

## Using Patient-Reported Outcomes in Registries

### Draft White Paper for Third Edition of

### “Registries for Evaluating Patient Outcomes: A User’s Guide”

## 1. What are patient-reported outcomes and why all the recent interest?

As the medical system refocuses on delivering patient-centered care, the importance of measuring and reporting those aspects of health and well-being that are best described by patients themselves, whether related to disease, treatment, or both is increasingly recognized.<sup>1-4</sup> Discrepancies exist between patient and clinician estimates of both the prevalence and severity of patients’ symptoms as well as functional impairments, highlighting the importance of direct patient reporting.<sup>3,5-9</sup> Collectively, such reports of health status taken directly from patients without interpretation by clinicians are known as patient-reported outcomes (PROs) [Table 1]. PROs are a subgroup of patient outcomes, which are more general and reflect any outcome related to a patient, whether reported by the patient or described by a third party (e.g., imaging, laboratory evaluation, clinician assessment).

Over the past 20 years, an expanding body of literature has demonstrated that PROs are associated with traditional outcomes, such as overall survival<sup>10-15</sup> and tumor response.<sup>16</sup> PROs themselves are increasingly recognized as valid outcomes (e.g., quality of life, pain, breathlessness, physical functioning).<sup>17-26</sup> Systematic collection of PROs in clinical trials, patient registries, and usual clinical care is feasible and efficient.<sup>27-31</sup> PROs are more reflective of underlying health status than physician reporting<sup>32</sup> and facilitate discussion of important symptoms and quality of life (QoL) with clinicians.<sup>33</sup> Additionally, they have been shown to serve as supporting documentation,<sup>28</sup> improve symptom management,<sup>34</sup> and potentially impact clinical decision making,<sup>29,35</sup> all of which are viewed favorably.<sup>29</sup> As a matter of terminology, the term “health-related quality of life” (HRQoL) has emerged as the preferential choice in recent literature, and there are cogent arguments surrounding its use. However, the more general “QoL” reflects the fact that health status impacts numerous aspects of daily life and impacts overall QoL. Thus, further discussions in this chapter will consistently use the term QoL.

While widespread adoption of PROs as a key component in clinical research has not occurred, there is increasing recognition of their role in complementing traditional clinical and administrative data. To this

end, the importance of incorporating PROs into clinical research has been highlighted by a number of national policy-making organizations.<sup>2,36</sup> Recently, the United States Food and Drug Administration (FDA) identified PROs as the regulatory standard for supporting subjective endpoints, like symptoms, in drug approval and labeling, and their updated guidance distributed in December 2009 provides clear instructions on PRO measurement in drug development trials.<sup>37</sup> Certainly, the purposes of PROs in registry studies and in support of labeling claims do not align perfectly, and registry studies are generally not viewed as an avenue for product labeling, but the guidance provided by the FDA has helped refine the definition of PROs and expand the sphere of interest surrounding their use. Most importantly, the FDA guidance document has established a benchmark, albeit a high one, for PRO data and has been the focus of much recent PRO-related literature (references too numerous to list). For this reason, the standards set by the FDA are heavily referenced in the following discussion.

Presently, there are no evidence-based guidelines for inclusion of PROs in registries, leading to substantial heterogeneity in capture and reporting of PROs in this setting (see, for example, the review about some large registries in rheumatoid arthritis).<sup>38</sup> Recent initiatives to define how PROs should be used in oncology comparative effectiveness research (CER) are instructive,<sup>39</sup> as they reflect current, collaborative opinions of many different stakeholders, and may serve as a template for inclusion of PROs in registries [Table 2].

## 2. What is the role of PROs in registries?

### 2.1. Relationship between PROs and CER

Comparative effectiveness research was recently defined by the Institute of Medicine as:

“... the generation and synthesis of evidence that compares the benefits and harms of alternative methods to prevent, diagnose, treat, and monitor a clinical condition or to improve the delivery of care. The purpose of CER is to assist consumers, clinicians, purchasers, and policy makers to make informed decisions that will improve health care at both the individual and population levels.”<sup>40</sup>

Central to this definition is that the information generated by CER should assist consumers of health care (i.e., patients) in making decisions. Of great interest to patients are factors like QoL, symptom burden, and functional status, which are best described directly by patients, thereby implicitly emphasizing the importance of PROs to CER.<sup>41,42</sup> The strength of this relationship is furthered by the term patient-centered outcomes research (PCOR), which has emerged after passage of the Patient Protection and

Affordable Care Act that established the Patient-Centered Outcomes Research Institute (PCORI).

According to PCORI:

“Patient-centered outcomes research helps people make informed health care decisions and allows their voice to be heard in assessing the value of health care options. This research answers patient-focused questions: (1) ‘Given my personal characteristics, conditions and preferences, what should I expect to happen to me?’ (2) ‘What are my options and what are the benefits and harms of those options?’ (3) What can I do to improve the outcomes that are most important to me?’ (4) ‘How can the health care system improve my chances of achieving the outcomes I prefer?’”<sup>43</sup>

By definition, PCOR is impossible to pursue without including the patient voice and PROs are an important tool for capturing the patient voice. As PCOR is effectively a subset of CER (and will not be referred to independently from this point), PROs are therefore critical components of CER, as well. The importance of PROs in CER is highlighted by the interest in the patient experience of the multiple stakeholders who ultimately utilize results of CER.<sup>42</sup>

## 2.2. Relationship between CER and registries

While clinical trials are generally felt to represent the gold standard of evidence to support clinical decisions, many clinical trials are conducted under conditions that limit generalizability or do not emphasize factors that are important to patients and clinicians in the course of actual practice. Clinicians and patients face challenging decisions regarding treatment choices and toxicity profiles that are unaddressed by traditional clinical trials, and these are exactly the types of questions that CER is intended to address. Registries are important tools for answering such questions. They can evaluate effects in a more “real-world” population, improving generalizability. In uncommon diseases, where traditional clinical trials are unrealistic because of small numbers, registries can help fill the information void on any number of issues, including treatment options and responses, natural history, and QoL. Registries can be designed to answer specific questions that impact clinical practice, but were unaddressed by pivotal clinical trials. Importantly, when partnered with electronic health records (EHRs), registries can capitalize on the massive amounts of data collected as part of routine clinical care to create datasets that more realistically replicate the array of inputs that clinicians and patients assimilate in almost every clinical encounter. Electronic PRO instruments that are directly incorporated into routine clinical care, and thus directly into an EHR, are potentially important sources of PRO data for registry studies. Collection and analysis of such datasets, in the form of registries, offers the opportunity to inform clinical care in ways that are meaningful to all stakeholders in the health care system.

## 2.3. Importance of PROs in registries

Having established the centrality of PROs to CER and the role of registries in CER, the importance of PROs to registries is apparent. Inclusion of PROs in prospectively collected registries is almost always appropriate. PROs contribute information across the spectrum of registry purposes described in Chapter 1 (Patient Registries)<sup>1</sup> including describing the natural history of disease, determining effectiveness, measuring or monitoring safety or harm, and measuring quality. As one walks down the list of nominated purposes of registries, the substantive role of PROs in registry design becomes increasingly important.

### 2.3.1 Describing natural history of disease

A requirement of registries intended to describe natural history of disease is adequate information about symptom burden and related QoL trajectories, especially in the setting of rare diseases, inherited diseases with increasing life span (e.g., cystic fibrosis, sickle cell disease), and heterogeneous diseases (e.g., chronic obstructive pulmonary disease, breast cancer). Registries can provide useful information on the expected course of health, even in the absence of treatment, which could provide useful information regarding need for and timing of treatment. Understanding how new therapies impact patient experience can also be captured under this rubric. For example, metastatic renal cell carcinoma is a relatively uncommon malignancy for which the FDA has approved six targeted therapies within the past decade. All have different toxicity profiles and different symptom alleviation profiles; insufficient information can be derived from the pivotal clinical trials to develop optimal strategies for sequencing and timing of these therapies.<sup>44</sup> Registries of patients receiving routine care on these different agents (i.e., “real-world” registries), especially when containing PRO data, can help inform sequencing, timing and impact of treatments, providing critical information where there is an explosion of treatment options but a dearth of comparative information.

### 2.3.2. Determining effectiveness

In registries designed to determine effectiveness, PROs also figure prominently, especially considering the importance placed upon the patient experience as a meaningful outcome in the IOM’s definition of CER. Beyond traditional outcome measures such as overall survival and risk reduction, QoL is a valid marker of efficacy by itself and is best captured by PRO measures. Patient-reported symptoms can be indicators of adverse consequences of therapy (e.g., toxicity monitoring), targets for meaningful intervention (e.g., symptom control intervention), and means of understanding how patient perceptions of

---

<sup>1</sup> Chapters referenced in this document can be found in the second edition of “Registries for Evaluating Patient Outcomes: A User’s Guide,” available at: <http://www.effectivehealthcare.ahrq.gov/ehc/products/74/531/Registries%20nd%20ed%20final%20to%20Eisenberg%209-15-10.pdf>.

toxicities or effectiveness impact effectiveness (e.g., through adherence behavior). Consider a prospective registry intended to support CER for the management of early stage prostate cancer. For these patients, differentiating between and comparing surgery and radiation is best achieved from patient-reported information on symptoms of radiation proctitis, sexual health, pain, and urinary function, as well as the relationship of these factors to overall QoL and patient preference.

Within the area of toxicity monitoring, PROs are likely to take a place on center stage. The National Cancer Institute has recently developed a patient-reported version of its Common Terminology Criteria for Adverse Events, PRO-CTCAE<sup>45</sup> for use in cancer clinical trials. Pharmacovigilance studies provide another fertile area for PRO implementation. Perhaps even more powerful are efforts to link PROs to genomic and proteomic data in order to understand the biologic basis for toxicity phenotype. Registries intended for safety monitoring offer potential for a much more robust understanding of long-term safety than typical clinical efficacy trials and when coupled with data on effectiveness may help answer difficult questions such as “Was the intervention worth it?” especially as viewed through the patient’s lens.

### **2.3.3. Quality measurement**

Registries intended to measure quality can incorporate PROs in numerous ways and PROs can contribute to quality assessment. In some instances, established quality standards do not exist, and registries can be used to establish realistic and acceptable standards. For example, there is an impetus to initiate quality monitoring in palliative medicine programs, but the evidence base is insufficient to establish benchmarks to define quality.<sup>46</sup> In such a setting, registries incorporating PROs would serve an important role in establishing definitions for quality and could then be used in real-time to monitor quality. However, some quality metrics focused on the patient experience already exist. For example, in the American Society of Clinical Oncology’s Quality Oncology Practice Initiative assessment and management of pain, nausea/vomiting and dyspnea are core metrics; this requires both PRO assessment and response to findings.<sup>47,48</sup>

## **2.4. PROs in prospective versus retrospective registries**

Having established the role for PROs across a spectrum of registries, it is important to consider the roles of PROs in prospective and retrospective registries. Patients’ experiences are transient and are best captured “in the moment.” They cannot be recreated or recalled precisely, thus highlighting the need to routinely and systemically capture PROs for prospective registries. Further, abundant evidence demonstrates that third party assessments (most notably clinicians) do not adequately reflect patients’ subjective experience with care.<sup>7,8,27,49</sup> For example, in patients with lung cancer receiving chemotherapy, Basch et al showed that, when compared to physician assessments, patient reports of symptoms were

more reflective of daily health status, as measured by EuroQoL EQ-5D.<sup>32</sup> As rapid-learning healthcare systems<sup>50-52</sup> become standard, routine capture of longitudinal and systematic PROs will happen as part of routine care, thereby simplifying the process to prospectively capture PROs for registry support.

As opposed to prospective registries, which can be designed to collect PROs as data accrue, registries constructed by manual chart extraction or from EHR queries should not attempt to retrospectively add PRO data that was not originally collected. Additionally, researchers should not ask patients to provide recalled/recreated PROs for missing data in such registries, as this may introduce recall bias. The exact length of time over which recall bias develops is unclear, and seems to vary for different experiences.<sup>53</sup> For pain, single-item assessments reflecting the prior week do not seem to represent actual pain levels as well as a mean of daily pain levels collected for the same one week period.<sup>54</sup> Thus, asking patients to precisely recall their symptom experience associated with a clinic visit at some arbitrary point in the past is fraught with pitfalls.

## 2.5. Other general considerations on inclusion of PROs in registries

Including PROs in registries offers numerous advantages. First, incorporation of the patient voice helps keep care and research patient-centered, acknowledging the balance and tension between traditional outcomes and PROs. Further, symptom burden, QoL, and satisfaction with care are dynamic variables that cannot be recreated accurately through retrospection; they are essentially lost if not captured “in the moment.” For this reason, routine, systematic, and longitudinal collection is recommended and should be a standard of practice. The importance of longitudinal collection cannot be overstated; it allows patients to serve as their own control; that is, each patient serves as his or her own experiment. Changes from baseline are tracked over time and linked to other interventions, such as initiation or discontinuation of a drug, or outcomes, such as change in disease status (e.g., cancer progression, cardiac event). Serial PROs address a number of critical issues. They: (1) improve our understanding of the trajectory of individual patient’s symptom burden and QoL over the course of disease (or treatment); (2) remind clinicians of the variability between patients; (3) provide information on the value that the individual patient places on their health state; and, (4) are central to the efforts of CER, pharmacovigilance studies, and quality monitoring. When routine and systematic collection of PROs is incorporated into registries, the healthcare community can improve efficiency of routine care through support of billing and clinical documentation functions.

Certainly, including PROs in registries poses challenges. Collection of PROs can generate significant amounts of data and adds another layer of complexity to already complex datasets. Clinician acceptance may lag slightly for several reasons.<sup>55</sup> Although the history (patient reports filtered through a clinician’s

lens) and physical exam are central to clinical diagnosis and decision making,<sup>56</sup> long-standing and deeply ingrained beliefs persist that clinician assessment alone is objective and unbiased, casting doubt upon the value and validity of unfiltered, direct patient reports. Regardless, collection of PROs generates more data for clinicians to consider and incorporate into care, which could be viewed as onerous and burdensome, especially since PROs are not yet ubiquitous or the standard of care. More importantly, it is largely unclear how PROs collected within the context of clinical research should be used to inform care and change daily practice patterns. Without appropriate infrastructure for responding to critical reports, collection of PROs may pose a liability if critical data do not receive appropriate clinician attention and response. For example, significant liability could result if a patient reports a constellation of symptoms known to be strongly associated with suicidal behavior and there is inadequate clinical intervention. Further, it is possible that PROs could lead to decreased satisfaction with care if patients expect that their PROs will be reviewed and addressed, but are unmet or unacknowledged in the clinical encounter.<sup>28</sup>

### **3. What methods are available to collect PROs and which is best?**

Often, choice of PRO instrument and mode of administration are considered jointly, however, they need not be, as administration methods simply provide a platform for collecting and presenting information. There are two main ways of collecting PRO data – on paper and electronically.

#### **3.1. Paper-based methods**

Historically, PROs were collected via paper forms and were developed based on this collection method. From a practical standpoint, collection of PRO data via paper-based methods is relatively straightforward. After selecting the instrument(s) to be used (discussed further in Section 4), consistency is the guiding principle. Items should be presented in the same order for every collection. If the PRO measurement selected is a single-item tool, this is automatic, but if multiple instruments are employed, presenting them in the same order is important. Patients should complete forms in a confidential space, without fear that “wandering eyes” will see responses. Once forms are completed, they should be reviewed multiple times for completeness. For those instruments completed in clinic, this review should be done by staff collecting the instruments, nurses involved in patient intake and rooming, and clinicians reviewing responses. Once forms are submitted to the research team for data entry, completeness should be reassessed. Patients who fail to complete a pre-defined percentage of questions (there is no consensus on an acceptable percentage), should receive a follow-up telephone inquiry to attempt to minimize missing data. Finally, data should be entered into electronic forms using double data entry techniques to enhance transcription accuracy, ideally augmented with near real-time exploratory analyses to examine the believability of the data within the clinical context.<sup>57</sup>

Paper forms are the historical gold standard for PRO collection. For this reason, patients are inherently familiar with them. Their use is not limited by unfamiliarity or unease with new technologies, although unfamiliarity with new technology dissipates quickly and patients are increasingly familiar with technology as advances continue to disseminate. They do not require significant upfront capital investment, in terms of devices or software. There are a plethora of measurement instruments across a variety of disease states that have been extensively evaluated and are available for immediate use.

However, paper forms have many limitations. They require research personnel to sort, distribute, and collect, introducing risk for inconsistencies and a source of ongoing cost. Paper forms collected as part of routine/scheduled clinic visits are generally straightforward, but this approach systematically misses participants unwilling or unable to attend a clinic appointment. Collection between visits is logistically difficult with paper forms; delivery of the paper forms either requires that participants take paper booklets home with them or that research personnel coordinate timely delivery of booklets through the postal service. With either approach, obtaining a time/date stamp for at-home, paper-based administration remains a challenge. Relying on at-home paper booklets risks participants completing multiple days of reporting all at once (i.e., the so-called “parking lot” effect<sup>58</sup> in which all responses for the past month are completed immediately before a visit while sitting in the parking lot). Paper forms often include illegible or uninterpretable responses and require manual data entry, which is administratively burdensome and subject to transcription errors. Manual entry also generates a lag time in monitoring response rates, complicating the process of reducing missing data.<sup>59</sup> Overall, there is a threshold beyond which the continuing data collection and quality assurance costs of paper-based PROs surpass the upfront technology costs for electronic data capture, making electronic PROs the cheaper and more reliable approach.

### **3.2. Electronic capture methods**

With the advent of portable and more cost-effective electronic capture methods, the presence of such methods within the literature has grown recently. Similar to traditional paper-based collection, electronic collection begins with instrument(s) selection. Integral to the choice of instruments is the choice of platform, as not all instruments are tested across multiple platforms, nor is every instrument amenable to every platform. Electronic PRO (ePRO) capture has been demonstrated on a variety of platforms, including web-based, electronic tablets, interactive voice response system (IVRS), handheld device, and digital pen. For ePRO collection using tablet computers or handheld devices in the clinic setting, patients are provided the device at the time of check-in to clinic with pre-loaded PRO measures such that patients simply select their response to each item as it is presented. With the digital pen, patients select responses on a specially designed paper survey, with responses electronically recorded by the pen. With IVRS,

patients call a telephone number and are prompted, via an automated transcript, to select a preferred language, provide an identifier and then are guided through the PRO measure, providing verbal responses to each item. Access to web-based platforms can be provided at “confidential” computer stations in clinic waiting rooms, or in the exam room itself, as well as from any web-enabled device including home computers, handheld devices, and mobile telephones. Regardless of platform, data are transmitted to a central, secure repository immediately upon submission and can be accessed for “real-time” incorporation into routine care, if desired. Both web-based and IVRS collection platforms can extend beyond the clinic and capture PROs between visits. Factors influencing platform selection include budget and technical support, technology literacy of the registry’s target population, collection logistics (in-clinic, between-visit, or combination), and the instrument(s) chosen.<sup>59</sup>

Electronic methods of PRO capture have been widely shown to be feasible in a variety of practice settings, disease states, and age ranges.<sup>28,29,60</sup> Recently developed PRO measures have either been created specifically for electronic data capture or include features to capitalize on electronic capture technologies, such as the Patient-Reported Outcomes Measurement Information System (PROMIS),<sup>61-63</sup> the PRO-CTCAE,<sup>45</sup> and the Patient Care Monitor, version 2 (PCM).<sup>64</sup> The PROMIS and PRO-CTCAE tools take advantage of electronic functionalities such as skip logic or computerized adaptive testing, which can reduce the number of items patients have to complete, while the PCM also fulfills clinical documentation needs for clinical review of systems and triggers for accompanying patient education.

In terms of obtaining hardware or software for these purposes, hardware often requires an upfront investment. Again, the size of the investment depends largely upon the scope and scale of the registry. Some software packages are publicly available (e.g., PROMIS Initiative items) while others are proprietary. Third party commercial vendors specializing in design and implementation of PROs offer a variety of products. The decision to involve a commercial vendor depends upon factors like the rationale for including PROs in the registry, the size of the registry and number of involved sites, local technological expertise and support, whether the data will be collected as part of routine care or just for research purposes, and the degree of psychometric analysis needed. Although registry studies are not viewed as sufficiently rigorous for product labeling, exploratory analyses of PROs from a registry may serve as the basis for a subsequent trial for labeling purposes, in which case having a sound PRO measure in the registry could simplify the trial process. In such a scenario, using a commercial vendor to ensure adequate audit trails and compliance with all FDA guidance for PROs would be prudent. Alternatively, consider a healthcare system with an extensive EHR which plans a registry to monitor the impact of a series of clinical pathways to lessen the debilitation following major abdominal surgery; they may elect to

develop or modify a PRO system to be directly integrated with their EHR without involving an ePRO vendor.

Compared to paper methods, delivery of ePROs can be automated, minimizing the risk of inconsistent presentation of materials or mishandling paper forms. Electronic collection of responses provides immediate and accurate time/date stamps, and facilitates real-time monitoring of response rates and review for missing data.<sup>59</sup> Additionally, electronic platforms may provide a safer environment for patients to disclose sensitive concerns, such as sexual function.<sup>65</sup>

Not all PRO measures were developed for, or have been tested on, electronic administration platforms. The transition of paper-based measures to electronic platforms is referred to as “migration” and guidelines were recently developed to assess the equivalence of measures that have migrated from one collection mode to another.<sup>66</sup> In general, paper to electronic migration yields between-mode equivalence comparable to the test-retest reliability of the original mode, but this is not always the case and should be tested.<sup>67</sup> When incorporating a migrated PRO measure into a registry, registry developers should verify that the ePRO measure has demonstrated validity in the intended mode of administration or reasonable equivalence with the mode for which validity, reliability, and sensitivity were initially demonstrated.<sup>39</sup>

Although electronic capture provides substantive advantages over paper-based methods, enthusiasm must be tempered on several fronts. First, completion of electronically delivered PRO measures requires some level of comfort with and access to newer technologies, which may prove challenging in certain situations. For example, in rural areas, using web-based methods to collect PROs between visits may be impractical due to unpredictable internet access, while some geriatric populations may be uncomfortable with tablet or handheld technologies. Second, if paper-electronic equivalence has not already been verified for a migrated PRO instrument, the process of documenting equivalence can be time-consuming and expensive. Finally, electronic methods require greater up-front investment in terms of the devices and software, electronic storage (meeting appropriate security standards), training, and technical support. Depending upon the scale of the registry, these issues may render electronic methods too burdensome.

### **3.2.1. Specific considerations on software selection**

Software selection is a common question. While outside the scope of this chapter, some broad advice can be provided. First, there are many companies that offer software to collect ePROs. Publicly available software is also in production (e.g., PROMIS) or being developed (e.g., ePRO CTCAE). The software solution itself is relatively simple and expensive systems are not needed, unless specific features are required (e.g. requirement to be compliant with the FDA’s CFR Part 11). Software should be from a credible vendor, with available security documentation. Since patients will likely enter Protected Health

Information (PHI), the system should be appropriately compliant with the Health Insurance Portability & Accountability Act (HIPAA). Avoid using survey software where HIPAA compliance and other requirements cannot be documented.

In general, patients should report one item per screen, the screen should be clear and move to the next item when the answer is provided, and there shouldn't be any software delays between questions. Visually the software should present questions and response "buttons" in large enough font for easy reading by mildly visually impaired individuals. Validation code and verifications should be built into the software, as well as any required clinical triggers. It should be easily adaptable, and easily integrated into the registry workflow. Reports (e.g., for clinicians) should be visually appealing, efficient and informative. Whenever possible, software should connect into the EHR workflow, including embedding data into the EHR for clinical documentation and/or contributing to an enterprise data warehouse.

Finally, ensure that the software has been tested before full-scale implementation with the registry. Request testing documentation from the vendor, who should have completed this. Both usability and feasibility should be considered, and it should be conducted with the planned population for the registry. As elaborated on [www.usability.gov](http://www.usability.gov), usability is not a single, one-dimensional property of the interface, but rather a synthesis of:

- Ease of learning - How fast can a user who has never seen the user interface before learn it sufficiently well to accomplish basic tasks?
- Efficiency of use - Once an experienced user has learned to use the system, how fast can they accomplish tasks?
- Memorability - If a user has used the system before, can they remember enough to use it effectively the next time or does the user have to start over again learning everything?
- Error frequency and severity - How often do users make errors while using the system? How serious are these errors, and how do users recover from these errors?
- Subjective satisfaction - How much does the user like using the system?

The degree of usability testing should match the complexity of the task. For an ePRO system, this process minimally includes documentation of respondents' ability to navigate the electronic platform, follow instructions, and answer questions, with an overall goal of demonstrating that respondents can complete the computerized assessment as intended. Generally, less than ten representative patients are required to verify usability. If the system is not usable, then it should be iteratively updated until it is usable.

Feasibility extends usability and establishes the practical implementation of the software system in the local setting (e.g., clinic, home, hospital). Assessment approaches are similar and the software goes through iterative updates until feasible. During this process, patients can contribute critical advice for the “help” manual and instruction sets.

Although most often associated with questionnaire development, cognitive debriefing is also appropriate for usability and feasibility assessment through verbal probing by the interviewer (e.g., “What does the instruction ‘skip item’ mean to you here?”) and “thinking aloud” in which the interviewer asks the respondent to verbalize whatever comes to mind as they conduct a task. Incorporated in usability and feasibility testing, cognitive debriefing helps to assess whether the ePRO system influences the way respondents interpret the questions, decide on an answer, and respond. In addition, it can help to determine whether the instructions are clear or if anything is confusing.

### 3.3. Which method is best?

As with most other aspects involving PROs in registries, the choice of PRO capture method is highly dependent upon the design and purpose of the registry. Both paper-based and electronic platforms offer advantages and disadvantages, as outlined above. Ideally, when either method is shown to be valid for an instrument, both methods of PRO data collection should be available in a study. Providing an interface familiar to or preferred by particular patients or populations may reduce missing data not at random. Modes may be mixed across patients in a study (e.g., each patient selects a specific mode at baseline and continues to report via that mode throughout a study), or within patients (e.g., a patient reports by web until he becomes symptomatically ill, at which point IVRS becomes preferable). One mode may be preferred at a particular site, for example in multinational studies where IVRS or web access are heterogeneous across countries. “Real-world” registries are likely to enroll patients from a variety of settings (e.g., home, hospital, assisted living facility) and circumstances (e.g., independent, caregiver-assisted), such that flexibility in mode of administration facilitates capturing a broad mix of patients. Mixing modes is generally viewed as acceptable if a reasonable level of between-mode equivalence has been demonstrated.<sup>39</sup>

In general, electronic capture is preferred to paper because of its flexibility and its ability to reduce the chance that the PRO data in a registry will be missing. In contemporary research, paper methods are usually most cost effective until registries start to grow in size or number of sites. When the registry is going to be intentionally small (e.g., less than 100 patients), paper methods will likely suffice. When the registry is going to be large, upfront investments in electronic approaches will realize substantial downstream gains in efficiency, cost, and data quality. Regardless of the ultimate choice of

administration method, clear documentation of the rationale for the choice and clear evidence of appropriate psychometric assessment is strongly recommended. Assistance with this process may arise from internal expertise (as in many academic institutions) or may rely upon input from a commercial vendor, whose involvement can range from consulting only to nearly full control of the development and implementation process.

#### 4. Which PRO measure should be selected?

The process of choosing which PRO(s) to include in a registry can be challenging, largely because the plethora of available measures is overwhelming. In 2007, a PubMed search for PRO instrument development articles since 1995 resulted in more than 2000 citations.<sup>68</sup>

Existing PRO measures assume a variety of forms:

- general assessment scales (e.g., health-related QoL)
- disease-specific scales [e.g., chronic obstructive pulmonary disease, cancer (including scales for individual tumor types), arthritis, or psoriasis]
- symptom-specific scales (e.g., pain, breathlessness, distress)
- evaluations of functioning across a variety of domains (e.g., physical, social, emotional)
- scales assessing satisfaction with care received
- other (e.g., adherence with therapy)

Some PRO measures are extensive, with dozens of items related to a single concept (e.g., breathlessness), while others have 80 or more items reflecting many different patient-reported concerns constituting an entire clinical review of systems, and yet others are single-item instruments measuring a single construct in a single question.

Further, there is extensive literature describing the important characteristics (i.e., conceptual framework, content validity, reliability, ability to detect change) of PRO measures, but consolidating this information into practical guidance for selecting among existing PRO measures is difficult. The FDA Guidance document has outlined a standard for evaluating PRO measures for labeling claims that encompasses the salient points regarding development history, conceptual framework and psychometric evaluation. The standards outlined by the FDA may be more stringent than is necessary for certain registry purposes, but nevertheless serve as an important and well-conceived framework for discussion and conform to accepted best practices.<sup>27</sup> While a comprehensive review of PRO development and psychometric evaluation is

beyond the scope of this chapter, below is a concise overview of the process and concepts. For more information, several texts provide detailed descriptions.<sup>25,26,69</sup>

#### 4.1. Getting started and the importance of clarity

The key to successfully navigating this process is to clearly define the following aspects of the registry:

- population of interest (e.g., cancer patients receiving radiotherapy for painful bony metastases, individuals with oxygen-dependent chronic obstructive pulmonary disease, children with rhinoconjunctivitis, United States veterans with rheumatoid arthritis)
- outcomes of interest, also known as the concept (e.g., specific symptom severity, overall symptom burden, treatment-related toxicities, physical functioning, social functioning, QoL)
- intended users of the registry (e.g., clinicians, patient advocacy groups, pharmaceutical companies, insurance companies, governmental agencies)
- the purpose(s) of the registry (e.g., pharmacovigilance, establish symptom trajectories, correlate survival benefit with QoL or symptom benefit).

As with any research activity, *a priori* specific aims and hypotheses to be tested must be outlined up front, and PRO selection appropriately aligned. Registry studies, in particular, are susceptible to poorly defined outcomes; PRO instruments may be chosen because they are general in nature and capture a broad range of patient-reported concerns, meet a target goal of demonstrating that PROs are captured rather than capturing specific PRO concepts of interest. If the objectives of the registry, intended hypotheses, and outcomes of interest are clearly defined, the desired characteristics of the PRO instrument become more clearly delineated, facilitating a search of existing measurement instruments.

#### 4.2. Potential sources for identifying PRO instruments

Once these issues are clearly defined, identification of candidate PRO measures can begin in earnest. In general, the process of PRO development is time- and resource-intensive and using existing measures whenever possible is best. It is highly unlikely that any existing instrument will perfectly suit the needs of a registry study, or that a “perfect” instrument can be developed, further underscoring the importance of clearly defining the population, outcomes of interest, and purpose of the registry. Such clarity will allow more appropriate assessment of the relative strengths and weaknesses of existing PRO measures. In many cases, modifications to existing measures will improve the measure for use in a registry. These modifications can include changes in wording or order of questions, adding specific questions, or altering the method of administration. In general, such modifications require some degree of psychometric

reassessment, though the degree to which instrument modification requires psychometric reassessment varies and is discussed by Snyder et al.<sup>70</sup>

Traditional literature searches can yield results, but may be quite time-consuming. The Mapi Institute maintains the Patient-Reported Outcome and Quality of Life Instruments Database (<http://www.proqolid.org>), allowing users to search a large and relatively comprehensive database for PRO instruments that best address the specific needs identified. The Online Guide to Quality-of-life Assessment (<http://www.olga-qol.com>) is another database of existing QoL instruments. Additionally, the US National Institutes of Health PROMIS Initiative (<http://www.nihpromis.org>) has been tasked with developing rigorously tested item banks across a broad range of domains and subdomains (functioning, disability, symptoms, distress, and role participation).<sup>61</sup> The PROMIS Initiative is also actively evaluating methods to achieve brevity in instruments through techniques such as computer adaptive testing. Importantly, these measures are publicly available through the PROMIS Assessment Center (<http://www.assessmentcenter.net>). Commercial vendors can also aid in identifying appropriate measures; as with selecting a mode for administering the PRO measure, the decision to involve a commercial vendor is multifactorial, depending on the factors described in Section 3.2.

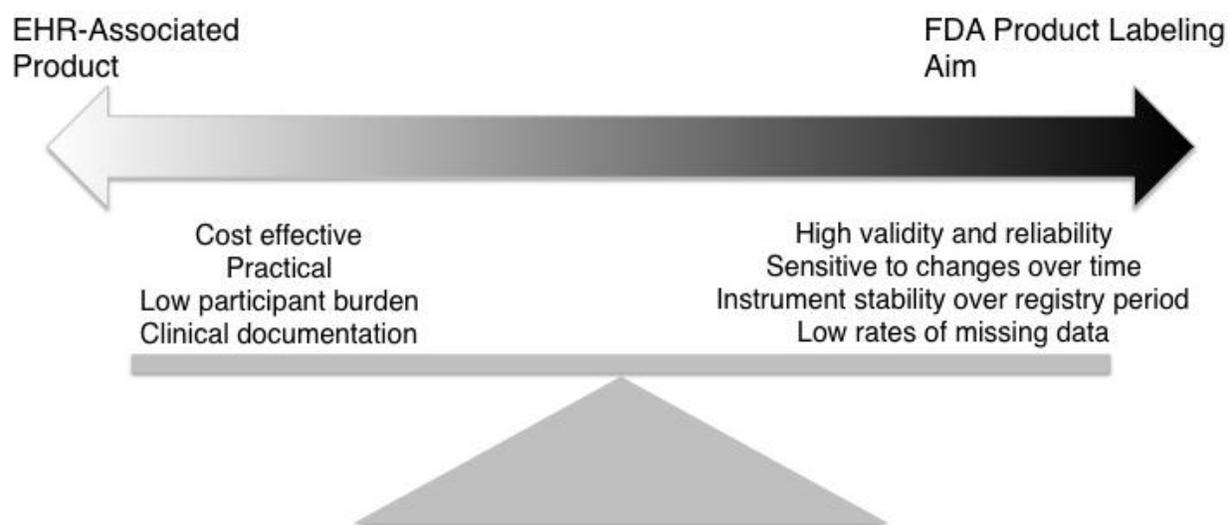
Item banks represent another option for developing PRO surveys. In general, item banks contain comprehensive collections of items that pertain to a particular construct (e.g., dyspnea).<sup>71</sup> Item banks generally rely on item response theory (IRT), in which the unit of focus is the item, rather than the entire instrument. As such, instruments can be constructed using IRT that employ only those items which provide the most useful and relevant information, eliminating questions with little added value, without compromising psychometric qualities.<sup>72</sup> The PROMIS Initiative is an example of an item bank.<sup>71,72</sup> Item banks may represent the future of PRO collection, but they are currently limited by logistical issues, questions about whether IRT-based item banks represent an improvement over existing PRO instruments, concerns over regulatory acceptance, and limited data about psychometric properties of item banks in specific populations.<sup>71</sup> However, IRT-based item banks represent a promising approach, especially in light of the emphasis on limiting respondent burden.

### 4.3. Choice of the best PRO for the registry

Section 4.4 describes many of the properties of PRO instruments that should be considered when choosing the appropriate instrument for each unique registry scenario. Whether to adhere closely to the conservative FDA recommendations is a frequent source of question, if not frank tension. While there is no formal avenue through which registries can support product-labeling claims, if the registry is in any way tied to trials with aspirations of product-labeling then the answer is straightforward and the FDA

PRO guidance should be followed. Anchoring the FDA threshold as a “maximally conservative” (and therefore usually least practical) state, there is a continuum of scenarios and a continuum of practical allowances to the ideal state where the need for precision and reduction of bias is balanced with the need for practical solutions and the reduction of missing data (Figure 1). Explicitly outlining the registry objectives, population, outcomes and intended uses as described in Section 4.1 will help to define where the registry is on the continuum and guide decision making.

**Figure 1. Psychometric properties and logistical considerations exist along a spectrum. The tension between psychometric desirability and logistical considerations of PRO collection in registries requires a careful balance, driven primarily by the goals of the registry.**



#### 4.4. Development history and conceptual framework

The PRO development history and conceptual framework are inextricably linked and are discussed in close proximity for this reason.

##### 4.4.1. Development history

The FDA Guidance document strongly recommends transparency with respect to development history. “Development history” explicitly refers to the entire process of developing and psychometrically evaluating a patient reported outcome measure, including the conceptual framework, item development and revision history, and evidence of patient input. For newly developed PRO instruments, clearly documenting the development history is straightforward and can be integrated into the development process. Contrast this to using an existing measure, where the development history may be very difficult, if not impossible, to obtain. Ideally, the development history is well vetted in the literature, but if the history is somewhat opaque, the FDA has indicated that demonstration of content validity with specific

examples, including direct patient input from the appropriate population, is an acceptable alternative. For newly developed PROs it is imperative, from an FDA and product-labeling standpoint, that the entire development history be well documented. The cornerstone of the development history is the conceptual framework.

#### 4.4.2. Conceptual framework

Clear identification of the target population, purpose of the registry, and outcomes of interest greatly facilitates developing a conceptual framework. According to the FDA Guidance document, a conceptual framework “explicitly defines the concepts measured by the instrument in a diagram that presents a description of the relationships between items, domain (subconcepts), and concepts measured and the scores produced by a PRO instrument.”<sup>37</sup> Initially, the conceptual framework arises out of expert opinion and literature review. The framework is then refined by qualitative methods of patient input, such as patient interviews and focus groups, which ensures that *a priori* hypotheses are consistent with patient experiences and descriptions. The conceptual framework will be modified iteratively.<sup>73</sup> For complex concepts, such as breathlessness, multiple domains impact the overall concept, so identifying appropriate domains and then assessing these is paramount to assessing the overarching concept.

#### 4.5. Psychometric properties

Entire texts are written on psychometrics and there is an extensive literature on psychometric properties of PRO measures. An excellent series arising from the Mayo/FDA Patient-Reported Outcomes Consensus Group focused on PRO development in advance of the anticipated FDA Guidance; it was published in a special supplement of the November/December 2007 issue of the journal *Value in Health* and provides more detailed descriptions of processes and procedures needed to implement PRO systems to meet FDA expectations.

Almost every guideline regarding utilizing PRO measures recommends selecting measures that have demonstrated content validity, criterion validity, reliability and sensitivity (including the ability to detect change over time) in the target population.<sup>39</sup> It is important to note that psychometric properties are not dichotomous and instruments are not completely “valid” or “reliable.” These properties are continuous variables relaying incremental information. Additionally, it is inappropriate to refer to an instrument as “validated” as this simply means it has been subjected to psychometric analysis, but conveys no information regarding the measure’s performance.<sup>74</sup> For this reason, instruments are reflected at varying points on our continuum in Figure 1 to demonstrate that differing states of reliability and validity may be appropriate depending upon the context of the registry and the PROs to be captured within it. The goal,

ultimately, is to identify or develop instruments with acceptable psychometric properties in the population of interest.

#### 4.5.1. Validity

From a psychometric standpoint, validity has three main forms: content, construct, and criterion validity. Content validity is the extent to which the instrument actually measures the concepts of interest. The FDA Guidance understandably places significant emphasis on content validity, consistent with other groups,<sup>75</sup> even stating that without adequate content validity, labeling claims cannot be supported. At face value, the importance of content validity is intuitive; it is important that an instrument assess those concepts it was designed to measure. In general, qualitative evidence, in the form of documented patient input through focus groups, is an important standard in the view of the FDA.<sup>74</sup> Construct validity describes the degree to which what was measured reflects the *a priori* conceptualization of what should be measured.<sup>76</sup> Subcomponents of construct validity are convergent and discriminant validity, which assess the degree of similarity between measures that are theoretically similar (convergent validity) or the extent to which measures that are theoretically different actually differ (discriminant validity). For example, a new measure of anxiety would be expected to have high convergent validity with the anxiety subscale of the Hospital Anxiety and Depression Scale.<sup>77</sup> To that end, the FDA would expect comparisons of new PRO measures with similar existing measures to support construct validity. Criterion validity describes the extent to which the scores of PRO measure reflect the gold standard measure of the same concept.<sup>37</sup> Criterion validity is often difficult to assess in the PRO arena because identifying gold standard measures for many PRO concepts is difficult and the FDA therefore deemphasizes criterion validity.

#### 4.5.2. Reliability

Reliability reflects the ability of an instrument to yield the same result on serial administrations when no change in the concept being measured is expected. The reliability of an instrument is typically assessed via test-retest methods and by measuring the internal consistency.<sup>74</sup>

##### 4.5.2.1. Test-retest reliability

Test-retest reliability describes the ability of an instrument to generate the same results in the same respondent over a period of time during which no change is reasonably expected.<sup>2,37,74</sup> Thus, test-retest reliability assesses the intra-individual variability. Identifying the optimal timeframe for retesting can be challenging, and may vary by disease state and target population.<sup>74</sup>

#### **4.5.2.2. Internal consistency reliability**

Internal consistency reliability reflects the degree to which items within a scale measure the same concept. It can be quantitatively assessed with Cronbach's alpha, which measures the internal consistency of an instrument. Well-established thresholds for interpreting Cronbach's alpha are available; in general, coefficient alpha greater than 0.7 is the minimum acceptable threshold for comparisons between groups.<sup>74</sup>

#### **4.5.3. Ability to detect change**

The ability of a PRO measure to detect change is intuitively important. Demonstration of this ability, according to the FDA, requires that changes in the PRO instrument parallel changes in other factors that indicate a change in the status of the concept of interest. For example, in patients receiving a new treatment for opioid-induced constipation, changes in a PRO instrument designed to assess overall bowel health may be linked with use of certain other bowel products, such as enemas, to establish the ability to detect change. The measure must demonstrate ability to detect both improvements and losses in health status. Further, it is important to detect changes throughout the range of possible values. In registry studies, where longitudinal collection and analysis are critical, understanding the concept of minimally important change detected,<sup>78</sup> rather than establishing that number explicitly, may be sufficient.

#### **4.5.4. Areas of controversy**

The emphasis placed upon content validity has generated some controversy as PRO developers attempt to improve content validity, in part by meticulously wording items and instructions to minimize variations in interpretation between patients. However, the ability to improve content validity likely is asymptotic, in that individual variability undoubtedly influences interpretation of questions in ways that cannot be accounted for, meaning that responses to an instrument capture the patient's true (and unique) perceptions. There are concerns that in the pursuit of greater content validity, other important characteristics of PRO instruments may be underdeveloped or underappreciated.<sup>76</sup> For example, in pursuing greater content validity, the constraints placed upon questions may actually limit patient perspective by forcing some degree of conformity, or may result in misinterpretation of results. Consider a registry of patients with advanced cancer designed to assess the impact of certain interventions upon the development of disability. Upon entering the registry, a patient rates his disability as severe because his reference point is a previously healthy state. Four months later, he rates his disability as mild, though on more open-ended questioning, notes he can simply sit on the front porch and watch his grandchildren as he knows that any other activities are unrealistic and his goal is to simply make it to the front porch. Even though the instrument measures disability from the view of the patient and would thus have adequate content validity, the interpretation regarding the merits of the intervention would be erroneous, as the

patient has clearly become more disabled, but has shifted his frame of reference, a fact which is not captured by content validity. This phenomenon is commonly referred to as “response shift” and has long been recognized as a challenge in QoL research.<sup>79</sup> Alternatively, all measures with marginal content validity may be cast aside without consideration of other properties. Consider two new measures for the same concept tested in different studies with different methodologies, resulting in different content validities. The measure with higher content validity is likely to propagate, even if it is more flawed, simply because of methodological issues.

These arguments on content validity are not intended to undermine the standards established by the FDA, nor should they be viewed as rationale for not adhering to these standards, but are meant to prompt careful consideration of all the psychometric properties of PRO measures, especially in the context of the specific registry. Remember first principles – before anything else, it needs to make good sense, have face validity, be doable, and limit patient burden.

## 4.6. Non-psychometric considerations

Beyond identifying a PRO instrument with desirable psychometric properties, consideration must be given to the *people* that are closely tied to completing and acting upon PRO data and the tension that can exist between impacts on people and psychometric desirability.

### 4.6.1. Patient factors

In designing registries and considering PROs for inclusion, it is important to consider the burden to the patient the PRO measures represent. For instance, lengthy questionnaires may result in increasing missing data over time, as patients grow weary of serially completing such questionnaires. The capacity to answer lengthy instruments cannot be predicted *a priori* and differs between groups. At Duke Cancer Institute, patients in a variety of solid tumor clinics routinely complete 80-86 item instruments without significant fatigue or burnout;<sup>64</sup> median time to complete the survey is 11 minutes, reducing to <8 minutes after several visits in the clinic using the same instrument. While the FDA did not offer specific recommendations on questionnaire length, an upcoming guidance document from the Center for Medical Technology Policy will recommend that, for patients with cancer, completion of PRO instruments take no more than 20 minutes at the initial visit and less than 10 minutes at subsequent visits.<sup>39</sup> Patients should be offered a private space for completing instruments, to minimize concerns regarding confidentiality, especially for sensitive questions. Instructions should be provided for every item, even if it only frames the recall period. The instrument should be delivered with adequate font size and at appropriate literacy levels. Additionally, physical assistance should be provided if needed, such as reading items aloud to patients with visual impairments. While most pilot studies of PRO instruments provide a small amount of

remuneration,<sup>28,64</sup> these studies have demonstrated that the collection of PROs made patients feel encouraged that their clinicians were seeking additional information and felt that the ePRO instrument facilitated communication between patient and clinician.<sup>28</sup> Outside the pilot testing phase, it is not advisable to provide remuneration to patients for completing PRO instruments, even in the setting of a registry study. PRO responses should be shared with clinicians, as this has been shown to be an important aspect of PROs to patients.<sup>28</sup>

#### **4.6.2. Clinician factors**

Even within the research setting, assessing the impact of PRO collection on routine care is important. Will the PRO results be made available immediately as part of routine care or only available to research personnel? Whether or not PRO data are shared with clinicians in real time should be explicitly addressed in the informed consent process. If data are to be made available to clinicians, are appropriate support services available to assist in managing newly identified concerns or issues? Are there mechanisms to support incorporation of PRO data into clinical care, if it will be made available, or will it be “one more thing” for which clinicians are responsible? What will be the impact of the PRO collection on workflow?

Many recent guidelines recommend providing clinician feedback of concerning patient-reported information, such as reports of new chest pain. The thresholds for triggering a clinical alert, components of the alert message, and method of delivering the notice to the clinician must be carefully considered. What are the risk management concerns? How will the clinician’s response be verified? Though often mundane, these factors are important to consider in the implementation phase. Teams experienced in embedding PROs into registries and clinical workflow can provide sage advice as to how to navigate these pathways (e.g., Duke Cancer Care Research Program, <http://www.cancer.duke.edu/dccrp/>); clear guidelines do not exist. See further discussion in Section 4.7 below.

#### **4.6.3. Ensuring data quality**

Collecting quality data is an implicit necessity of any registry. Although assessing data quality can assume many forms, for the purposes of registries, there are two concepts that are critical. The first is to minimize missing data. Missing data are anathema to quality data. Missing data degrades the quality of the information, thereby decreasing its analytic potential. It is essential to anticipate missing data and to plan interventions to reduce missingness. This is especially important in registry studies where time horizons are long and the potential for missing data great. There are a number of steps that can be taken to minimize missing data during the implementation phase of the registry. The most important step is to make sure that the PRO instrument chosen is meaningful, and the role in the registry and related work is

well described, especially to patients and families. Ideally, the PRO measures should be implemented as standard of care, such that they become ubiquitous and desired, not only by patients, but also by clinicians.<sup>51</sup> If this occurs, missing data should decrease. Electronic data collection practically supports real-time, or near real-time, quality monitoring of information being collected in order to identify patterns of missing data, leading to development of targeted interventions to reduce missingness. Additionally, with near real-time quality analysis, backup data collection methods can—and should—be deployed. For example, a central telephone interviewer can contact individuals who did not respond to items (either individually or entire instruments) to both obtain the data and ascertain why the item was omitted. Analytic approaches must include a plan for managing the unavoidable occurrence of missing data; importantly, a “last observation carried forward” approach to handling missing data should be avoided.

The second issue related to data quality is consistency. In registries with long time horizons, it is not uncommon for measurement items, or instruments, to evolve or change entirely. Unfortunately, it is equally uncommon for notations of such changes to be embedded within the data structure, as metadata, such that future analyses can quickly and readily identify which iteration of an instrument was completed at which point in time. Metadata is essentially data about data. More precisely, it is “...structured information that describes, explains, locates, or otherwise makes it easier to retrieve, use, or manage an information source.”<sup>80</sup> Consider a long-term registry where the primary measurement instrument undergoes an iterative update to “version 2” to reflect new knowledge in the field and is quickly implemented into the registry. Though the two versions are likely very similar, they also likely have slightly different questions (in terms of structure or order), psychometric properties, and scoring algorithms. In such a scenario, it is imperative that the version of the instrument completed at any given point in time be identified within the dataset. Further, there may be cases where the person completing the questionnaire may not always be the patient (see discussion in Section 4.6.4). For example, in a palliative care registry, patients are not always able to complete a PRO instrument, even with assistance. The ability of the person to complete the instrument may change over time as cognition wanes. In these settings, proxy-reports involving close family or caregivers may become the only available measures and the only available data to be incorporated into registries; therefore, it is essential to identify, via metadata, who is completing the instrument.

#### **4.6.4. Special populations: Are proxy-reports ever appropriate?**

There are numerous situations in which patients are not physically or cognitively able to provide direct assessment of their experience. Obvious examples include infants and small children, individuals with significant cognitive impairment (congenital or acquired), and those at end-of-life. In such settings, proxy-reports of QoL are often collected,<sup>19</sup> though the literature suggests that proxy-reports demonstrate

moderate agreement, at best, with patient-reports.<sup>81-83</sup> Nevertheless, proxy-reports are viewed as valuable in many of these settings because caregiver or family perception is also an important consideration. The FDA strongly discourages proxy-reports in product-labeling claims.<sup>37</sup> Unfortunately, such an extreme stance leaves these vulnerable populations marginalized. By not considering proxy-reports, symptom-based research and other lines of inquiry in these populations face considerable obstacles with a potential end-result that drugs or products that could improve symptom burden or QoL never have the opportunity to gain FDA approval for such indications. The FDA's position on proxy-reports is emphasized because of the rigorous standard the FDA guidance document establishes, but that position should not devalue the potential role for proxy-reports. Ideally, the extent of agreement between patient- and proxy-reports can be established in advance of use of proxy reports. The PROMIS Initiative is investigating application of existing methods for PROs to proxy-reports to improve performance.<sup>84,85</sup>

#### 4.7. Implementation issues

Upon successful navigation of the challenging process of selecting PRO instruments and the mode of administration comes the daunting task of implementing the selected instruments. Below is a practical framework for successful implementation, centered on achieving data quality and consistency.

Just as with mode of administration, implementing PRO data collection is best achieved if consistency is a central tenet, especially if the registry study is multicenter. In this setting, consistency refers to processes. Standard operating procedures should be established for each site of data collection that delineate, to the extent possible, how patients, researchers, and clinicians interact with the collection system (paper or electronic). As part of standard operating procedures, specific training should be provided, with accessible and easy-to-use manuals available (preferably in both text and video format). Every aspect of the process that can be standardized should be standardized, including the dataset itself. That is, the datasets should include metadata that describe key components important for subsequent analyses and end-users, including who completed the instrument (patient or proxy), where it was completed (e.g., outpatient clinic, home, inpatient ward), which version was administered, and a flag for irregularities identified as part of internal quality control.

Ideally, for multisite studies, these standard operating procedures are the same at each site, with another set of standard operating procedures for the central repository (or coordinating) site that delineates how often data from cooperating sites should be transmitted, how it should be compiled and stored, how often it should undergo quality assessment, and how it should be accessed and distributed for analysis. Within multisite registries, and even within some single site registries, it may be necessary to select an instrument that has been translated into and validated in other languages besides English. It is not adequate to simply

translate an instrument into another language, as the psychometric properties obtained within an American population of patients with disease X are unlikely to be reproduced in a population of Japanese patients with the same disease. Thus, formal assessment of the psychometric properties of the instrument is necessary when translating to another language.

Another aspect of consistency in this setting reflects administering the same instrument over the lifespan of the registry. The strength of this recommendation depends partly upon the purpose of the registry; for registries comparing effectiveness, this consistency is essential, while for a registry focused on quality and embedded within an EHR, this recommendation is less stringent. Nevertheless, if the data are collected prospectively, the strong preference is for consistency in PRO instrument administered. Regardless of purpose, collected data should include metadata labels.

Further, involving the entire healthcare team (physicians, mid-level providers, nurses, administrators, and other support staff) in the development process is essential, especially with respect to integrating the PRO instruments into the clinical workflow and providing clinician feedback. As part of this integration, clinical triggers should be established (and standardized) that explicitly force acknowledgement of a patient report by a provider (e.g., a pain score of 8 out of 10) or initiate some standardized intervention (e.g., a patient reporting a high distress level might be automatically contacted by a psychosocial care support team).<sup>84,85</sup> Such standardized triggers will only be embraced if there is inclusion of the healthcare team in the implementation process. This inclusive implementation process will also help shape the perception of the PRO data, in that buy-in from the healthcare team will make the PRO collection process a necessary and desired component of care, rather than simply an extra task to complete.<sup>84</sup>

Finally, explicitly including the patient voice in the form of PROs has been shown to improve patient well-being and enhance patient-provider communication.<sup>34</sup> Building on this premise, inclusion of PROs in observational studies may improve patient engagement, recruitment, and retention, though there are no data directly supporting this. The experience of the Duke Cancer Care Research Program with ePRO collection as part of routine cancer care has shown remarkable response and participation rates, with rates of missing data, even for sensitive questions such as level of sexual enjoyment, routinely less than 5% (manuscript in preparation). Certainly, more rigorous documentation of improved long-term patient participation with inclusion of PROs is needed before more ardent assertions can be made.

#### **4.8. Summary regarding selecting PRO instruments**

Selecting PRO instruments for inclusion in registry studies is not a one-size-fits-all process. The Center for Medical Technology Policy is preparing a guidance document for inclusion of PROs in adult oncology trials<sup>39</sup> and these recommendations are included as an example [Table 2]. Clear and careful definition of

the target population, concept to be measured, and purpose of the registry is an important first step. For a given population or context, even in a registry, it is important to have some *a priori* hypotheses and justification for outcomes being measured, or the study risks becoming a prospective fishing expedition. As such, there needs to be a systematic approach to selecting salient outcomes (to the extent possible in a registry, which admittedly is sometimes exploratory by nature). In CER, the process of identifying meaningful outcomes requires upfront patient input. But regardless of how the outcomes are selected, there must be a systematic approach to determining whether an outcome is best reported by a patient (i.e., if information about a particular symptom or overall health state or satisfaction is sought, it is best reported from the patient/surrogate perspective, thus a PRO instrument is appropriate). Far too frequently the tail wags the dog in registry studies; that is, PRO instruments are selected first, prior to identifying outcomes of interest. Thus, the rational identification of outcomes of interest early in the process of registry development is important. Such an approach will quickly identify if PROs are appropriate and will produce a sound base for evaluating PRO instruments and administration methods. If this process is navigated effectively, the stage will be set for successful incorporation of PROs into the registry.

After the arduous process of clearly defining the population and outcomes of interest, search for existing PRO instruments that will assess the outcomes of interest. If a suitable measure is not identified, options include modifying an existing measure or developing a new measure. In general, development of new PRO instruments is resource intense, so it is preferable to use an existing measure whenever possible. After identifying (or developing) a measure, administration mode should be selected. Electronic administration is preferred, but not all instruments have been evaluated using electronic administration, though this can be accomplished. Important to the scientific basis of the registry are the psychometric properties of the instrument. While the FDA highly values content validity, it is possible to effectively use an instrument with modest content validity, depending on the purpose of the registry, highlighting the importance of understanding and defining the purpose of the registry.

In most registry studies, the purposes of the study and outcomes of interest will necessitate inclusion of PRO data. Careful planning is essential, in identifying appropriate PRO instruments for inclusion, selecting modes of instrument administration, and implementing the PRO collection system, and when done effectively, this generally produces more complete datasets that truly include the voices of all stakeholders in the healthcare system and are meaningful to all stakeholders.

## EXAMPLE

Consider the division of pulmonary medicine at an academic university. Within the division is a growing multidisciplinary cystic fibrosis (CF) program with a large catchment area and approximately 250 patients ranging in age from 21-65 years, though most patients are younger than age 35. As the program develops, the team plans to implement a series of initiatives targeting not only improved survival, but also improved functioning for patients with CF. Proposed interventions include routine endocrinology consultations for all CF-related diabetes mellitus, improved psychological services, and standardized exercise regimens during hospitalizations. The outcomes of interest for these interventions are equally broad ranging, but include traditional measures such as pulmonary function (as measured by pulmonary function tests), end-organ damage (diabetes, chronic kidney disease), resistant organism colonization rates, hospitalization utilization, symptom burden (including breathlessness, weight change, worry, and fatigue) and quality of life (QoL). The team plans to use a registry for this because they do not feel that they can reasonably test the effectiveness of these interventions through parallel or sequential randomized, controlled trials, but do wish to systematically capture outcomes of interest in a longitudinal manner as the interventions are introduced.

In considering the outcomes of interest, symptom burden and health-related QoL merit closer inspection for inclusion of PROs. Certainly, patients are better positioned to report breathlessness, worry, fatigue and QoL. In fact, most argue that patients are the only valid source of information on these issues, thus inclusion of PROs in this registry is appropriate.

In considering which instruments to use, it is important for the team to consider the relationships between the symptoms under consideration and QoL [Figure 2]. Specifically, the influence of symptom burden on QoL must be weighed carefully, to help determine if a series of single-item instruments is most appropriate or if a multi-item, disease-specific instrument (of which there are several in CF) or another approach is most appropriate. As the team plans to use this registry in a longitudinal fashion for numerous planned interventions and because they want to understand how specific interventions impact certain domains impacted by CF, they select an established, multi-item, multi-domain, CF-specific measure that incorporates an aggregate assessment of QoL, as well as several component domains of well-being.

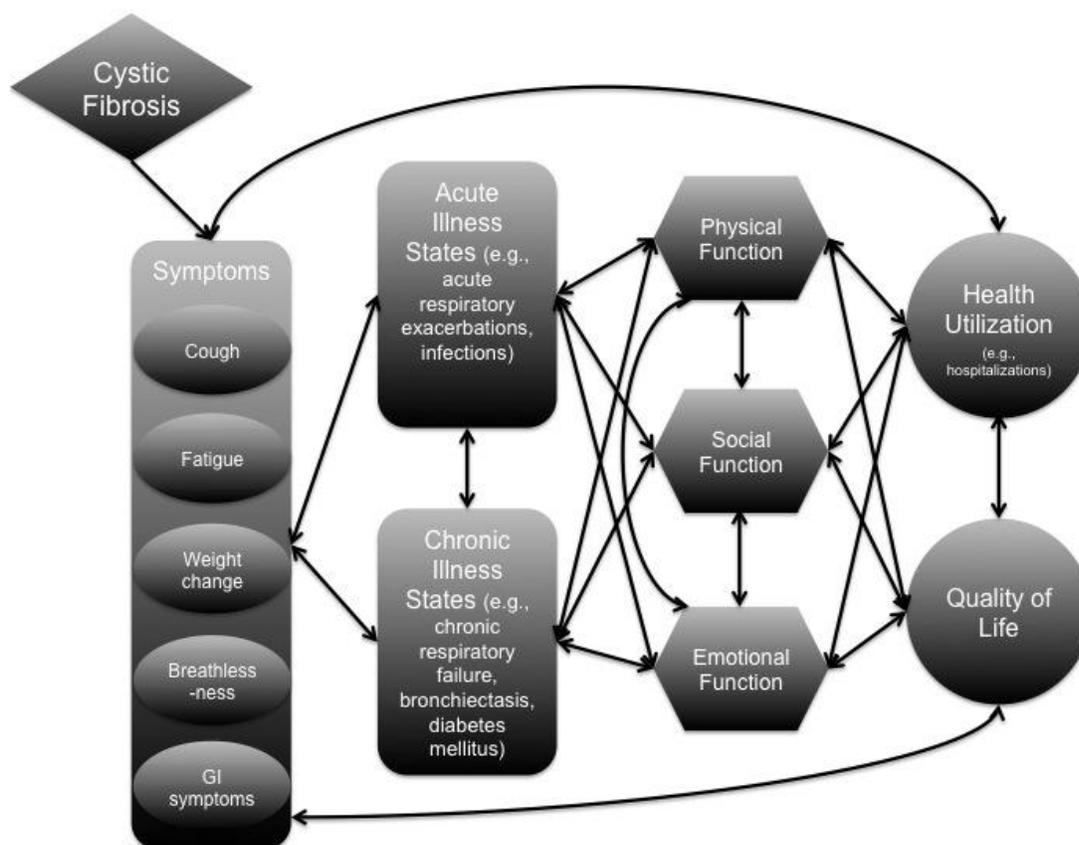
Since the planned settings of intervention include both inpatient and outpatient settings and because of travel issues related to the catchment area, the team also plans to capture reports between visits, such that no more than two months elapse between PRO data collection. For this reason, the team prefers to use electronic methods, but the instrument they selected has only been psychometrically assessed via paper-

based methods. They collaborate with the institutional expert on PROs to document paper-electronic equivalence, and to perform usability and feasibility testing for web-based administration. This pilot study demonstrates that it is reasonable to use a web-based approach for PRO assessments.

From a health-care team standpoint, implementation goes smoothly, since the entire CF team was involved in developing the registry and PRO system. Missing data are minimal for inpatient and clinic appointment collection, as the team heavily advertised the PRO collection system to the patients prior to implementation, provided in-clinic teaching, and used reports during the clinical visits; the instrument quickly becomes viewed as a necessary component to the healthcare encounter. However, as data accumulates, the team identifies patterns in missing data for between-visit administrations. They identify that at-home internet access remains a problem for a small but significant portion of their patients. They receive a grant from the local CF foundation to support internet access for vulnerable patients, with subsequent reduction in missing data.

This example highlights several key points: 1) the importance of understanding the target population; 2) the need to identify outcomes of interest prior to selecting PRO instruments as the outcomes of interest should dictate the instrument, not vice versa; 3) the benefit of incorporating PRO instruments into longitudinal, routine care.

Figure 2. Simplified concept map illustrating some of the relationships that exist around health-related quality of life in cystic fibrosis.



## TABLES

Table 1. Definitions of commonly encountered terms within PRO-related literature

Term	Definition
Ability to detect change	Evidence that a PRO instrument can identify differences in scores over time in individuals or groups who have changed with respect to the measurement concept. <sup>37</sup>
Clinician reported outcome (ClinRO)	Outcomes that are either observed by the physician (e.g., cure of infection and absence of lesions) or require physician interpretation (e.g., radiologic results and tumor response). In addition, ClinROs may include formal or informal scales completed by the physician using information about the patient. <sup>86</sup>
Concept	The specific measurement goal, or the thing that is measured by a PRO. <sup>37</sup>
Conceptual framework	Explicitly defines the concepts measured by the instrument in a diagram that presents a description of the relationships between items, domain (subconcepts), and concepts measured and the scores produced by a PRO instrument. <sup>37</sup>
Construct validity	The degree to which what was measured reflects the a priori conceptualization of what should be measured. <sup>76</sup>
Content validity	The extent to which the instrument actually measures the concepts of interest. <sup>74</sup>
Criterion validity	The extent to which the scores of PRO measure reflect the gold standard measure of the same concept. <sup>37</sup>

<b>Term</b>	<b>Definition</b>
Domain	A subconcept represented by a score of an instrument that measures a larger concept comprised of multiple domains. <sup>37</sup>
Health-related quality of life	The subjective assessment of the impact of disease and treatment across the physical, psychological, social and somatic domains of functioning and well-being. <sup>87</sup>
Instrument	A means to capture data (i.e., a questionnaire) plus all the information and documentation that supports its use. Generally, that includes clearly defined methods and instruction for administration or responding, a standard format for data collection, and well-documented methods for scoring, analysis, and interpretation of results in the target population. <sup>37</sup>
Item	An individual question, statement, or task (and its standardized response options) that is evaluated by the patient to address a particular concept. <sup>37</sup>
Item bank	A comprehensive collection of questions (and their response options) designed to measure an underlying construct across its entire continuum. <sup>71</sup>
Metadata	Structured information that describes, explains, locates, or otherwise makes it easier to retrieve, use, or manage an information source. <sup>80</sup>
Patient-reported outcome (PRO)	A measurement based on a report that comes directly from the patient (i.e., study subject) about the status of a patient's health condition without amendment or interpretation of the patient's response by a clinician or anyone else. <sup>37</sup>
Proxy-reported outcome	A measurement based on a report by someone other than the patient reporting as if he or she is the patient. <sup>37</sup>
Quality of life	An individual's perception of their position in life in the context of the culture and value systems in which they live and in relation to their goals, expectations, standards and concerns. It is a broad ranging concept affected in a complex way by the person's physical health, psychological state, level of independence, social relationships, personal beliefs and their relationship to salient features of their environment. <sup>88</sup>
Recall period	The period of time patients are asked to consider in responding to a PRO item or question. <sup>37</sup>
Reliability	The ability of an instrument to yield the same result on serial administrations when no change in the concept being measured is expected. <sup>74</sup>
Scale	The system of numbers of verbal anchors by which a value or score is derived for an item. Examples include VAS, Likert scales, and rating scales. <sup>37</sup>
Score	A number derived from a patient's response to items in a questionnaire. A score is computed based on a prespecified, validated scoring algorithm and is subsequently used in statistical analyses of clinical results. <sup>37</sup>

**Table 2. Example guidelines for PRO incorporation into product-labeling claims in oncology from the Center for Medical Technology Policy<sup>39</sup> (used with permission).**

<b>Selection of Measures</b>
1. Include patient-reported outcomes in all prospectively designed comparative effectiveness research and post-marketing studies in adult oncology (including registries, observational cohorts, and controlled trials).
2. Include systematic assessment of the following 14 patient-reported symptoms ("Core" symptom set) in all CER and post-marketing clinical studies in adult oncology: anorexia, anxiety, constipation, depression, diarrhea, dyspnea, fatigue, insomnia, mucositis, nausea, pain, sensory neuropathy, rash, vomiting.

<b>Selection of Measures</b>
3. Include additional patient-reported symptoms as appropriate to a specific study's population, intervention, context, objectives, and setting (in addition to the Core symptom set), and incorporate a process that allows individual patients to report unsolicited symptoms.
4. Measure quality of life (QOL), either via a single-item or multi-item questionnaire, in all prospective CER and post-marketing clinical studies. Inclusion of a measure which enables cost-utility analysis is encouraged.
5. Selected measures to assess symptoms or QOL should have demonstrated content validity (based on direct patient input), criterion validity, reliability, and sensitivity in the intended patient population (including assessment of the meaningfulness of specific score changes and the ability to detect change over time), as well as an appropriate recall period. Linguistic translations should be conducted in accordance with existing methodological standards.
<b>Implementation Methods</b>
6. Limit PRO data collection so that the average patient can complete the process within 20 minutes at the initial (baseline) visit and within 10 minutes at any subsequent time points.
7. Collect PROs as frequently as necessary to meet research objectives, without overburdening patients. When using PROs to assess potential treatment benefits, collection of PROs at baseline and following treatment completion or study withdrawal as well as at selected long-term time points should be considered a minimum standard. When using PROs to assess treatment toxicities/harms or comparative tolerability, more frequent assessment is merited such as at baseline and every 1-4 weeks during active therapy as well as at selected long-term time points.
8. Collect PROs via electronic means whenever possible.
9. Establish measurement equivalence when mixing modes of PRO measure administration in a study (e.g., web, telephone/interactive voice response [IVRS], handheld device, and/or paper).
10. Employ methods to minimize missing PRO data including education of local site personnel, training of patients, and real-time monitoring of adherence with backup data collection.
<b>Data Analysis and Reporting</b>
11. Include a plan for analyzing and reporting missing PRO data in the protocol.
12. Report the proportion of patients experiencing a change from baseline demonstrated as being meaningful to patients for each PRO measure.
13. Evaluate the cumulative distribution of responses for each PRO measure and include cumulative distribution curves in reports and publications.
14. Include a mechanism for alerting clinical staff in real-time about symptoms of concerning severity reported by patients during study participation.
15. Analyze and publish results of PRO analyses simultaneously with other clinical outcomes.

## REFERENCES

<sup>1</sup> Fairclough DL. *Design and Analysis of Quality of Life Studies in Clinical Trials, Second Edition (Chapman & Hall/CRC Interdisciplinary Statistics)*. 2nd ed. Chapman and Hall/CRC; 2010:424.

<sup>2</sup> Lipscomb J, Gotay CC, Snyder CF. Patient-reported outcomes in cancer: a review of recent research and policy initiatives. *CA Cancer J Clin*. 2007;57(5):278–300.

<sup>3</sup> Hewlett SA. Patients and clinicians have different perspectives on outcomes in arthritis. *J Rheumatol*. 2003;30(4):877–879.

- 
- <sup>4</sup> Flynn KE, Lin L, Ellis SJ, et al. Outcomes, health policy, and managed care: Relationships between patient-reported outcome measures and clinical measures in outpatients with heart failure. *Am Heart J*. 2009;158(4):S64–S71.
- <sup>5</sup> Cleeland CS, Sloan JA, ASCPRO Organizing Group. Assessing the Symptoms of Cancer Using Patient-Reported Outcomes (ASCPRO): searching for standards. *J Pain Symptom Manage*. 2010;39(6):1077–1085.
- <sup>6</sup> Bushmakina AG, Cappelleri JC, Taylor-Stokes G, et al. Relationship between patient-reported disease severity and other clinical outcomes in osteoarthritis: a European perspective. *J Med Econ*. 2011;14(4):381–389.
- <sup>7</sup> Pakhomov SV, Jacobsen SJ, Chute CG, Roger VL. Agreement between patient-reported symptoms and their documentation in the medical record. *Am J Manag Care*. 2008;14(8):530–539.
- <sup>8</sup> Basch E, Iasonos A, McDonough T, et al. Patient versus clinician symptom reporting using the National Cancer Institute Common Terminology Criteria for Adverse Events: results of a questionnaire-based study. *Lancet Oncol*. 2006;7(11):903–909.
- <sup>9</sup> Stacy M, Bowron A, Guttman M, et al. Identification of motor and nonmotor wearing-off in Parkinson's disease: comparison of a patient questionnaire versus a clinician assessment. *Mov Disord*. 2005;20(6):726–733.
- <sup>10</sup> Norekvål TM, Fridlund B, Rokne B, et al. Patient-reported outcomes as predictors of 10-year survival in women after acute myocardial infarction. *Health Qual Life Outcomes*. 2010;8:140.
- <sup>11</sup> Wang XS, Shi Q, Lu C, et al. Prognostic value of symptom burden for overall survival in patients receiving chemotherapy for advanced nonsmall cell lung cancer. *Cancer*. 2010;116(1):137–145.
- <sup>12</sup> Singh JA, Nelson DB, Fink HA, Nichol KL. Health-related quality of life predicts future health care utilization and mortality in veterans with self-reported physician-diagnosed arthritis: the veterans arthritis quality of life study. *Semin Arthritis Rheum*. 2005;34(5):755–765.
- <sup>13</sup> Cunningham WE, Crystal S, Bozzette S, Hays RD. The association of health-related quality of life with survival among persons with HIV infection in the United States. *J Gen Intern Med*. 2005;20(1):21–27.
- <sup>14</sup> Mapes DL, Lopes AA, Satayathum S, et al. Health-related quality of life as a predictor of mortality and hospitalization: the Dialysis Outcomes and Practice Patterns Study (DOPPS). *Kidney Int*. 2003;64(1):339–349.
- <sup>15</sup> Gotay CC, Kawamoto CT, Bottomley A, Efficace F. The prognostic significance of patient-reported outcomes in cancer clinical trials. *Journal of Clinical Oncology*. 2008;26(8):1355–1363.
- <sup>16</sup> Victorson D, Soni M, Cella D. Metaanalysis of the correlation between radiographic tumor response and patient-reported outcomes. *Cancer*. 2006;106(3):494–504.
- <sup>17</sup> Rolfson O, Kärrholm J, Dahlberg LE, Garellick G. Patient-reported outcomes in the Swedish Hip Arthroplasty Register: results of a nationwide prospective observational study. *J Bone Joint Surg Br*. 2011;93(7):867–875.
- <sup>18</sup> Castellano D, del Muro XG, Perez-Gracia JL, et al. Patient-reported outcomes in a phase III, randomized study of sunitinib versus interferon- $\alpha$  as first-line systemic therapy for patients with metastatic renal cell carcinoma in a European population. *Ann Oncol*. 2009;20(11):1803–1812.
- <sup>19</sup> Thornburg CD, Calatroni A, Panepinto JA. Differences in Health-Related Quality of Life in Children With Sickle Cell Disease Receiving Hydroxyurea. *J Pediatr Hematology/Oncology*. 2011;33(4):251–254.
- <sup>20</sup> Molnar-Varga M, Molnar MZ, Szeifert L, et al. Health-related quality of life and clinical outcomes in kidney transplant recipients. *Am J Kidney Dis*. 2011;58(3):444–452.
- <sup>21</sup> Pronzato P, Cortesi E, van der Rijt CC, et al. Epoetin alfa improves anemia and anemia-related, patient-reported outcomes in patients with breast cancer receiving myelotoxic chemotherapy: results of a European, multicenter, randomized, controlled trial. *Oncologist*. 2010;15(9):935–943.
- <sup>22</sup> Abernethy AP, McDonald CF, Frith PA, et al. Effect of palliative oxygen versus room air in relief of breathlessness in patients with refractory dyspnoea: a double-blind, randomised controlled trial. *Lancet*. 2010;376(9743):784–793.
- <sup>23</sup> Deshpande AD, Sefko JA, Jeffe DB, Schootman M. The association between chronic disease burden and quality of life among breast cancer survivors in Missouri. *Breast Cancer Res Treat*. 2011;129(3):877–886.

- 
- <sup>24</sup> Sprangers MAG. Disregarding clinical trial-based patient-reported outcomes is unwarranted: Five advances to substantiate the scientific stringency of quality-of-life measurement. *Acta Oncol.* 2010;49(2):155–163.
- <sup>25</sup> Lipscomb J, Gotay CC, Snyder C. *Outcomes assessment in cancer.* Cambridge Univ Pr; 2005:662.
- <sup>26</sup> Streiner DL, Norman GR. *Health Measurement Scales: A practical guide to their development and use.* 4th ed. Oxford University Press, USA; 2008:428.
- <sup>27</sup> Basch E, Iasonos A, Barz A, et al. Long-term toxicity monitoring via electronic patient-reported outcomes in patients receiving chemotherapy. *J Clin Oncol.* 2007;25(34):5374–5380.
- <sup>28</sup> Abernethy AP, Herndon JE, Wheeler JL, et al. Feasibility and acceptability to patients of a longitudinal system for evaluating cancer-related symptoms and quality of life: pilot study of an e/Tablet data-collection system in academic oncology. *J Pain Symptom Manage.* 2009;37(6):1027–1038.
- <sup>29</sup> Basch E, Artz D, Dulko D, et al. Patient online self-reporting of toxicity symptoms during chemotherapy. *J Clin Oncol.* 2005;23(15):3552–3561.
- <sup>30</sup> Snyder CF, Jensen R, Courtin SO, Wu AW, Website for Outpatient QOL Assessment Research Network. PatientViewpoint: a website for patient-reported outcomes assessment. *Qual Life Res.* 2009;18(7):793–800.
- <sup>31</sup> Dudgeon DJ, Knott C, Chapman C, et al. Development, implementation, and process evaluation of a regional palliative care quality improvement project. *J Pain Symptom Manage.* 2009;38(4):483–495.
- <sup>32</sup> Basch E, Jia X, Heller G, et al. Adverse symptom event reporting by patients vs clinicians: relationships with clinical outcomes. *J Natl Cancer Inst.* 2009;101(23):1624–1632.
- <sup>33</sup> Berry DL, Blumenstein BA, Halpenny B, et al. Enhancing patient-provider communication with the electronic self-report assessment for cancer: a randomized trial. *J Clin Oncol.* 2011;29(8):1029–1035.
- <sup>34</sup> Velikova G, Booth L, Smith AB, et al. Measuring quality of life in routine oncology practice improves communication and patient well-being: a randomized controlled trial. *J Clin Oncol.* 2004;22(4):714–724.
- <sup>35</sup> Detmar SB, Muller MJ, Schornagel JH, Wever LDV, Aaronson NK. Role of health-related quality of life in palliative chemotherapy treatment decisions. *J Clin Oncol.* 2002;20(4):1056–1062.
- <sup>36</sup> National Quality Forum. *A National Framework and Preferred Practices for Palliative and Hospice Care Quality: A Consensus Report.* Oxford University Press, USA; 2006:124. Available at: <http://www.qualityforum.org>.
- <sup>37</sup> US Department of Health and Human Services, Food and Drug Administration. *Guidance for industry: Patient-reported outcome measures: Use in medical product development to support labeling claims.* 2009:1–43. Available from: <http://www.fda.gov/downloads/Drugs/GuidanceComplianceRegulatoryInformation/Guidances/UCM193282.pdf>. Accessed July 26, 2011.
- <sup>38</sup> Curtis JR, Jain A, Askling J, et al. A comparison of patient characteristics and outcomes in selected European and U.S. rheumatoid arthritis registries. *Semin. Arthritis Rheum.* 2010;40(1):2–14.e1.
- <sup>39</sup> Center for Medical Technology Policy (CMTPT). *Effectiveness Guidance Document: Recommendations for Incorporating Patient-Reported Outcomes into the Design of Post-Marketing Clinical Trials in Adult Oncology.* Available at [www.cmtptnet.org](http://www.cmtptnet.org). November 2011; Forthcoming.
- <sup>40</sup> Institute of Medicine (U.S.). Committee on Comparative Effective Research Prioritization, Institute of Medicine (U.S.). *Initial national priorities for comparative effectiveness research.* National Academies Press; 2009:227.
- <sup>41</sup> Wu AW, Snyder C, Clancy CM, Steinwachs DM. Adding the patient perspective to comparative effectiveness research. *Health Aff (Millwood).* 2010;29(10):1863–1871.
- <sup>42</sup> Hirsch BR, Giffin RB, Esmail LC, et al. Informatics in action: lessons learned in comparative effectiveness research. *Cancer J.* 2011;17(4):235–238.
- <sup>43</sup> Patient-Centered Outcomes Research Institute. <http://www.pcori.org/patient-centered-outcomes-research/>. Accessed September 20, 2011.
- <sup>44</sup> Plimack ER, Hudes GR. Selecting targeted therapies for patients with renal cell carcinoma. *J Natl Compr Canc Netw.* 2011;9(9):997–1006.

- 
- <sup>45</sup> Basch EM, Reeve BB, Mitchell SA, et al. Electronic toxicity monitoring and patient-reported outcomes. *Cancer J*. 2011;17(4):231–234.
- <sup>46</sup> Kamal AH, Currow DC, Ritchie C, et al. The value of data collection within a palliative care program. *Curr Oncol Rep*. 2011;13(4):308–315.
- <sup>47</sup> Neuss MN, Desch CE, McNiff KK, et al. A process for measuring the quality of cancer care: the Quality Oncology Practice Initiative. *J Clin Oncol*. 2005;23(25):6233–6239.
- <sup>48</sup> Champion FX, Larson LR, Kadlubek PJ, Earle CC, Neuss MN. Advancing performance measurement in oncology: quality oncology practice initiative participation and quality outcomes. *J Oncol Pract*. 2011;7(3 Suppl):31s–5s.
- <sup>49</sup> Fromme EK, Eilers KM, Mori M, Hsieh Y-C, Beer TM. How accurate is clinician reporting of chemotherapy adverse effects? A comparison with patient-reported symptoms from the Quality-of-Life Questionnaire C30. *J Clin Oncol*. 2004;22(17):3485–3490.
- <sup>50</sup> Abernethy AP, Etheredge LM, Ganz PA, et al. Rapid-learning system for cancer care. *J Clin Oncol*. 2010;28(27):4268–4274.
- <sup>51</sup> Abernethy AP, Herndon JE, Wheeler JL, et al. Improving health care efficiency and quality using tablet personal computers to collect research-quality, patient-reported data. *Health Serv Res*. 2008;43(6):1975–1991.
- <sup>52</sup> Etheredge LM. A rapid-learning health system. *Health Aff (Millwood)*. 2007;26(2):w107–18.
- <sup>53</sup> Stone AA, Broderick JE, Schwartz JE. Validity of average, minimum, and maximum end-of-day recall assessments of pain and fatigue. *Contemp Clin Trials*. 2010;31(5):483–490.
- <sup>54</sup> Stone AA, Broderick JE, Kaell AT. Single momentary assessments are not reliable outcomes for clinical trials. *Contemp Clin Trials*. 2010;31(5):466–472.
- <sup>55</sup> Bainbridge D, Seow H, Sussman J, et al. Multidisciplinary health care professionals' perceptions of the use and utility of a symptom assessment system for oncology patients. *J Oncol Pract*. 2011;7(1):19–23.
- <sup>56</sup> Paley L, Zornitzki T, Cohen J, et al. Utility of clinical examination in the diagnosis of emergency department patients admitted to the department of medicine of an academic hospital. *Arch. Intern. Med*. 2011;171(15):1394–1396.
- <sup>57</sup> Day S, Fayers P, Harvey D. Double data entry: what value, what price? *Control Clin Trials*. 1998;19(1):15–24.
- <sup>58</sup> Rand CS, Wise RA, Nides M, et al. Metered-dose inhaler adherence in a clinical trial. *Am Rev Respir Dis*. 1992;146(6):1559–1564.
- <sup>59</sup> Rose M, Bezjak A. Logistics of collecting patient-reported outcomes (PROs) in clinical practice: an overview and practical examples. *Qual Life Res*. 2009;18(1):125–136.
- <sup>60</sup> Salaffi F, Gasparini S, Grassi W. The use of computer touch-screen technology for the collection of patient-reported outcome data in rheumatoid arthritis: comparison with standardized paper questionnaires. *Clin. Exp. Rheumatol*. 2009;27(3):459–468.
- <sup>61</sup> Cella D, Yount S, Rothrock N, et al. The Patient-Reported Outcomes Measurement Information System (PROMIS): progress of an NIH Roadmap cooperative group during its first two years. *Med Care*. 2007;45(5 Suppl 1):S3–S11.
- <sup>62</sup> Garcia SF, Cella D, Clauser SB, et al. Standardizing patient-reported outcomes assessment in cancer clinical trials: a patient-reported outcomes measurement information system initiative. *J Clin Oncol*. 2007;25(32):5106–5112.
- <sup>63</sup> Reeve BB, Burke LB, Chiang Y-P, et al. Enhancing measurement in health outcomes research supported by Agencies within the US Department of Health and Human Services. *Qual Life Res*. 2007;16 Suppl 1:175–186.
- <sup>64</sup> Abernethy AP, Zafar SY, Uronis H, et al. Validation of the Patient Care Monitor (Version 2.0): a review of system assessment instrument for cancer patients. *J Pain Symptom Manage*. 2010;40(4):545–558.
- <sup>65</sup> Dupont A, Wheeler J, Herndon JE, et al. Use of tablet personal computers for sensitive patient-reported information. *J Support Oncol*. 2009;7(3):91–97.

- 
- <sup>66</sup> 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(4):419–429.
- <sup>67</sup> Gwaltney CJ, Shields AL, Shiffman S. Equivalence of electronic and paper-and-pencil administration of patient-reported outcome measures: a meta-analytic review. *Value Health*. 2008;11(2):322–333.
- <sup>68</sup> Turner RR, Quittner AL, Parasuraman BM, Kallich JD, Cleeland CS, Mayo/FDA Patient-Reported Outcomes Consensus Meeting Group. Patient-reported outcomes: instrument development and selection issues. *Value Health*. 2007;10 Suppl 2:S86–93.
- <sup>69</sup> Fayers P, Machin D. *Quality of Life: The Assessment, Analysis and Interpretation of Patient-reported Outcomes*. 2nd ed. Wiley; 2007:566.
- <sup>70</sup> Snyder CF, Watson ME, Jackson JD, Cella D, Halyard MY, Mayo/FDA Patient-Reported Outcomes Consensus Meeting Group. Patient-reported outcome instrument selection: designing a measurement strategy. *Value Health*. 2007;10 Suppl 2:S76–85.
- <sup>71</sup> Flynn KE, Dombeck CB, Dewitt EM, Schulman KA, Weinfurt KP. Using item banks to construct measures of patient reported outcomes in clinical trials: investigator perceptions. *Clin Trials*. 2008;5(6):575–586.
- <sup>72</sup> Cella D, Riley W, Stone A, et al. The Patient-Reported Outcomes Measurement Information System (PROMIS) developed and tested its first wave of adult self-reported health outcome item banks: 2005–2008. *J Clin Epidemiol*. 2010;63(11):1179–1194.
- <sup>73</sup> Wilson IB, Cleary PD. Linking clinical variables with health-related quality of life. A conceptual model of patient outcomes. *JAMA*. 1995;273(1):59–65.
- <sup>74</sup> Frost MH, Reeve BB, Liepa AM, Stauffer JW, Hays RD, Mayo/FDA Patient-Reported Outcomes Consensus Meeting Group. What is sufficient evidence for the reliability and validity of patient-reported outcome measures? *Value Health*. 2007;10 Suppl 2:S94–S105.
- <sup>75</sup> Terwee CB, Bot SDM, de Boer MR, et al. Quality criteria were proposed for measurement properties of health status questionnaires. *J Clin Epidemiol*. 2007;60(1):34–42.
- <sup>76</sup> McClimans LM, Browne J. Choosing a patient-reported outcome measure. *Theor Med Bioeth*. 2011;32(1):47–60.
- <sup>77</sup> Snaith RP. The Hospital Anxiety And Depression Scale. *Health Qual Life Outcomes*. 2003;1:29.
- <sup>78</sup> Norman GR, Sloan JA, Wyrwich KW. Interpretation of changes in health-related quality of life: the remarkable universality of half a standard deviation. *Med Care*. 2003;41(5):582–592.
- <sup>79</sup> Sprangers MA, Schwartz CE. The challenge of response shift for quality-of-life-based clinical oncology research. *Ann Oncol*. 1999;10(7):747–749.
- <sup>80</sup> NISO. *Understanding Metadata*. NISO Press; 2004:1–20.
- <sup>81</sup> Varni JW, Thissen D, Stucky BD, et al. PROMIS(®) Parent Proxy Report Scales: an item response theory analysis of the parent proxy report item banks. *Qual Life Res*. 2011 Oct. 5; Epub Ahead of Print:DOI 10.1007/s11136-011-0025-2.
- <sup>82</sup> Kutner JS, Bryant LL, Beaty BL, Fairclough DL. Symptom distress and quality-of-life assessment at the end of life: the role of proxy response. *J Pain Symptom Manage*. 2006;32(4):300–310.
- <sup>83</sup> Jones JM, McPherson CJ, Zimmermann C, et al. Assessing agreement between terminally ill cancer patients' reports of their quality of life and family caregiver and palliative care physician proxy ratings. *J Pain Symptom Manage*. 2011;42(3):354–365.
- <sup>84</sup> Basch E, Abernethy AP. Supporting clinical practice decisions with real-time patient-reported outcomes. *J Clin Oncol*. 2011;29(8):954–956.
- <sup>85</sup> Cleeland CS, Wang XS, Shi Q, et al. Automated symptom alerts reduce postoperative symptom severity after cancer surgery: a randomized controlled clinical trial. *J Clin Oncol*. 2011;29(8):994–1000.
- <sup>86</sup> Willke RJ, Burke LB, Erickson P. Measuring treatment impact: a review of patient-reported outcomes and other efficacy endpoints in approved product labels. *Control Clin Trials*. 2004;25(6):535–552.

<sup>87</sup> Revicki DA, Osoba D, Fairclough D, et al. Recommendations on health-related quality of life research to support labeling and promotional claims in the United States. *Qual Life Res.* 2000;9(8):887–900.

<sup>88</sup> The WHOQOL Group. The World Health Organization Quality of Life assessment (WHOQOL): position paper from the World Health Organization. *Soc Sci Med.* 1995;41(10):1403–1409.