Chapter 9
Symptoms as Patient-Reported Outcomes in Cancer Patients Undergoing Immunotherapies

Tito R. Mendoza

Abstract Cancer therapies are toxic. Newer oncological treatments such as immunotherapy produce unconventional adverse events that are collectively referred to as immune-related adverse events (irAEs). These irAEs are clinician-rated and typically reported via tabulation of adverse events from the National Cancer Institute's Common Terminology Criteria for Adverse Events (CTCAE). However, the symptomatic effects of treatment and the severity of disease are best reported by the patient themselves. Although many pivotal trials for immunotherapeutic agents include health-related quality-of-life measures, symptom-focused assessments are more proximal to the effects of treatment and disease burden. This chapter discusses how best to measure symptoms, describes the desirable properties of a psychometrically valid symptom assessment tool, reviews available symptom assessment tools, provides methods to assist in the interpretation of PRO data, elucidates the feasibility and benefit of incorporating PRO in several cancer cohorts, describes the current use of PROs in immunotherapy, and identifies areas where further research are needed to enhance the use of PROs in cancer patients undergoing immunotherapy.

Keywords Patient-reported outcomes · Symptoms · Immunotherapy · Cancer

Introduction

Cancer is a disease with symptoms that profoundly impair a patient's quality of life and ability to function. Symptoms are further exacerbated by newer cancer treatments such as immunotherapies that have revolutionized the treatment of various cancers by reinvigorating a suppressed immune system. Because of this disruption in immune balance, a unique set of side effects referred to as immune-related adverse events (irAEs) have emerged. These irAEs are typically clinician-rated and may not

T. R. Mendoza (✉)
Department of Symptom Research, The University of Texas MD Anderson Cancer Center, Houston, TX, USA
e-mail: tmendoza@mdanderson.org

A. Naing, J. Hajjar (eds.), *Immunotherapy*, Advances in Experimental Medicine and Biology 995, https://doi.org/10.1007/978-3-030-02505-2_9

be consistent with patient's reports of their symptoms. In order to accurately measure symptoms, we must rely on the use of patient-reported outcomes (PRO).

Symptoms, like health-related quality of life, is a PRO because the patients themselves are the best source of information. However, unlike health-related quality of life, symptom is more proximal to the effect of treatment and the disease. Health-related quality of life is a much broader concept than symptom.

This chapter describes how best to measure symptoms using patient-reported outcomes, discusses the desirable properties of a psychometrically valid symptom assessment tool, reviews available symptom assessment tools, provides methods to assist in the interpretation of PRO data, elucidates the feasibility and benefit of incorporating PRO in several cancer cohorts, describes the current use of PROs in immunotherapy, and identifies areas where further research are needed to enhance the use of PROs in cancer patients undergoing immunotherapy.

Importance of Symptom Assessment

Patient's inability to tolerate treatment-related symptoms often precludes full and effective treatment, and residual symptoms of treatment may limit the functioning of those who may be in remission. Most symptom-focused interventions are typically designed with the goal of usually reducing the severity and impact of symptoms. Because patients commonly face choices among treatments that are similarly effective for tumor control and prolonging survival, differences in the patient's symptoms during the survival period is a major factor in making individualized treatment choices and in developing new therapies. Hence, the ability to compare treatment-related symptoms provides a benchmark for evaluating various cancer treatments. Quality assurance also depends on information about the extent and severity of symptoms. All of these approaches require accurate symptom measurement.

Symptoms and Patient-Reported Outcomes

A symptom report is the patient's statement of their perception of disturbance in normal function that is caused by disease or treatment of disease. Although symptoms are based on complex biological and behavioral phenomena, as subjective experiences their measurement is typically restricted to self-report. Because a symptom can only be known through the patient's *subjective* report, it is by definition a patient-reported outcome (PRO). In contrast, a sign or laboratory value, such as elevated white blood cell count or reduced hemoglobin, is *objective* evidence of the presence of a disease or toxicity of therapy.

The use of PROs continues to increase over the years for several reasons. First, the National Institutes of Health (NIH), as part of its Roadmap Program, has made a significant investment in the development of a measurement system called the

Patient-Reported Outcomes Measurement Information System to increase the measurement precision of patient self-report questionnaires [1]. Second, the US Food and Drug Administration (FDA) has issued guidance for the pharmaceutical industry entitled *patient-reported outcome measures: use in medical product development to support labeling claims*, which provides guidance on how self-report measures are to be used for making claims about the effectiveness of agents for which approval is being sought [2]. Third, the National Cancer Institute realized the shortcomings of their Common Terminology Criteria for Adverse Events (CTCAE) and therefore commissioned contract work to develop a patient-reported outcome version of the CTCAE coined as the PRO-CTCAE [3].

Symptom Reports as Proximal Measure of Disease and Treatment

Patient-reported outcomes can assume many forms such as health status, patient satisfaction, symptom severity, and functional impact. As alluded to earlier, symptoms are generally seen as a subset of health-related quality of life (HRQOL). HRQOL is a multidimensional construct comprising at least four dimensions: physical function (e.g., daily activities, self-care), psychological function (e.g., emotional or mental state, mood), social role function (e.g., social interactions, family dynamics), and disease-related or treatment-related symptoms (e.g., pain, nausea) [4]. Commonly used HRQOL measures, including the Medical Outcomes Study Short Form-36 (SF-36) [5], the Functional Assessment of Cancer Therapy (FACT) [6], and the European Organisation for Research and Treatment of Cancer Quality of Life Questionnaire (EORTC QLQ-C30) [7], address major symptoms such as pain, depression, fatigue, and nausea. In the EORTC QLQ-C30, 18 of 30 items are self-reported symptoms. HRQOL measures also ask questions about various dimensions of patient perception, such as societal role function and concerns about social support. In most conceptualizations of HRQOL, symptoms can be viewed as the patient report closest to the physical and psychological perceptions of the disease process and the immediate effects of treatment on these perceptions [8].

Symptom Measurement

Symptoms are only known by what people tell us. Statements about symptoms (such as, “I have terrible back pain”) are reports of experiences that have common meaning to the person (patient) reporting the symptom and to the person (clinician or caregiver) receiving the report. A person who has never experienced pain might find a pain report hard to comprehend. Unlike height or weight, pain, fatigue, or feeling sad cannot be measured with a measuring stick or a weighing scale.

Symptom measurement depends on our understanding of how symptoms are communicated between the person experiencing the symptom and the people who need to know about it. Because self-reported symptoms are subjective, they are typically described using "constructs," or internal mental states that we cannot measure directly. Rather, we deduce that construct through a set of questions or items that underlie that construct. For example, to understand the construct of pain, we ask questions about the severity of pain and how pain impact daily functions. The measurement of such constructs as symptoms depends on the science of psychometrics, a field of study that originated in educational testing because of the need to know how best to measure intelligence and educational achievement. We can ask many questions with some seemingly more relevant than others. The primary goal of psychometrics is in managing the precision of self-report. Psychometrics concerns itself with reducing the measurement error so that each item provides maximum information about the construct that we are trying to approximate [9]. Two commonly used psychometric metrics are reliability and validity of a scale.

Desirable Properties of a Symptom Measure

Measures of Reliability

Test-retest reliability. If patients are asked about their symptoms more than once within a short time frame and symptoms are not expected to change, symptom ratings should be very similar each time. In general, the correlations between the ratings of the same item at these various times are considered adequate if they equal or exceed 0.70 [10]. This type of reliability is known as "test-retest reliability." Because the symptoms of patients with cancer can change quite rapidly, test-retest reliability should be assessed in patients whose symptoms and disease status are relatively stable during the specified assessment times.

Internal consistency reliability. Another measure of reliability is internal consistency, or the degree to which individual items in a measure correlate with the total score to which the item contributes. One of the most widely used measures of internal consistency reliability is the Cronbach alpha [11]. The Cronbach alpha can be thought of as the average correlation calculated from all possible combinations of items when split into two half-tests.

Measures of Validity

Content validity. Self-report measures need to be more than just stable or reliable. The term "validation" is sometimes used broadly to include all the steps used to evaluate a self-report instrument. However, in a more technical psychometric sense, "validity"

refers to evidence that the assessment instrument is actually capturing the concept or concepts it is designed to measure. An assessment instrument has content validity if it appears to measure the construct of interest. Content validity is related to face validity, which reflects the judgment of stakeholders who will use the measurement tool (health care professionals and patients) that the instrument appropriately represents what it is intended to measure. Experts and clinicians have long been traditionally consulted on item selection, but the incorporation of patient input into the measurement process is becoming a new standard of validation not found in educational measurement standards [12]. The FDA's guidance imposes the common-sense criteria that a PRO measurement needs to "make sense" to the patients who will be asked to complete the measure and should incorporate symptoms relevant to the disease/treatment to be evaluated [2, 13]. This typically involves patient interviewing and commenting at several steps in the item-development process, a method known as "qualitative research" or "cognitive interviewing." If a new measurement tool is being created, this partially assures that the items and scales are meaningful and understood by patients [14]. If an existing assessment tool is to be used in a study, cognitive debriefing supports the appropriateness of using the tool in that particular study or trial. The FDA guidance recommends that cognitive debriefing studies be included in the medical product's dossier including those of new immunotherapeutic agents to support labeling claims [2].

Convergent validity. Convergent validity indicates whether scores agree with results from a similar-but-independent measure. Convergent-related validity is determined by correlating the new assessment measure with a known "gold standard" for assessing the variable of interest (the symptom). Unfortunately, few gold standards are available for measuring symptoms. Some studies of convergent validity have used previously validated measures of the symptom or symptom-specific subscales from validated HRQOL measures, such as the pain items from the SF-36 or the fatigue subscale of the Profile of Mood States, to estimate measurement convergence.

Known-group validity. Known-group validity refers to the ability of the instrument to differentiate between groups in a predictable way. For example, cancer patients with poor performance status or late-stage disease should demonstrate higher symptom burden on the measurement instrument compared with patients who have good performance status or early-stage disease, respectively. Similarly, patients undergoing aggressive therapy should have higher severity levels of treatment-related symptoms (such as fatigue) later on in their treatment, compared with pretreatment symptom severity.

Sensitivity to Change

Whereas known-group validity is cross-sectional in nature, a measure's sensitivity is assessed repeatedly over the time that symptoms are expected to change. Sensitivity always includes a time component in that changes can be demonstrated in the expected direction. For example, pain severity ratings should improve when

the patient receives appropriate analgesics for pain in a pre-post study design. Similarly, patients undergoing aggressive cancer therapy are expected to experience worsening symptoms as they progress through their treatment regimen, and a symptom assessment tool should be able to detect those expected changes.

Practical Characteristics of a Symptom Measure

In addition to being sensitive to change and having acceptable reliability and validity, an ideal symptom assessment measure should also be brief and easy to complete, so as to reduce patient burden. Conciseness is particularly important if the symptom measure is to be used repeatedly to monitor changes in symptoms over time. A symptom measure must also be easy to understand, preferably written at around fifth grade level so that a patient with poor education can still complete it with minimal assistance. Availability in multiple languages is also important, especially in settings where patients come from different countries and linguistic background. Finally, the scores derived from the measure should be easy to interpret and intuitively meaningful to both patients reporting symptoms and to the clinicians and researchers making decisions about them.

Commonly Used Symptom Assessment Tools

Pain assessment instruments. A measure of pain should reflect (1) important aspects of what a person with pain experiences, and (2) how pain is expected to change as a result of the study to be conducted or the treatment to be administered. These issues have been the focus of a long-standing working group called the Initiative on Methods, Measurement, and Pain Assessment in Clinical Trials (IMMPACT, see www.immpact.org). The collective publications of this working group, available on its Web site, are an important resource for persons planning symptom trials. IMMPACT has specified domains of measurement that should be considered in a clinical pain study, such as pain severity, pain interference, and effects of the treatment on other symptoms, including mood [15]. One single-symptom, multi-item measure that assesses these recommended dimensions is the Brief Pain Inventory (BPI) [16, 17].

Other tools that are commonly used for pain assessment in cancer are the Short-Form McGill Pain Questionnaire (recently revised) [18], the bodily pain subscale of the SF-36 [5], and the EORTC QLQ-C30 pain scale [7].

Fatigue assessment instruments. Fatigue, the most common symptom described by patients with cancer, is endemic during cancer treatment and in advanced disease. Substantial debate is being waged over how best to measure fatigue, which many agree is multidimensional, having physical, mental, and, perhaps, emotional components.

It has been argued that single-item fatigue measures and short single-symptom, multi-item measures are too simplistic to represent the complex construct of fatigue; conversely, measures that attempt to capture the complexity of fatigue have many more items and take longer to complete, making them more burdensome for longitudinal administration than the shorter measures.

The Brief Fatigue Inventory (BFI) [19] is a single-symptom, multi-item measure that evolved from the Brief Pain Inventory. The BFI is useful for rapid assessment of fatigue severity in clinical screening and clinical trials. We developed the BFI along the lines of the BPI and examined its psychometric properties in inpatients and outpatients with cancer and in a comparison sample of community-dwelling adults. As with the BPI, the BFI asks patients to rate their fatigue or tiredness on three items assessing fatigue severity and six items assessing how much fatigue interferes with daily functioning. Although our aim in constructing the BFI was to capture both fatigue severity and interference, several studies have demonstrated that the underlying structure of the BFI items suggests a single dimension underlying all items. This single-factor result for the BFI is consistent with the report of Lai et al. [20] that, on the basis of results from 555 patients with cancer who responded to 72 fatigue items, cancer-related fatigue can be considered unidimensional.

Other single-symptom, multi-item measures for fatigue include the Cancer Fatigue Scale [21], Fatigue Symptom Inventory [22], the FACT fatigue [23], Lee Fatigue Scale [24], Multidimensional Fatigue Inventory [25], the revised Piper Fatigue Scale [26], and the Schwartz Cancer Fatigue Scale [27].

Item banks for individual symptoms. The Patient-Reported Outcomes Measurement Information System (PROMIS) is an NIH-funded initiative tasked with developing a more fluid, yet consistent, measurement system for PROs. PROMIS has developed and continues to test a large bank of items that measure various PROs that allows for efficient, psychometrically robust assessment of PROs in clinical research [1]. PROMIS is using item response theory (IRT) to generate a list of patient self-report questions based on initial cues.

Although the PROMIS measures represent a major advance in the development of PROs because of item banking and its methodical IRT approach, much work remains to be done to provide evidence for the utility of the PRO measures that would lead to clinicians' acceptance of their use.

Item library for adverse events reporting. In order to complement the Common Terminology Criteria for Adverse Events (CTCAE), the US National Cancer Institute contracted work to develop its patient-reported outcomes version (PRO-CTCAE). The validated PRO-CTCAE consists of 124 items reflecting 78 symptomatic adverse events, and each adverse event is assessed relative to one or more attributes, specifically presence or absence, frequency, severity, and/or interference with usual or daily activities [28]. PRO-CTCAE captures a full range of symptomatic treatment effects across a full range of cancer treatment modalities. Frequency, severity, and interference with daily activities are scored using a 0–4 rating scale (i.e., frequency: 0 indicates never, 1 rarely, 2 occasionally, 3 frequently, and 4 almost constantly; severity: 0 indicates none, 1 mild, 2 moderate, 3 severe, and 4 very severe; and interference with

daily activities: 0 indicates not at all, 1 a little bit, 2 somewhat, 3 quite a bit, and 4 very much). The response options for presence or absence are 0 for no or 1 for yes. The recall period for all items is the past 7 days. Intended to complement the CTCAE, the PRO-CTCAE is primarily used to describe and elucidate the toxicity profile of an investigational agent. The PRO-CTCAE has been shown to be feasible to use in large multicenter trials [29] but because the PRO-CTCAE was only recently developed, work remains to be done to determine clinically meaningful differences in PRO-CTCAE scores.

Multisymptom assessment tools. Immunotherapies produce a host of symptoms. An ideal multisymptom assessment tool should include the symptoms that occur most frequently and are most distressing to patients. At the same time, the assessment should be short, easy to understand. Multisymptom inventories can be used to identify symptoms that are prevalent and distressing across various cancers and treatments. For example, the M. D. Anderson Symptom Inventory (MDASI*)* is a brief measure of the severity and impact of cancer-related symptoms regardless of cancer or treatment type [30]. The MDASI was developed on the basis of our previous efforts to assess the severity and interference of single symptoms, including the development of the Brief Pain Inventory and the Brief Fatigue Inventory [16, 19]. The MDASI asks patients to rate the severity of 13 symptoms that are common in patients with cancer once treatment begins: fatigue, disturbed sleep, pain, drowsiness, poor appetite, nausea, vomiting, shortness of breath, numbness, difficulty remembering, dry mouth, distress, and sadness. Patients rate each symptom's presence and greatest severity in the previous 24 h on an 11-point (0–10) scale, with 0 representing "not present" and 10 representing "as bad as you can imagine." The MDASI also contains six items that assess the degree to which symptoms have interfered with aspects of the patient's life in the previous 24 h: general activity, mood, walking ability, normal work (including work outside the home and housework), relations with other people, and enjoyment of life. Each interference item is also rated on an 11-point scale, with 0 signifying "did not interfere" and 10 signifying "interfered completely."

Other most commonly used multisymptom assessment tools include the EORTC QLQ C30 [7], the Rotterdam Symptom Checklist [31], the Symptom Distress Scale [32], the MSAS [33], the ESAS [34], and the symptom monitor [35].

Interpretation of Patient-Reported Symptom Data and Methods of Determining Minimally Important Difference

Widespread use of an instrument depends on how well clinicians and researchers can use and interpret scores derived from the tool. Once a tool's validity has been established, the next step is to determine the instrument's minimal clinically important difference (MCID; or minimally important difference, MID) in symptom scores. With large enough sample sizes, very small differences in symptom ratings

can be statistically significant, yet offer little value to patients and health care providers making treatment decisions. Determining the MCID in the field of health-related quality of life can facilitate the interpretability of symptom scores. Two approaches are used to determine the MCID: distribution-based methods and the anchor-based methods [36]. One approach is not preferred over the other, and one clinical significance consensus panel [37] suggested that the procedures within each method are not sufficient by themselves but are complementary, especially when their respective results are consistent.

Distribution-based methods. Distribution-based methods compare the change in symptom scores seen in a clinical trial to measures of variability in score distributions, such as the standard deviation, the effect size, or the standard error of measurement (SEM). For effect sizes, variability of symptom reports at baseline for all trial patients is typically used. However, estimates of variability can potentially vary from one study to another depending on the heterogeneity of the patient sample.

One approach for the distribution-based method is to set the MCID as one-half standard deviation of the symptom scores at baseline [38, 39]. Cohen's effect-size guidelines, which attach values to the magnitude of an effect, can also be used to aid interpretation of symptom scores [40]. The SEM can be calculated to further minimize the impact of population heterogeneity. This is computed as the baseline standard deviation multiplied by the square root of (1 − the reliability of the symptom scores); for any longitudinal study, either of two estimates of reliabilities, internal consistency and test-retest reliability, can be used. Wyrwich et al. [41] demonstrated that a criterion of 1 SEM was closely related to the anchor-based approach when determining the MCID for the Chronic Respiratory Questionnaire and the Chronic Heart Failure Questionnaire.

Anchor-based methods. As the name implies, this method requires the use of an "anchor," which typically is a question or set of questions designed to compare the patient's judgment of degree of change in a variable (e.g., a rating of health status) that is logically associated with the change. The anchor can either be individual-focused (single anchor) or population-focused (multiple anchors). Both approaches require that the anchor by itself is interpretable and that the anchor is related to symptoms. An example of the single-anchor method might be an item stating, "Compared with your last treatment, how do you rate your symptom now?" with possible response options of "better," "no change," or "worse." The average symptom score that falls into each value of this item constitutes an MCID. This strategy is consistent with approaches used in developing MCIDs for the Chronic Heart Failure Questionnaire [42]. For the multiple-anchor method, this procedure can be extended by using candidate variables such as disease severity, disease progression, response to treatment, or treatment discontinuation.

Using cut points to determine treatment responders. Categorizing symptoms as mild, moderate, or severe may be useful for interpreting clinically significant changes in symptom levels in the clinic and in determining the amount of change that constitutes a response to treatment in a clinical trial. Serlin et al. [43] showed how

cancer "pain at its worst" measured on a 0–10 NRS can be categorized into mild (1–4), moderate (5, 6), or severe (7–10) levels using cut points determined by multivariate analysis of variance. Previous studies have shown that patients whose pain is moderate to severe (i.e., 5 or greater on the 0–10 NRS) report significantly greater pain-related interference with function than do patients with mild or no pain. The derivation of cut points has also been applied to fatigue using the 0–10 NRS scale of the Brief Fatigue Inventory. Several researchers have employed this methodology using "average pain" rather than "pain at its worst" and with non-malignant disease conditions (e.g., diabetic neuropathy [44], low back pain [45]). Cut-point-defined categories such as mild, moderate, and severe are a simple way for clinicians to assess patient symptoms within the practice setting.

This cut-point method can also be used to compare treatment groups in clinical trials [46, 47]. For example, a responder can be defined as a patient whose "pain at its worst" changed from moderate or severe at intake to none or mild at follow-up after an intervention.

Feasibility and Utility of Incorporating PRO in Different Cancer Cohorts

This section discusses the feasibility and added benefit of including PRO objectives specifically the MDASI, presented earlier in this chapter, in evaluating the toxicity of treatment and understanding symptom trajectory over the treatment period. The patient cohorts include lung, hematological and head and neck cancers receiving various cancer treatments.

Symptom severity is predictive of the development of radiation-induced pneumonitis. In a study of 152 patients with non-small cell lung cancer treated with concurrent chemoradiation, the MDASI was administered before the start of chemoradiation and then weekly up to 6 months after therapy was completed. After controlling for the effects of sex, age, and radiation dose/volume, the authors found that increases in the severity levels of shortness of breath and coughing were associated with high-grade radiation-related pneumonitis at 6 months after therapy completion [48]. In short, concurrent chemoradiation therapy for locally advanced non-small cell lung cancer was found to be associated with the development of clinically significant radiation-related pneumonitis.

Symptom severity and symptom interference predict survival in advanced lung cancer. In a study in which we followed 94 patients with advanced-stage non-small cell lung cancer, we collected symptom data with the MDASI before and after the first cycle of chemotherapy [49]. We found that moderate to severe levels of cough (ratings $\geq$5 on a 0–10 scale) at baseline predicted poor overall survival. In addition, increases in fatigue and shortness of breath from baseline to the end of the first chemotherapy cycle predicted poor overall survival. In a separate cohort of patients with advanced-stage non-small cell lung cancer, we found that patient-

reported symptom interference with daily activities, as measured by the MDASI, added prognostic information to Eastern Cooperative Oncology Group performance status and cancer stage in the prediction of overall survival [50].

Symptom burden in hematopoietic stem cell transplantation recipients. We used the blood and marrow transplantation module of the MDASI (i.e., MDASI-Bone Marrow Transplantation) in 192 patients who had undergone hematopoietic stem cell transplantation to assess symptom severity and symptom interference with daily activities. Data were collected at 20 time points from the day of stem cell infusion to 100 days after hematopoietic stem cell transplantation. Symptom severity and symptom interference with daily activities were calculated using the arithmetic average of MDASI-Bone Marrow Transplantation items for symptom severity or symptom interference with daily activities. Those who had acute graft-versus-host disease (GVHD) had higher symptom severity and greater symptom interference with daily activities than patients without GVHD [51]. Symptoms are initially expected to increase but will eventually decrease over time. These changes in symptoms can be reliably and validly measured using MDASI-Bone Marrow Transplantation. It is worth noting the commonality between GVHD and immunotherapy. GVHD is one of the major complications of allogeneic hematopoietic stem cell transplantation [52]. For both GVHD and immunotherapy, symptoms are reported because of the immune response.

We have also shown that long-term collection of symptom data is feasible. In a study of patients with chronic myeloid leukemia, symptoms were assessed via MDASI-Chronic Myeloid Leukemia every 2 weeks for 1 year using an interactive voice response system. Compliance was excellent: 80% of patients completed at least 50% of assessments and 51% of patients completed 80% of the assessments [53].

Symptom burden in patients with head and neck cancer. In a prospective study [54], we examined the pattern of patient-reported symptoms during radiation therapy and concurrent chemotherapy for patients with head and neck cancer so that future symptom interventions and clinical investigations could be more effectively designed. A cohort consisting of 149 patients completed the head and neck module of the MDASI weekly during the course of radiation therapy-based treatment. Overall symptom severity ($p < 0.001$) and symptom interference with daily activities ($p < 0.001$) became progressively worse over the treatment course and were worse for those receiving concurrent chemotherapy ($p < 0.001$). Fatigue, drowsiness, lack of appetite, mouth and throat mucus, and problems tasting food were more severe for those receiving concurrent chemotherapy. By the end of 6–7 weeks of treatment, about 67% of patients experienced high symptom burden. Multivariable analysis showed that low patient baseline performance status and receipt of concurrent chemotherapy were associated with increased symptom burden. In conclusion, the study identified the pattern of both local and systemic symptoms, and the degree of symptom interference with daily activities was temporally distinct, marked by increased magnitudes and shifts in individual symptom rankings, as well as identifiable symptom clusters.

Symptom PRO and Immunotherapies

While there are multiple ongoing clinical trials that are testing the safety and efficacy of immunotherapy either singly or in combination with other forms of therapy, patient-reported symptom data related to new immune-based oncology treatments are lacking. Although a few studies [55, 56] reported HRQOL associated with immunotherapy, symptom-focused PRO is more relevant owing to its proximity to the effects of immunotherapy. A recent study by Bordoni et al. [57] did use the EORTC-QLQ-C30 that includes many symptoms as a PRO measure. However, the frequency of assessments may not lend itself to precise symptom tracking. In Bordoni et al. study, PROs were collected on day 1 of every cycle up to the end of treatment visit. Weekly PRO assessments up until the first restaging may provide useful data. As presented later in this chapter, PRO assessments do not have to coincide with clinic visits but can be accomplished through various modes of administration. This frequent assessment is vital for clinicians because it allows tracking of the patient's ability to tolerate the intended oncologic therapies and allows for improved patient-centered care [58]. Because the FDA is also concerned on how cancer patients feel and function, in addition to prolonging survival of cancer patients, the role of symptom PRO is even more critical in drug development especially for newer immunotherapeutic agents. However, the lack of symptom data collected rather frequently over time for patients undergoing treatment with immunotherapy hinders our understanding of these changes in symptoms and their associated interference with daily functions.

PRO in patients in early-phase trials. In 52 patients with advanced cancer enrolled in a phase I clinical trial of the first-in-human true human monoclonal antibody, MABp1, patients completed the European Organization for Research and Treatment of Cancer Quality of Life Questionnaire, a PRO measure, at three time points over the course of the trial [59]. The PRO measure was able to capture longitudinal changes in symptoms over time. PRO assessments at baseline and week 8 showed significant improvements on day 1 of cycle 3 in social ($p = 0.042$), emotional ($p = 0.032$), and role function scores ($p = 0.006$). Fatigue ($p = 0.0084$), pain ($p = 0.025$), and appetite loss ($p = 0.020$) also improved. Patients reported a significant improvement in global quality-of-life scores, from 4.8 to 5.4 ($p = 0.021$). These results indicate that PRO changes can be observed in patients in phase I clinical trials undergoing treatment with monoclonal antibodies.

In a recent cross-sectional study, George et al. [60] explored symptom patterns and patient clusters based on symptom severity and examined associated factors. The researchers approached 248 patients in phase I clinical trials and only two patients declined to participate. Patients in a phase I clinical trial reported less dyspnea ($p < 0.001$) and vomiting ($p < 0.029$) than did patients who were not enrolled, but the patient groups did not differ in terms of other symptoms. The researchers also assessed the relationships among sleep quality, symptom burden,

and mood in patients with advanced cancer who were enrolled in early-phase clinical trials. Results showed that sleep quality was poor among most patients, and poor sleep was associated with an increased likelihood of high symptom burden and symptom-related interference with daily activities.

Feasibility of obtaining multiple baseline symptom assessments and frequent assessments in patients in phase I clinical trial settings. In a recent study of cancer patients enrolled in phase I clinical trials at MD Anderson [61], 37 patients receiving immunotherapy were assessed daily for about 2 weeks before beginning treatment and twice per week for 4–6 weeks before the end of cycle 2 or disease progression. Patients were given the option to respond on paper, through an interactive voice response system, or electronically through web-based platforms. Most patients preferred responding electronically. With 15 potential maximum baseline assessments, the mean was 10.2 and the standard deviation was 2.8. The median number of baseline assessments was 11 with a mode of 12 from 8 patients. With 22 potential maximum on-treatment assessments, the mean was 11.8, standard deviation 6.1, median 13, and mode 15.

Mode of PRO Administration

With technological advancement, there are many options to collect PRO data. Patient reports can be obtained either via the use of interactive voice response system or various web-based version of data collection. A major advantage of these various options is the ability to collect more frequent and real-time assessment and without having the need for the patient to be in the clinic or hospital. In addition, missing data is minimized which is critically important in longitudinal studies.

Potential Issues in the Incorporation of PRO in Immunotherapy Studies

Issues of practicability, ease of administration, level of patient (assessment) burden, and interpretability are critical factors to consider in considering the use of PRO in immunotherapy studies. Immunotherapy is known to prolong survival in many cases, but the patient's experience and function with this survival benefit is less clear. PRO focusing on symptom burden will improve understanding of the impact of immunotherapy. Many symptom measures are available to suit a variety of needs but require critical thinking about how they will be used. We can ask similar questions to those used for other treatment modalities. Will the treatment reduce symptoms that are present (e.g., shortness of breath in lung cancer) or prevent symptoms normally expected to occur (e.g., neuropathy from certain cancer treatments)? Will the treatment have rapid effects on symptoms, requiring repeated assessments over

a short period, perhaps daily or three times per week? Or will the treatment have more gradual effects on the symptom, such as the pain reduction associated with palliative radiotherapy? If the effects on symptoms are rapid, repeated use of a brief and easily administered symptom measure is probably the best choice, whereas if symptoms change more gradually, assessment should be less frequent and might include additional symptom items.

Selection of symptom items for assessment in immunotherapy poses another challenge. Many symptom measures, including the MDASI, were further improved by including items specific to the disease or treatment. For example, the head and neck module of the MDASI included items such as difficulty swallowing and problems with mouth sores to underscore the nature of the cancer affecting the head and neck region. However, a comprehensive list of symptoms associated with immunotherapy has yet to be uncovered. Although the list of immune-related adverse events provides a good indication of the symptomatic effects of immunotherapy, we need to ask the patients themselves via qualitative interviewing, a well-accepted approach favored by regulatory agencies.

Conclusions

We have discussed how symptom or collectively symptom burden is more proximal to the effect of the disease and treatment compared to the more general health-related quality of life. In developing or even using symptom measures, we need to be cognizant of the desirable properties of a psychometrically valid symptom assessment tool. We reviewed available symptom assessment tools focusing first singly on pain and fatigue and then emphasizing the need for a multisymptom assessment because cancer and its treatment produce multiple symptoms. We described two main methods in deriving minimally important difference, anchor-based and distribution-based methods, to help in the interpretation of PRO data.

We have shown the importance of symptom assessment. It can no longer be argued that we cannot use patient report to represent patients' symptoms with a relatively high degree of precision or to meet the standards of "assay sensitivity" that are expected of standard clinical assessments and laboratory tests. Changes in symptom status as measured by patient report are critical for clinical care and for implementation of clinical guidelines for symptom control. Quality assurance and clinical effectiveness research increasingly demand assessment of symptom status as a representation of what the patient experiences in a clinical trial or clinical encounter.

Finally, we described the utility of incorporating PROs in several cancer cohorts, discussed the current use of PROs in immunotherapy and identify areas where further research is needed to enhance the use of PROs in cancer patients undergoing immunotherapy. With the emergence of immunotherapies, regulatory agencies such as the FDA are increasingly interested not only in prolonging survival of cancer

patients but also in how these patients feel and function while undergoing cancer treatment. Understanding patient's experiences is best accomplished by directly asking them about their symptoms with the use of PRO. Many studies involving the use of immunotherapeutic agents have started to incorporate PRO in the study design. However, many of these studies are still in their infancy. Many issues involved in symptom assessment have yet to be resolved, such as frequency of administration and adequacy of the chosen symptom list to cover both known and unknown effects of immunotherapy. These areas offer a potentially rich agenda for future research.

Conflicts of Interest The author reports no conflict of interest in this work.

References

1. Patient-Reported Outcomes Measurement Information System. Welcome to PROMIS. Available from: http://www.nihpromis.org. Accessed 26 June 2018.
2. US Food and Drug Administration, Center for Drug Evaluation and Research, Center for Biologics Evaluation and Research, Center for Devices and Radiological Health. Guidance for industry. Patient-reported outcome measures: use in medical product development to support labeling claims. Available from: http://www.fda.gov/downloads/Drugs/GuidanceComplianceRegulatoryInformation/Guidances/UCM071975.pdf. Accessed 26 June 2018.
3. Basch E, Reeve BB, Mitchell SA, Clauser SB, Minasian LM, Dueck AC, et al. Development of the National Cancer Institute's Patient-Reported Outcomes Version of the Common Terminology Criteria for Adverse Events (PRO-CTCAE). J Natl Cancer Inst. 2014;106(9). pii: dju244.
4. de Haes JC. Quality of life: conceptual and theoretical considerations. In: Watson M, Greer S, Thomas C, editors. Psychosocial oncology. Oxford: Pergamon Press; 1988. p. 61–70.
5. Ware JE Jr, Sherbourne CD. The MOS 36-item short-form health survey (SF-36). I Conceptual framework and item selection. Med Care. 1992;30(6):473–83.
6. Cella DF, Tulsky DS, Gray G, et al. The Functional Assessment of Cancer Therapy Scale: development and validation of the general measure. J Clin Oncol. 1993;11(3):570–9.
7. Aaronson NK, Ahmedzai S, Bergman B, et al. The European Organization for Research and Treatment of Cancer QLQ-C30: a quality-of-life instrument for use in international clinical trials in oncology. J Natl Cancer Inst. 1993;85(5):365–76.
8. Cleeland CS. Symptom burden: multiple symptoms and their impact as patient-reported outcomes. J Natl Cancer Inst Monogr. 2007;(37):16–21.
9. Crocker LM, Algina J. Introduction to classical and modern test theory. Pacific Grove: Wadsworth Publishing; 2006.
10. Litwin MS. How to measure survey reliability and validity, vol. 7: The survey kit. Thousand Oaks: Sage; 1995.
11. Nunnally JC, Bernstein IH. Psychometric theory, vol. 3. New York: McGraw-Hill; 1994. McGraw-Hill Series in Psychology
12. American Educational Research Association, American Psychological Association, National Council on Measurement in Education, Joint Committee on Standards for Educational and Psychological Testing. Standards for educational and psychological testing. 2nd ed. Washington, DC: American Educational Research Association; 1999.
13. Turner RR, Quittner AL, Parasuraman BM, Kallich JD, Cleeland CS. Patient-reported outcomes: instrument development and selection issues. Value Health. 2007;10(s2):S86–93.

14. Willis GB, Reeve BB, Barofsky I, Invited Paper C. The use of cognitive interviewing techniques in quality of life and patient-reported outcomes assessment. In: Lipscomb J, Gotay CC, Snyder C, editors. Outcomes assessment in cancer: measures, methods and applications. Cambridge: Cambridge University Press; 2004. p. 610–22.
15. Dworkin RH, Turk DC, Farrar JT, et al. Core outcome measures for chronic pain clinical trials: IMMPACT recommendations. Pain. 2005;113(1–2):9–19.
16. Cleeland CS, Ryan KM. Pain assessment: global use of the Brief Pain Inventory. Ann Acad Med Singap. 1994;23(2):129–38.
17. Cleeland CS. Measurement of pain by subjective report. In: Chapman CR, Loeser JD, editors. Issues in pain measurement, vol. 12: Advances in pain research and therapy. New York: Raven Press; 1989. p. 391–403.
18. Dworkin RH, Turk DC, Revicki DA, et al. Development and initial validation of an expanded and revised version of the Short-form McGill Pain Questionnaire (SF-MPQ-2). Pain. 2009;144(1–2):35–42.
19. Mendoza TR, Wang XS, Cleeland CS, et al. The rapid assessment of fatigue severity in cancer patients: use of the Brief Fatigue Inventory. Cancer. 1999;85(5):1186–96.
20. Lai JS, Crane PK, Cella D. Factor analysis techniques for assessing sufficient unidimensionality of cancer related fatigue. Qual Life Res. 2006;15(7):1179–90.
21. Okuyama T, Akechi T, Kugaya A, et al. Development and validation of the cancer fatigue scale: a brief, three-dimensional, self-rating scale for assessment of fatigue in cancer patients. J Pain Symptom Manage. 2000;19(1):5–14.
22. Hann DM, Jacobsen PB, Azzarello LM, et al. Measurement of fatigue in cancer patients: development and validation of the Fatigue Symptom Inventory. Qual Life Res. 1998;7(4):301–10.
23. Cella D. The Functional Assessment of Cancer Therapy-Anemia (FACT-An) Scale: a new tool for the assessment of outcomes in cancer anemia and fatigue. Semin Hematol. 1997;34(3 Suppl 2):13–9.
24. Lee KA, Hicks G, Nino-Murcia G. Validity and reliability of a scale to assess fatigue. Psychiatry Res. 1991;36(3):291–8.
25. Smets EM, Garssen B, Bonke B, de Haes JC. The Multidimensional Fatigue Inventory (MFI) psychometric qualities of an instrument to assess fatigue. J Psychosom Res. 1995; 39(3):315–25.
26. Piper BF, Dibble SL, Dodd MJ, Weiss MC, Slaughter RE, Paul SM. The revised Piper Fatigue Scale: psychometric evaluation in women with breast cancer. Oncol Nurs Forum. 1998;25(4):677–84.
27. Schwartz AL. The Schwartz Cancer Fatigue Scale: testing reliability and validity. Oncol Nurs Forum. 1998;25(4):711–7.
28. Dueck AC, Mendoza TR, Mitchell SA, Reeve BB, Castro KM, Rogak LJ, et al. Validity and reliability of the US National Cancer Institute's Patient-Reported Outcomes Version of the Common Terminology Criteria for Adverse Events (PRO-CTCAE). JAMA Oncol. 2015; 1(8):1051–9.
29. Basch E, Pugh SL, Dueck AC, Mitchell SA, Berk L, Fogh S, et al. Feasibility of patient reporting of symptomatic adverse events via the Patient-Reported Outcomes Version of the Common Terminology Criteria for Adverse Events (PRO-CTCAE) in a Chemoradiotherapy Cooperative Group Multicenter Clinical Trial. Int J Radiat Oncol Biol Phys. 2017;98(2):409–18.
30. Cleeland CS, Mendoza TR, Wang XS, et al. Assessing symptom distress in cancer patients: the M.D. Anderson Symptom Inventory. Cancer. 2000;89(7):1634–46.
31. de Haes JC, van Knippenberg FC, Neijt JP. Measuring psychological and physical distress in cancer patients: structure and application of the Rotterdam Symptom Checklist. Br J Cancer. 1990;62(6):1034–8.
32. McCorkle R, Young K. Development of a symptom distress scale. Cancer Nurs. 1978;1(5):373-378.
33. Portenoy RK, Thaler HT, Kornblith AB, et al. The Memorial Symptom Assessment Scale: an instrument for the evaluation of symptom prevalence, characteristics and distress. Eur J Cancer. 1994;30A(9):1326–36.

34. Bruera E, Kuehn N, Miller MJ, Selmser P, Macmillan K. The Edmonton Symptom Assessment System (ESAS): a simple method for the assessment of palliative care patients. J Palliat Care. 1991;7(2):6–9.
35. Hoekstra J, Bindels PJ, van Duijn NP, Schadé E. The symptom monitor. A diary for monitoring physical symptoms for cancer patients in palliative care: feasibility, reliability and compliance. J Pain Symptom Manage. 2004;27(1):24–35.
36. Revicki D, Hays RD, Cella D, Sloan J. Recommended methods for determining responsiveness and minimally important differences for patient-reported outcomes. J Clin Epidemiol. 2008;61(2):102–9.
37. Guyatt GH, Osoba D, Wu AW, Wyrwich KW, Norman GR. Methods to explain the clinical significance of health status measures. Mayo Clin Proc. 2002;77(4):371–83.
38. 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):582–92.
39. Sloan JA, Dueck A. Issues for statisticians in conducting analyses and translating results for quality of life end points in clinical trials. J Biopharm Stat. 2004;14(1):73–96.
40. Cohen J. Statistical power analysis for the behavioral sciences. 2nd ed. Hillsdale: Lawrence Earlbaum; 1988.
41. Wyrwich KW, Tierney WM, Wolinsky FD. Further evidence supporting an SEM-based criterion for identifying meaningful intra-individual changes in health-related quality of life. J Clin Epidemiol. 1999;52(9):861–73.
42. Guyatt G, Walter S, Norman G. Measuring change over time: assessing the usefulness of evaluative instruments. J Chronic Dis. 1987;40(2):171–8.
43. Serlin RC, Mendoza TR, Nakamura Y, Edwards KR, Cleeland CS. When is cancer pain mild, moderate or severe? Grading pain severity by its interference with function. Pain. 1995;61(2):277–84.
44. Zelman DC, Gore M, Dukes E, Tai KS, Brandenburg N. Validation of a modified version of the brief pain inventory for painful diabetic peripheral neuropathy. J Pain Symptom Manag. 2005;29(4):401–10.
45. Keller S, Bann CM, Dodd SL, Schein J, Mendoza TR, Cleeland CS. Validity of the brief pain inventory for use in documenting the outcomes of patients with noncancer pain. Clin J Pain. 2004;20(5):309–18.
46. Cleeland CS, Portenoy RK, Rue M, et al. Does an oral analgesic protocol improve pain control for patients with cancer? An intergroup study coordinated by the Eastern Cooperative Oncology Group. Ann Oncol. 2005;16(6):972–80.
47. Wong GY, Schroeder DR, Carns PE, et al. Effect of neurolytic celiac plexus block on pain relief, quality of life, and survival in patients with unresectable pancreatic cancer: a randomized controlled trial. JAMA. 2004;291(9):1092–9.
48. Wang XS, Shi Q, Yue J, Chen TY, Xu T, Komaki R, et al. Predictive value of patient-reported outcomes on radiation pneumonitis related normal lung uptake of FDG on PET in patients with non-small cell lung cancer treated with concurrent chemoradiation [abstract 3015]. International Society for Quality of Life Research 23rd annual conference; Copenhagen, Denmark; Oct 19-22, 2016. Qual Life Res. 2016;25(Suppl 1):157–8.
49. Wang XS, Shi Q, Lu C, Basch EM, Johnson VE, Mendoza TR, et al. Prognostic value of symptom burden for overall survival in patients receiving chemotherapy for advanced nonsmall cell lung cancer. Cancer. 2010;116(1):137–45.
50. Barney BJ, Wang XS, Lu C, Liao Z, Johnson VE, Cleeland CS, et al. Prognostic value of patient-reported symptom interference in patients with late-stage lung cancer. Qual Life Res. 2013;22(8):2143–50. PMCID: PMC3724766.
51. Williams LA, Giralt SA, Wang XS, Mobley GM, Mendoza TR, Cohen MZ, et al., editors. Measuring the symptom burden of allogeneic hematopoietic stem cell transplantation in patients with and without acute graft-versus-host disease [abstract 49]. Biol Blood Marrow Transplant. 2009;15(Suppl 2):20–1.
52. Stewart BL, Storer B, Storek J, Deeg HJ, Storb R, Hansen JA, et al. Duration of immunosuppressive treatment for chronic graft-versus-host disease. Blood. 2004;104(12):3501–6.

53. Williams LA, Garcia Gonzalez AG, Ault P, Mendoza TR, Sailors ML, Williams JL, Huang F, Nazha A, Kantarjian HM, Cleeland CS, Cortes JE. Measuring the symptom burden associated with the treatment of chronic myeloid leukemia. Blood. 2013;122(5):641–7.
54. Rosenthal DI, Mendoza TR, Fuller CD, Hutcheson KA, Wang XS, Hanna EY, et al. Patterns of symptom burden during radiotherapy or concurrent chemoradiotherapy for head and neck cancer: a prospective analysis using The University of Texas MD Anderson Cancer Center Symptom Inventory-Head and Neck Module. Cancer. 2014;120(13):1975–84.
55. Long GV, Atkinson V, Ascierto PA, Robert C, Hassel JC, Rutkowski P, et al. Effect of nivolumab on health-related quality of life in patients with treatment-naive advanced melanoma: results from the phase III CheckMate 066 study. Ann Oncol. 2016;27:1940–6.
56. Schadendorf D, Dummer R, Hauschild A, Robert C, Hamid O, Daud A, et al. Health-related quality of life in the randomised KEYNOTE-002 study of pembrolizumab versus chemotherapy in patients with ipilimumab-refractory melanoma. Eur J Cancer. 2016;67:46–54.
57. Bordoni R, Ciardiello F, von Pawel J, Cortinovis D, Karagiannis T, Ballinger M, Sandler A, Yu W, He P, Matheny C, Felizzi F, Rittmeyer A. Patient-reported outcomes in OAK: a phase III study of Atezolizumab versus Docetaxel in advanced non-small-cell lung cancer. Clin Lung Cancer. 2018. pii: S1525-7304(18)30131-1. https://doi.org/10.1016/j.cllc.2018.05.011.
58. Basch E, Iasonos A, McDonough T, Barz A, Culkin A, Kris MG, 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;7:903–9.
59. Hong DS, Hui D, Bruera E, Janku F, Naing A, Falchook GS, Piha-Paul S, Wheler JJ, Fu SQ, Tsimberidou AM, Stecher M, Mohanty P, Simard J, Kurzrock R. MABp1, a first-in-class true human antibody targeting interleukin-1 alpha in refractory cancers: an open-label, phase 1 dose-escalation and expansion study. Lancet Oncol. 2014;15:656–66.
60. George GC, Iwuanyanwu EC, Anderson KO, Yusuf A, Zinner RG, Piha-Paul SA, et al. Sleep quality and its association with fatigue, symptom burden, and mood in patients with advanced cancer in a clinic for early-phase oncology clinical trials. Cancer. 2016;122:3401–9.
61. Mendoza TR, George G, Williams L, Shi Q, Naing A, Hong DS, et al, editors. Feasibility and added value of multiple baseline symptom assessments in early-phase clinical trials [abstract]. Presented at the International Society for Quality of Life Research (ISOQOL) 24th annual conference, Philadelphia; 2017.