and train coders
   - Organize data using a qualitative research software program
     Assess saturation
   - Interpret results
5. Document concept development and elicitation methodology and results
   - Provide target claims and any other context for use
   - Describe target population
   - Provide hypothesized and revised disease model and any input from content experts
   - Provide endpoint model
   - Provide conceptual framework and revisions made from preliminary to revised
   - Provide study methods via protocols and guides
   - Provide summary of results, including evidence of saturation
   - Provide transcripts of interviews and focus group
   - Track origin and derivation of concepts captured in the patient-reported outcome instrument
   - Summarize qualitative data
   - Provide key references

Understand the disease or condition in the target population. During discussions of the disease or condition and the possible effects of treatment, consideration is given to the underlying pathophysiology, natural history, known signs and symptoms, and their effects in the target population. The entire range of effects can be potential candidate outcomes of treatment. All types of outcomes may be considered, including biomarkers, such as hemoglobin A1c or tissue samples; clinician-reported outcomes, such as standardized psychiatric assessments, cognitive function tests, or radiographic interpretations; observer-reported outcomes, such as parent-completed diaries of observed signs in children (e.g., frequency of coughing or diarrhea) and PROs [7].

The selection of outcomes appropriate for a given trial program is often informed by consultation with clinical, trial design, and measurement experts as well as an extensive literature review. Some sponsors and developers find it useful to create a disease model to inform internal discussions about context of use in medical product evaluation. Figure 1 is an example of a hypothetical disease model for psoriasis with a proposed pathway linking proposed risk factors, diagnosis, signs, symptoms, and effect as informed by prior research literature and clinical experts.

A hypothesized disease model can help organize and visualize the key features of the patient population, drug mechanism of action, underlying pathophysiology, and effects of the treatment, including PRO and non-PRO outcomes. The model may then be used to identify and prioritize trial endpoints, with consideration given to the unique features of a particular development program. Using Figure 1, for example, information on the external factors that can influence the manifestations of psoriasis, together with the clinician-observed signs and patient-reported signs and symptoms of this condition can inform the evaluation of the suitability of existing instruments for the development program and/or the development of a research protocol for eliciting concepts to be included in a new instrument.

Note that Figure 1 does not explicitly depict the frequency of assessment or the optimal outcome assessment window for the proposed clinical trials, although these are important considerations during the PRO development process. Potential patterns of change over time in the PRO concept(s) of interest and the type and expected effect of treatment are considered during the design of the clinical trials and qualitative elicitation research. In the example of psoriasis, a PRO measure for trials testing a slow onset treatment for a mild condition may take the form of a weekly questionnaire with a retrospective recall, whereas assessments for a trial evaluating a rapidly acting treatment for severe psoriasis may require short, early, and frequent assessments, such as a daily diary using a personal digital assistant. Trials demonstrating long-term benefit may involve a different type and frequency of PRO assessment than those with short-term or acute effects. Trial design factors, including frequency of assessments, potential issues around adherence, and the implications of missing data are important to consider when thinking about a new PRO instrument and planning the underlying qualitative work.

Develop an endpoint model for the context of use. Endpoint models specify the primary and secondary endpoints to be tested in the clinical trial(s) to support the targeted claims. Example endpoint models are provided in the FDA guidance document [3]. Even when a medical product cannot be fully specified (e.g., in multi-sponsor instrument development consortia), the anticipated role of the instrument can be shown in one or more hypothetical or illustrative endpoint models to specify the context of use, including targeted claims. In this case, the model(s) represent an educated prediction of the prioritization of study hypotheses in clinical trials in which the PRO instrument is to be used. A new PRO instrument may serve as an exploratory endpoint in early trials, with the data used to test reliability, validity, ability to detect change, and potential approaches to interpretation of observed differences between the treatment and comparison groups.

Consider the target population—cultural/language groups. As instrument development is planned, thought should be given to the characteristics of the target population, including the languages and cultures of patients likely to be enrolled in the clinical trial(s). The extent to which the disease, standard of treatment, and measurement concept(s) are the same or different across countries or cultures needs to be considered as early as possible and incorporated into the qualitative research effort. Literature as well as clinical and/or measurement experts can help in this discussion. If the development program is international and the concept is highly variable across countries with cultural and linguistic differences,
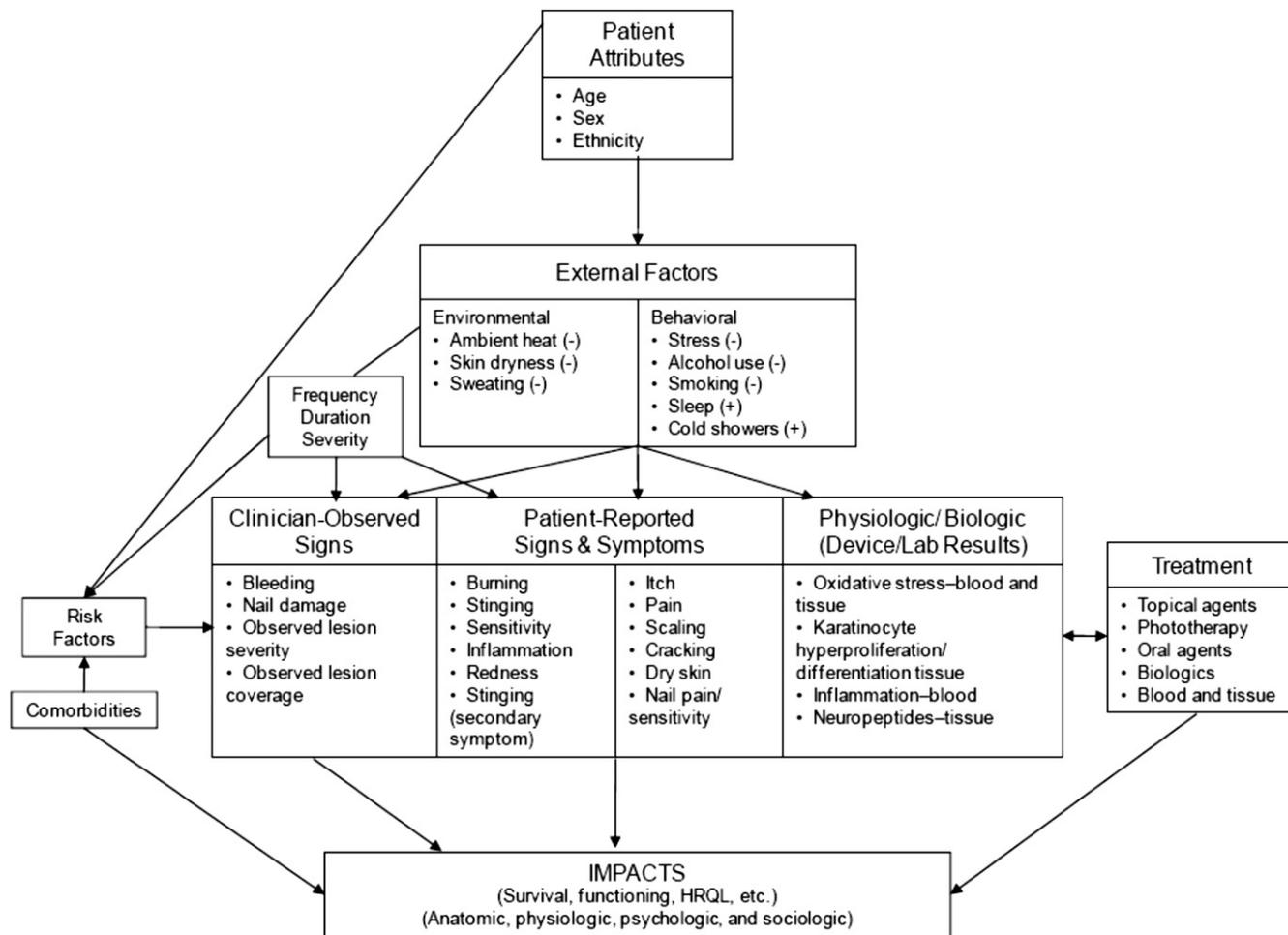
Fig. 1 – Hypothesized disease model for psoriasis for use in discussing context of measurement for the disease, target population, and treatment.

simultaneously developing an instrument internationally in two or more languages may strengthen the cultural and linguistic equivalence of the final instrument.

Consider preliminary issues related to instrument content and structure. As the context of use is identified and clarified, issues related to content dimensions and form may arise. Characteristics of the target population, for example, may indicate the extent to which respondents will be able to identify and describe symptoms that are specific to their condition. In Figure 1, if the signs and symptoms are clearly attributable to psoriasis alone and no other condition, it may be possible for patients to identify and rate symptoms of psoriasis. Patients with multiple chronic conditions may not be able to attribute their symptoms or health experience to one condition. For example, it would be difficult for patients to know if their breathlessness was the result of congestive heart failure as opposed to other causes, such as aging, anxiety, or infection.

If possible, patients should not be asked to attribute their symptoms to a cause. Patients also should not be asked to rate changes in their condition over time, particularly with long recall periods. Optimally, assessments taken at baseline and endpoint are used to compute change over the study period, rather than relying on patient recall and ratings of change. Information on patient perception of the condition, range of experiences, and their ability to attribute and recall are addressed during the qualitative research.

Another factor to consider early in the planning process is possible mode of data collection; that is, self vs. interviewer adminis-

tration or pen-and-paper, electronic, or interactive voice response, among others. Although the target population and target concept(s) should be the primary factors influencing mode of administration, trial constraints may also be taken into consideration, including global study sites and resources. Caution should be taken if switching methods or modes of administration between development and use of the instrument unless the methods or modes are adequately developed to assure score equivalence [8].

It is noteworthy that many of these issues will also be addressed during the design of the qualitative research protocol. Early and repeated consideration of these context-of-use issues contributes to a well-designed and executed PRO development process.

Consider the theoretical and qualitative methodologic approach. To ensure and document content validity of a new instrument, consultation with patients based on protocol-driven qualitative research methods is necessary. Detailed examination of the theories and overarching methodologic approaches that may inform qualitative methods in PRO development is outside the scope of this article. Nevertheless, it is important to note that the methods used in PRO development did not originate in this field alone. Many methods have come from multiple theoretical approaches developed in other contexts, such as phenomenology, grounded theory, content analysis, and thematic analysis [9–14]. All of these approaches are idiographic (i.e., focus on the individual) and based on an interpretive/constructivist paradigm. In contrast, the nomo-
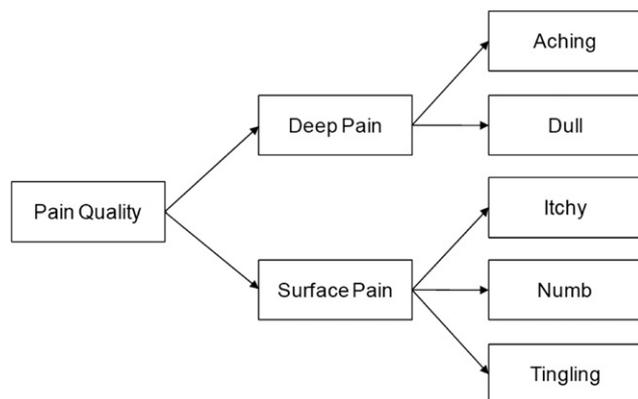
Fig. 2 – Example conceptual framework for a patient-reported outcome evaluating the concept of pain quality.

thetic paradigm (i.e., focus on the general) is founded in positivism [15,16] and is generally equated with quantitative research.

Some authors have recommended phenomenology as an overarching theoretical framework and grounded theory as a specific methodology as appropriate for the development of a new PRO instrument [11]. Grounded theory methods are characterized by inductive (i.e., from the particular to the general) rather than deductive (i.e., from the general to the particular) reasoning. In its pure application, no prior knowledge would be used to understand patient experience. An adaptation of grounded theory has also been proposed to incorporate prior clinical knowledge based on expert opinion and the scientific literature [17]. For the development of a PRO instrument with a specific intended use, existing information is used to focus the study design, inform the interview guide, identify themes and concepts in the data, and interpret the results in light of the ultimate goal.

The challenge is to develop methods that permit moving back and forth between hypothetic-deductive and inductive approaches. In this way, instrument development aims and prior knowledge are not ignored. Developers' understanding will change based on new information and/or observations. When developing new PRO instruments, the purpose across all qualitative methods is to understand patients' perspectives and experiences—including rare perspectives or uncommon experiences—because understanding the entire range of patient experience and perspective is crucial for developing sensitive and comprehensive PRO instruments.

Develop an hypothesized conceptual framework. The information described previously is used to inform the development of a hypothesized conceptual framework for the proposed instrument; that is, a diagram depicting relationships between the overarching concept, hypothesized domains or subconcepts, and candidate item content. This framework provides the initial structure for the measure based on existing information to inform the qualitative research protocol and interview guide.

An example of a hypothesized conceptual framework is shown in Figure 2. The concept of interest here is pain quality. In this hypothetical example, pain quality is divided into deep pain and surface pain: concepts perhaps suggested in published literature and prior qualitative work. These two concepts are further divided into aspects of pain quality. Some hypothesized frameworks for a specific condition consist of a single concept, such as pain on its own, and are not detailed prior to patient interviews. The essential elements of a hypothesized conceptual framework include specification of the potential concept (pain quality) and any subdomains (surface and deep) that may have been repeatedly established in the literature.

The hypothesized conceptual framework is adjusted throughout the collection of qualitative data. Qualitative data are necessary to find out how patients in the specified target population perceive their condition. In the case of Figure 2, the data will provide information on how patients characterize the quality of their pain, if indeed they make a hypothetical differentiation, and the words or concepts they use to describe their pain. Qualitative work may not substantiate the hypothesized framework—it may lead to revisions, including additional domains that are relevant and important and/or domains and items expressed in entirely different terms than those in the early hypothesized framework.

Concepts that are broad and perceived differently by people with different conditions, such as fatigue, may entail many domains different from those hypothesized. Moreover, the items need to reflect the language of the qualitative research participants. Similar to a disease model, the hypothesized framework is merely an organizing tool that summarizes what has been found in the literature and discussions with experts. Qualitative research is best conducted with openness to discarding or significantly revising the conceptual framework at all stages of the research, including going back to qualitative work when quantitative work shows that items do not tap the full continuum of the response variable; for example, severity.

Good practice 2: Develop the research protocol for qualitative concept elicitation and analysis

A detailed research protocol guides all aspects of the qualitative studies used to investigate, evaluate, and document content validity of the new PRO under development. The protocol includes a prespecified plan for identifying the study sample with inclusion and exclusion criteria, conducting interviews and/or focus groups, analyzing data, and documenting evidence that will inform the content and structure of the new instrument.

Study sample. Demographic and clinical characteristics (e.g., levels of severity, disease experience, and frequency or duration of symptoms) of the sample should closely mirror a wide range of members from the target population who will be enrolled in the planned clinical trials. When evaluating clinical sites and/or locations for possible participation, consideration should be given to geographic, educational, ethnic, and racial diversity, as well as to the availability of clinical information needed to characterize and evaluate sample characteristics in the final report, such as disease severity or stage, comorbidities, and time since onset of the condition. Sufficient time is required to identify, recruit, and interview participants and to analyze data from patients in a diverse sample that represents the clinical trial target population.

Newly diagnosed patients may have a different awareness of symptoms and thus provide different information from those who have had the opportunity to accommodate to their condition [18]. Patients with a long history with a condition may demonstrate response shift [18] that will be important to acknowledge if the clinical trial population will include those individuals. Patients who have participated in multiple clinical trials may or may not provide different information than those who have not. Researchers should be aware of these patient characteristics and seek a wide diversity of patient experience.

Estimating the sample size for a qualitative study can be challenging. In quantitative research protocols, sample size is estimated using analytical techniques requiring effect size projections, desired power, and a statistical significance criterion. In qualitative research, sample size estimation is based on projections of the number of participants needed to reach saturation of the concept. Saturation is the point in the data collection process after which no relevant information is elicited [1]. The notion of saturation was introduced as theoretical saturation in grounded theory [13]. The iterative process of sampling, data collection, and analysis is a defining feature of the grounded theory approach, which allows theory to be built, defined, and tested. Previously published articles on content validity and PRO qualitative research

Table 2 – Focus groups and interviews: advantages and disadvantages.		
	Focus groups	Interviews
Advantages	• Rich source of data • Allows to use ideas of others as cues to express their own views • Participants can compare their experiences with others • Able to reach many participants at once	• Get more in-depth and detailed information about an individual's experience • Can be useful for sensitive topics • Data can be easier to analyze • Scheduling can be easier
Disadvantages	• Data can be tough to analyze because talking can be in reaction to the comments of other group members • Moderators need to be highly trained and able to lead the group • One strong group member can sway tone of entire group	• It may take longer to collect the data • Limited to one participant's view at a time; no peer comparison • Interviewers need to be trained with excellent one-on-one communication skills • May be more costly (e.g., travel, room rental, transcription fees)

recommend the practice of gathering qualitative data to the point of saturation to ensure that the items in an instrument appropriately represent the relevant "universe of content" for the concept [1,11,17,19–21].

No rule can be provided to determine either the sample size or number of iterations needed to reach saturation in PRO instrument development. The sample needed to achieve saturation depends on the concept of interest and how it is perceived by patients from the target population. Heterogeneous patient samples and complex concepts generally require larger samples sizes. As noted in the FDA guidance document [3], "The number of patients is not as critical as interview quality and patient diversity included in the sample in relation to intended clinical trial population characteristics." Further discussion of saturation is included in the section on analyzing qualitative data.

Data collection method. Individual interviews and focus groups are the qualitative research data collection methods used in concept elicitation for instrument development purposes [11]. A brief summary of advantages and disadvantages of these methods illustrating the importance of this choice is shown in Table 2. Focus groups are economical and can stimulate discussion of topics and comparison of experiences across participants that cannot be captured in individual interviews [22–25]. Unfortunately, there are also risks associated with focus groups, particularly when run by inexperienced or untrained leaders. One example is a highly vocal, assertive participant who dominates or leads the discussion, minimizing participation of other group members. When such a participant dominates, content, tone, and/or perspectives in the data do not necessarily represent those of other individuals or the group as a whole.

Individual interviews are ideal for concepts that are sensitive or target populations/people unlikely to volunteer or share information in a group setting, and some aspects of transcript analysis can be less complex. Individual interviews, however, also have disadvantages. For example, interviews must be conducted sequentially or by multiple interviewers, both of which are more expensive and time-consuming [17]. Focus groups may augment individual interviews by sparking conversation and ideas arising from social interaction and seeking consensus on issues of importance.

Setting. Focus groups and interviews may be conducted in outpatient clinics, inpatient settings, dedicated research facilities, or participants' homes. In some situations, interviews may be conducted over the telephone, such as in rare, episodic, or contagious conditions or when illness severity, physical mobility, psychological state, or geographic location may inhibit participants' ability to travel. The setting should optimize the consistency of the sample with the target population by making participation as broadly accessible as possible.

Interview guide. The interview or focus group guide includes the procedures to be used and the questions to be asked to meet the goals of the study with optimal clarity and data quality. The development of a clear open-ended interview guide is essential to avoid undue influence of researcher, interviewer, or interview

guide bias [26]. An interview guide is not read verbatim during the focus group or interview. The actual situation often dictates how the dialogue unfolds.

The interview guide helps interviewers cover a range of topics and add topics mentioned by respondents. For example, if pain is identified by respondents as an important symptom, the interview/focus group guide might provide sample probes for exploring patient experiences of pain in greater depth. Questions about frequency, severity, duration, quality, or other attributes can increase understanding of a patient's experience. Whenever possible, participants should be allowed to speak spontaneously before probes are offered.

With this in mind, open-ended questions that are too broad can be confusing to participants. "Tell me about your condition," for example, lacks the specificity required for participants to address the concept of interest and can lead to wide-ranging information too broad to inform instrument development. Open-ended questions include parameters consistent with the concept of interest. If the concept of interest is knee pain, the interviewer could ask: "How did your knee feel in the last 24 hours? At the start of the day? When you went to bed?" Well-designed probes may be included to better understand the nature and characteristics of the experience offered by the participant. This approach provides spontaneous data on the words and phrases participants use to describe their condition that will inform instrument content.

Unless carefully worded and conducted, interview questions and procedures as well as interviewer body language and expression can also introduce bias into the data. For example, certain closed-ended or highly specific questions can be leading, such as "you experienced pain in your knee today?" or "how depressed were you during this event?" Questions should be open-ended whenever possible and worded to encourage spontaneous information from the participant without pointing them toward a specific response.

The reference timeframe—that is, the timeframe participants are asked to consider as they respond to the questions—will also depend on the PRO and measurement context. For example, when developing a measure for chronic heart failure patients, participants may be asked to recall and describe their experiences during the current day or past week. If acute heart failure is the target, patients may be asked to recall a recent acute episode or hospitalization. In general, it is desirable for the reference timeframe to be as close as possible to the interview or focus group session to diminish recall errors and bias. One method known as the day-reconstruction approach [27] can be used to direct a patient to a specific day as they describe symptoms, effects, or other experiences relevant to the target concept.

After a draft interview guide has been created, it should be reviewed by other qualitative researchers for flow, redundancy, poorly formulated questions, and the use of terminology and probes. The draft guide can be pretested with study naïve individuals or colleagues. Ideally, the guide should be pilot tested in the target population before primary data collection begins to identify areas that do not flow easily or may confuse respondents. Revi-

Table 3 – Sample form for evaluating core competencies in concept elicitation interviewing.				
Focus of evaluation	Criteria met?		Issues found in evaluation	Remedies needed
	Yes	No		
PREPARATION to start interview w/subject				
Familiar with interview content & recording				
Demonstrates understanding of participant responses				
PROTOCOLS followed as identified in interview guide				
Identified primary purpose of Interview/Focus Group to participant(s)				
Adhered to interview guide methods and covered all probed content				
Allowed time for participant to spontaneously respond to questions before offering examples or probing				
Encouraged additional comments at end				
COMPETENCIES demonstrated in conduct of interview				
Reframed question to assist understanding				
Allowed participant time to respond				
Did not interrupt or rush participant				
Appropriately explored responses, and promoted in depth discussions				
Offered multiple examples, not just one				
Kept participant on topic				
Stayed neutral				
Avoided confirming responses				
Avoiding leading questions				
Used participant language, as possible				
Recognized information already provided				
GENERAL COMMENTS				
Overall competency based on the criteria:				
Interview strengths:				
Interview skills in need of improvement:				

Note: Printed with permission from M. L. Martin, Health Research Associates.

sions are often made after the first few interviews because the researcher cannot fully anticipate how the targeted patients will interact and respond to the questions and probes.

Good practice 3: Conduct the concept elicitation interviews and focus groups

The approval of the research protocol by an appropriate ethical review board enables the start of patient recruitment and data collection. Participating sites or recruiters are provided with a copy of the study protocol and trained on inclusion/exclusion criteria, sample monitoring, recruitment processes, and informed consent procedures. Good interviewers and focus group facilitators are experienced in qualitative research methods and trained on the background and objectives of the protocol. Mock interviews or focus groups may be used to help the interviewers/facilitators develop a complete understanding of the questions and process. Sustained interaction with interviewers in the field is important to establish and maintain data quality.

A listing of the types of competencies optimal for focus group leaders and interviewers are shown in the sample evaluation form in Table 3. The concept elicitation process is intentionally broad to explore and define information from the perspective of the patient. A well-constructed interview guide defines the broad territory of discussion, leaving no need for the interviewer to censure or discount patient responses. Although discipline is needed to keep a participant or focus group on task, interviewers must avoid being overzealous in assuming irrelevance and favor an open dialogue among patients to encourage participation.

Body language and actions, such as nodding in agreement, frowning, or sighing, can communicate approval or disapproval of the pa-

tient's contribution, altering subsequent information. Interviewers should convey genuine interest to encourage open and honest communication. Hallmarks of interviewer skill rest on the ability to get the patient to talk about the areas and topics of interest in a natural conversational engagement, in which participants believe that they are being heard and respected. The skill set for successfully moderating focus group is different from that required for successfully conducting individual interviews. Regular monitoring and periodic training are good practices for all interviewers.

Concept elicitation interviews and focus groups are recorded by high quality audio or video equipment to fully capture the context and content of each session as well as produce transcripts that form the data for analysis. Audio recordings are easier to implement and transcribe, facilitate participant anonymity, and are generally more comfortable for participants, particularly when sensitive topics are being discussed. Video recordings of focus groups provide data going beyond the transcript, including group interaction. Regardless of recording method, participants are assured of confidentiality and limited use of the recorded materials from their interviews. Recordings should be monitored for quality by a senior interviewer who provides feedback to the lead interviewer to maintain or improve the quality of data collection throughout the duration of the study by improving question clarity, altering probes, and/or pursuing specific aspects in greater detail.

Recording frees the interviewer or moderator from note taking so that he or she may engage fully with participant(s). For focus groups, if video is not used, an assistant moderator is useful to observe the group and take notes to facilitate data interpretation. These notes include a seating chart and the key points contributed
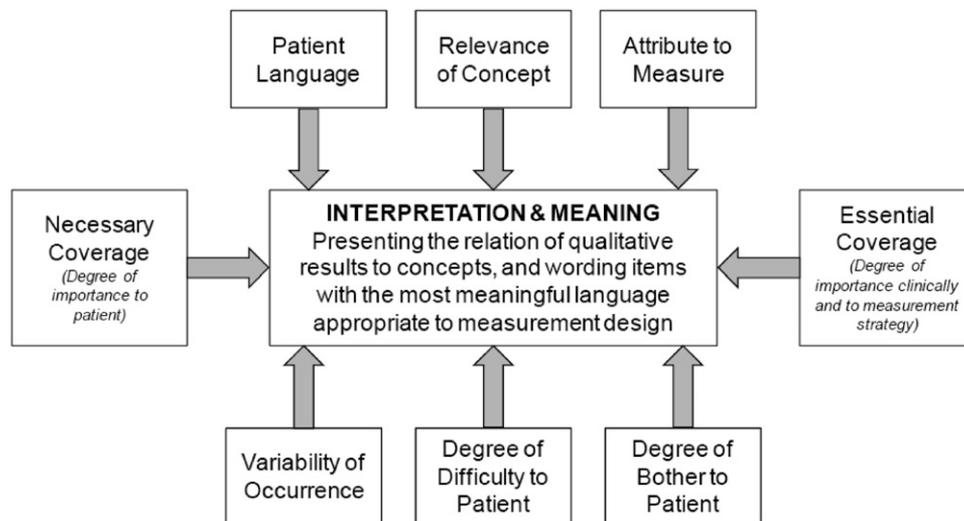
Fig. 3 – Example approach to analyzing qualitative data within instrument development process.

by members of the group. This also helps in checking the accuracy of transcriptions of focus group recordings.

Transcriptions of the audio/video recordings need to be 1) verbatim and reviewed as they are collected, not at the end of the study retrospectively; 2) quality checked to provide further training to interviewers or to make needed revisions in the process; and 3) cleaned by the facilitators/interviewers and associates. Cleaning involves comparing transcripts to the audio/video recordings. It includes removal of any personal identifiers and correction of any medical terms or words that the transcribers did not recognize or misspelled. Any dialogue that is related, but not central to the purpose of the interview, is retained in the transcript because such dialogue may provide important context. Transcript quality is assessed through the direct comparison of voice and transcript files, generally performed on a random sample of transcriptions. After transcripts have been quality checked and cleaned, qualitative analyses begin.

Good practice 4: Analyze the qualitative data

The analysis of qualitative data is not a quantitative process; there are no significance levels, effect sizes, or other quantitative metric. Qualitative analyses use respondent words and phrases as data, analyzing and classifying these data by concept and subconcept. Rather than a statistical analysis plan, the qualitative analysis process for instrument development is informed by context of use, measurement goal, the study protocol, and the qualitative interview guide.

Analyses of data from patient interviews and focus groups include data provided by study participants in response to direct queries, as well as responses arising spontaneously or as a result of specific probing. Respondent descriptions are considered carefully in relation to the attributes or characteristic features of the target concept, indicators of magnitude and variability, and information related to impact or bother and relevance or importance. The multiple aspects of data analyses are illustrated in Figure 3.

Patient quotes from all interviews and focus groups provide a picture of patients' experiences with the target concept. For instrument development, the goal is to understand, organize, and communicate the meaning of the data and translate that meaning into a set of items that can be scored to represent the targeted concept(s) quantitatively. The patient quotes show the relationship between the concepts, the words and phrases, and the final PRO instrument.

Ideally, analyses begin before the interviewing is completed. Existing guidelines for reporting qualitative research can aid in structuring the description, evaluating the process used, and determining how best

to present and discuss results [28–33]. Guidance on this process appears in the qualitative methodology literature [34–37].

Coding qualitative data for instrument development. The primary goal of transcript coding is to organize and catalog a patient's descriptions of their experiences within the context of use. Different qualitative methods for conducting the coding process share an emphasis on using and including the context of patients' discussions. Based on a phenomenologic approach, one can identify descriptions of the experience that are universal (phenomenology). Based on grounded theory, one can use three types of coding: open coding (examining, comparing, conceptualizing, and categorizing data); axial coding (reassembling data into groupings based on relationships and patterns within and among the categories identified in the data); and selective coding (identifying and describing the central phenomenon, or core category) [38–40].

Inductive coding [41] helps ensure that the ideas generated before the patient interviews are not superimposed on or adversely bias the data; that is, that concepts are generated from both the early work and patient data. Moreover, an inductive approach to developing concept codes ensures that the ideas generated from patients during the interview process have appropriate influence on the variety and labeling of the codes assigned and the overall organization of the qualitative results.

Coding is an iterative process with opportunities for data to be re-examined and reanalyzed until no new codes or code groupings are identified and all passages from the transcripts have been assigned one or more codes. The naming of codes and groups of codes is an important process to be conducted deliberately to ensure the concepts expressed by patients are accurately described. Naming is best done using the language of the patients as nearly as possible, because the names of the concepts represent the perspectives of patients and not the developers. Data should be transcribed and coded on a rolling basis with regular intervals of assessment to evaluate the consistency of the code assignment process and adequacy of the coding framework and to monitor the appearance and organization of newly appearing concept codes.

Figure 4 shows one possible approach for progressing from a coding framework to a completed coding dictionary. A coding dictionary is a document that contains all codes assigned to the transcripts with definitions as needed for standardization, clarity, and communication. The coding framework; that is, initial codes and starting structure for organizing these codes, is based on all sources of information up to this point, including early patient
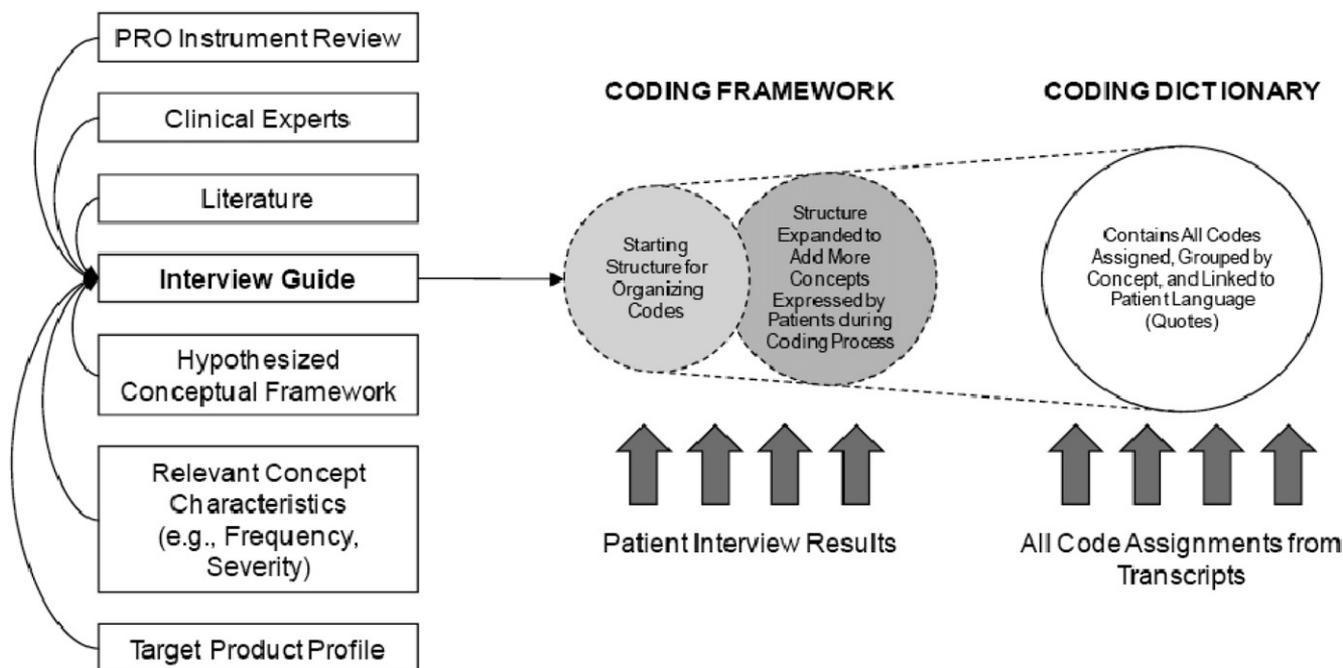
Fig. 4 – From coding framework to coding dictionary.

interviews. During the coding process, this framework is a living document that progresses to the final coding dictionary. That is, as the analyses progress and insight is gained, codes may be added and/or reorganized and definitions are added or clarified. Changes are discussed and agreed upon among the analysts to ensure coding consistency. As the coding dictionary takes shape, transcripts may be reanalyzed or double-coded by a second analyst to ensure consistency with the coding dictionary and across data analysts or coders. The final coding dictionary reflects patients' perspectives of the concepts, with the presentation of results showing how the codes and actual patient quotes interrelate.

Presentation of the coded qualitative data demonstrates and documents the language patients used to describe or characterize each coded concept, including the most frequently used words or phrases and the range of terminology. The presentation of codes and themes may differ depending on the qualitative approach, For example, a thematic map may be used showing an overall conceptualization of the data patterns and the relationships between them [12]. The data presentation may include a count of patients expressing the concepts both spontaneously and after probing; that is, thematic prevalence.

Computer-assisted qualitative data analysis software programs, such as Atlas.ti [42], can be used to organize interviewer notes, transcribed data, and the coding scheme for easier retrieval, analyses, and assessment of inter-coder reliability. These programs do not assign codes to the data; skilled decision-making is still needed to allocate patient expression of concepts to the appropriate code. Computer-assisted qualitative data analysis software programs aid in organizing data to assist in the determination and documentation of saturation and to facilitate quality assurance audits of the analysis process.

Assessing and documenting concept saturation. A good practice is to initiate assessments of saturation early in the concept elicitation process. It is recommended that saturation be assessed at multiple points during the data collection process. It is also recommended that a procedure for documentation of saturation be specified in the qualitative study protocol before conducting the study [20].

Careful monitoring during the coding process and a phased approach to assessing saturation provide researchers with insight into the data as the study progresses and an opportunity to return to the field if needed for further comprehensiveness or clarity. To assess saturation, transcripts and coding can be evaluated after a set of five to eight interview or focus group transcripts become available.

A coding dictionary can be used to guide the analyses and document the methods and results, including saturation [20,43]. This approach is also useful for replicating studies, providing a standardized, systematic approach for analyzing the data and documenting saturation; that is, when there are no new changes to the codebook/dictionary.

A common approach to analyzing data for saturation is to identify codes in each set of transcripts and compare these with the codes that appeared in previous groups. A saturation table organized by concept code can be used to document the elicitation of information by successive focus group or interviews. Table 4 shows an example saturation table used to track the appearance of new concepts. Alternatively, saturation tables can be completed that note all occurrences of the concept across the transcript groups; that is, thematic prevalence. Data are examined for either the continued identification of new concepts (newly appearing codes) or codes requiring further examination to confirm the relevance and accuracy of concept coding or the attainment of saturation.

In the best case scenario, saturation is confirmed when no new concepts whatsoever arise after the first set of interviews or focus groups. In reality, it is not uncommon for a new concept to arise late in the data collection process. Scientific judgment, including knowledge of the field and consultation with experts, is used to determine if this new concept is minor, technically unrelated, or an outlier—reflecting a relevant but unusual case—and further judgment is required to determine if additional data collection is warranted to reassess saturation following this late revelation.

Multiple coders. Although all codes represent clusters of information, ways of identifying and labeling clusters may differ across individuals. Good practice in analyses of qualitative data involves two or more coders thoroughly trained in the purpose of the study, target concept, nature of data itself, the coding framework, and the coding dictionary. Each coder completes one or two transcripts and meets with fellow coders to compare codes assigned, identify areas of consistency and inconsistency,

Concept codes	Transcript group where concept first appeared			
	Transcript Group 1 (n = 5 transcripts)	Transcript Group 2 (n = 4 transcripts)	Transcript Group 3 (n = 5 transcripts)	Transcript Group 4 (n = 5 transcripts)
Shortness of breath		X		
Difficult to breathe	X			
Not enough air	X			
Gasping			X	
Shallow/quick breathing	X			
Wheezing	X			
Coughing	X			
Chest tightness	X			
Chest pain		X		
Dizziness/lightheadedness	X			
No. of new concept codes appearing in each transcript group	7	2	1	0
% Of total new concept codes (total = 10)	70	20	10	0

Table 4 – Example symptom saturation table for concept elicitation interviews..

reconcile the codes used in transcripts, and revise the coding framework and dictionary for clarity and consistency in subsequent transcript coding. This process is repeated several times throughout data analyses.

Agreement between coders is defined as a set of words or phrases identified as reflecting the same code and/or subcode. Given the nature of qualitative data, flexibility is permitted around the words that represent larger wordsets or phrases. For example: two coders assigning codes "pain" and "pain with kneeling" to the transcript text, "You know I am always in pain when I kneel" would be considered in agreement, even though one is more specific than the other. A third person may be useful in reconciliation of coding differences.

Ensuring and documenting coding precision can take several forms. One approach is to have a supervising coder review all coded data to ensure consistency across coders. A second approach is to draw a random selection of transcripts that are coded by two coders and assessed for inter-rater agreement. These methods are similar to interviewer-coded audio recordings using psychiatric ratings scales in which inter-rater agreement is assessed until it reaches 90% or higher [44]. Typically, through ongoing discussion of coding and reconciliation, more than 90% agreement can be reached. Regardless of the approach used, the coding method and procedures for quality assurance should be carefully documented.

Good practice 5: Document concept development and elicitation methodology and results

The FDA PRO guidance document [3] outlines the information that sponsors should provide to document the content validity of a new PRO instrument. Documentation begins for both FDA and EMA reviews with the PRO instrument itself, followed by a description of the steps used to identify concepts and create the instrument. The FDA proposes an order and taxonomy designed to provides a logical flow for organizing the report [3].

Concept elicitation methods are the first part of the evidence supporting content validity. Essential documentation of content validity includes the concept elicitation methods discussed in this article and cognitive interviewing methods described in Part 2. This qualitative evidence of content validity may be accompanied by supplementary quantitative evidence that confirms or revises the proposed conceptual framework.

Essentially, the early content validity documentation provides evidence that the proposed instrument meets the definition of content validity; that is, the score represents the concept, and the instructions and item content are appropriate, comprehensive, and understandable to the targeted patient population. Further-

more, this evidence is specific to the planned clinical trial population and to the indication; that is, the context of use.

Table 5 shows the elements of documenting the concept elicitation phase of instrument development. Discussion of Item generation, cognitive interviewing, and documentation of these development and evaluation procedures, including the item tracking matrix, are contained in Part 2 of this report.

Conclusions

This article, the first of a two-part task force report, outlines steps for conducting qualitative research to inform the devel-

Table 5 – Documentation of concept elicitation. This Figure shows only documentation of concept elicitation. For further documentation, see Part II of this report [2].

Information for documentation	Check when completed
Target claims (i.e., from Target Product Profile)	
Description of the target population	
The hypothesized and any revisions to disease model	
The endpoint model	
Literature review and documentation of clinical and measurement expert input	
Preliminary and revised conceptual framework for the patient-reported outcome instrument based on qualitative studies conducted prior to testing of measurement properties	
Qualitative study methods and results, including protocols, interview guides, and summary of results	
Transcripts of interviews and focus groups	
Origin and derivation of concepts captured in the patient-reported outcome instrument	
Summary of qualitative data supporting the concepts, items, response options, and recall period	
Key references	

Note: The reader is cautioned not to confuse the hypothesized disease model with either the clinical trial endpoint model or the hypothesized conceptual framework of patient-reported outcome concepts that arise from qualitative work with patients.

opment of a new PRO instrument to be used in medical product development trials evaluating the benefits and risks of treatment. Context of use, qualitative study design, interview guide development, data collection methods, and procedures for data analyses each contribute to the identification, organization, and documentation of concepts that will form the basis of the new measure. Examples have been provided to clarify specific steps and inform the development of documentation of content validity. Part 2 of this report covers the drafting of the new PRO instrument. This includes evaluating and revising the instrument through cognitive interviewing, as well as documenting the methods and results to demonstrate that the instructions and item content are appropriate, comprehensive, and understandable to the target population.

Acknowledgements

REFERENCES

[1] Rothman M, Burke L, Erickson P, et al. Use of existing patient-reported outcome (PRO) instruments and their modification: ISPOR good research practices for evaluating and documenting content validity for the use of existing instruments and their modification PRO task force report. Value Health 2009;12:1075–83.
[2] Patrick DL, Burke L, Gwaltney CJ, et al. Content validity—establishing and reporting the evidence in newly developed patient-reported outcomes (PRO) instruments for medical product evaluation: ISPOR PRO good research practices task force report: Part 2—assessing respondent understanding. Value Health 2011;14:978–88.
[3] US Food and Drug Administration Guidance for Industry: Patient-Reported Outcome Measures: Use in Medical Product Development to Support Labeling Claims. Rockville, MD: Department of Health and Human Services, Food and Drug Administration, Center for Drug Evaluation and Research, 2009.
[4] European Medicines Agency. Reflection Paper on the Regulatory Guidance for the Use of Health Related Quality of Life (HRQL) Measures in the Evaluation of Medicinal Products. London: European Medicines Agency, 2005.
[5] Acquadro C, Berzon R, Dubois D, et al. Incorporating the patient's perspective into drug development and communication: an ad hoc task force report of the patient-reported outcomes (PRO) harmonization group meeting at the food and drug administration, February 15, 2001. Value in Health 2003;6:522–31.
[6] Patrick DL, Burke LB, Powers JH, et al. Patient-reported outcomes to support medical product labeling claims: FDA perspective. Value Health 2007;10(Suppl. 2):S125–37.
[7] Burke LB, Walton M. Outcomes targeted for labeling: what works and what doesn't? Available from: http://www.c-path.org/PROSlides/Workshop2/PROsTargetedforLabeling.pdf. [Accessed June 9, 2011].
[8] Coons SJ, Gwaltney CJ, Hays RD, et al. Recommendations on evidence needed to support measurement equivalence between electronic and paper-based patient-reported outcome (PRO) measures: ISPOR ePRO Good Research Practices Task Force report. Value Health 2009;12:419–29.
[9] Denzin N, Lincoln YS (Eds.). The SAGE Handbook of Qualitative Research. 3rd ed. Thousand Oaks, CA: Sage Publications, 2005.
[10] Starks H, Trinidad, SB. Choose your method: a comparison of phenomenology, discourse analysis, and grounded theory. Qual Health Res 2007;17:1372–80.
[11] Lasch KE, Marquis P, Vigneux M, et al. PRO development: rigorous qualitative research as the crucial foundation. Qual Life Res 2010;19:1087–96.
[12] Braun V, Clark V. Using thematic analysis in psychology. Qual Res Psychol 2006;3:77–101.
[13] Glaser BG, Strauss AL. The Discovery of Grounded Theory: Strategies for Qualitative Research. Chicago: Aldine Publishing Co., 1967.
[14] Strauss AL, Corbin J. Basics of Qualitative Research: Techniques and Procedures for Developing Grounded Theory. 2nd ed. Thousand Oaks, CA: Sage Publications, 1998.
[15] Ponterotto JG. Qualitative research in counseling psychology: a primer on research paradigms and philosophy of science. J Counseling Psychology 2005;52:126–36.
[16] Robson C. Real World Research: A Resource for Social Scientists and Practitioner-Researchers. 2nd ed. Oxford, UK: Blackwell Publishers, 2002.
[17] Brod M, Tesler LE, Christensen TL. Qualitative research and content validity: developing best practices based on science and experience. Qual Life Res 2009;18:1263–78.
[18] Sprangers MA, Schwartz CE. Integrating response shift into health-related quality of life research: a theoretical model. Soc Sci Med 1999;48:1507–15.
[19] Burke LB, Kennedy DL, Miskala PH, et al. The use of patient-reported outcome measures in the evaluation of medical products for regulatory approval. Clin Pharmacol Ther 2008;84:281–83.
[20] Kerr C, Nixon A, Wild D. Assessing and demonstrating data saturation in qualitative inquiry supporting patient-reported outcomes research. Expert Rev Pharmacoecon Outcomes Res 2010;10:269–81.
[21] Morse JM. Exploring the theoretical basis of nursing using advanced techniques of concept analysis. Adv Nurs Sci 1995;17:31–46.
[22] Lehoux P, Poland B, Daudelin G. Focus group research and "the patient's view." Soc Sci Med 2006;63:2091–104.
[23] Kitzinger J. Qualitative research: Introducing focus groups. Brit Med J 1995;311:299–302.
[24] Hollander J. The social contexts of focus groups. J Contemp Ethnogr 2004;33:602–37.
[25] Smithson J. Using and analysing focus groups: limitations and possibilities. Int J Soc Res Methodol. 2000;3:103–19.
[26] Lasch KE, Jernigan KA, Scott JA, et al. Interviews with cancer fatigue patients. J Clin Oncol 2009;27:e17577.
[27] Kahneman D, Krueger AB, Schkade DA, et al. A survey method for characterizing daily life experience: the day reconstruction method. Science 2004;306:1776–80.
[28] Cochrane Qualitative Research Methods Group. Cochrane Handbook for Systematic Reviews of Interventions. Oxford, UK: The Cochrane Collaboration, 2011.
[29] Critical Appraisal Skills Programme. Making sense of evidence: 10 questions to help you make sense of qualitative research. England: Public Health Resource Unit, 2006. Online publication available at http://www.sph.nhs.uk/sph-files/casp-appraisal-tools/Qualitative%20Appraisal%20Tool.pdf.
[30] Mays N, Pope C. Assessing quality in qualitative research. BMJ 2000;320:50–2.
[31] Tong A, Sainsbury P, Craig J. Consolidated criteria for reporting qualitative research (COREQ): a 32-item checklist for interviews and focus groups. Int J Qual Health Care 2007;19:349–57.
[32] Elliot B. Using narrative in Social Research: Qualitative and Quantitative Approaches. London, UK: Sage Publications, 2005.
[33] Clark J. How to peer review a qualitative manuscript. In: Godlee F, Jefferson T, ed. Peer Review in Health Sciences. London, UK: BMJ Books, 2003.
[34] Morse JM. Enhancing the usefulness of qualitative inquiry: gaps, directions, and responsibilities. Qual Health Res 2002;12:1419–26.
[35] Whittemore R, Chase SK, Mandle CL. Validity in qualitative research. Qual Health Res 2001;11:522–37.
[36] Cohen DJ, Crabtree BF. Evaluative criteria for qualitative research in health care: controversies and recommendations. Ann Fam Med 2008;6:331–9.
[37] Yardley L. Dilemmas in qualitative health research. Psychol Health 2000;15:215–28.
[38] Anger JT, Nissim HA, Le TX, et al. Women's experience with severe overactive bladder symptoms and treatment: insight revealed from patient focus groups. Neurourol Urodyn In press.
[39] Ivarsson AB, Soderback I, Ternestedt, BM. The mearning and form of occupational therapy as experienced by women with psychoses: a phenomenological study. Scand J Caring Sci 2002;16:103–10.
[40] Ungar WJ, Mirabelli C, Cousins M, et al. A qualitative analysis of a dyad approach to health-related quality of life measurement in children with asthma. Soc Sci Med 2006;63:2354–66.
[41] Thomas DR. A general inductive approach for analyzing qualitative evaluation data. Am J Eval 2006;27:237–46.
[42] Muhr T, ed. User's Manual for ATLAS.ti 5.0. 2nd ed. Berlin: Scientific Software Development, 2004.
[43] Guest G, Bunce A, Johnson, L. How many interviews are enough? An experiment with data saturation and variability. Field Methods 2006:18:59–82.
[44] Overall JE, Gorham DR. The brief psychiatric rating scale. Psychol Rep 1962;10:799–812.