Introduction to Quality of Life in Drug Development

21

Bellinda L. King-Kallimanis, Lee Jones, and Lynn Howie

Contents

B. L. King-Kallimanis (✉)
Patient-Focused Research Center, LUNGevity
Foundation, Bethesda, MD, USA
e-mail: bking-kallimanis@lungevity.org

L. Jones
Arlington, VA, USA

L. Howie
Asheville, NC, USA

21.1 Introduction

In developing anti-cancer therapies, the gold standard question clinical trials have historically sought to answer is: *what is the impact of the experimental therapy on patients' overall survival?* However, as sponsors have looked toward bringing new therapies to patients more quickly, this has translated into more frequent use of sur-

rogate endpoints as the primary clinical trial endpoint. A surrogate endpoint is defined as "an endpoint that is used in clinical trials as a substitute for a direct measure for how a patient feels, functions or survives" [1]. In other words, surrogate endpoints should reliably predict clinically meaningful effects. One of the most frequently used surrogate endpoints in oncology is progression-free survival (PFS). The concern with the use of PFS is that the relationship between PFS and overall survival, the clinical endpoint PFS is a surrogate for, is variable [2]. While overall survival is straightforward to capture, interpretation of the results can be complicated by crossover trial design, and in cancers with long natural histories, trials are expensive and can take decades to complete. This has led to increasingly stronger calls by oncologists and patient advocates to better understand *"feels and functions"* via patients' self-reported quality of life (QoL) to better assess the impact and clinical benefit of the therapy for patients and potentially identify issues with therapy toxicities [3].

Both the US Food and Drug Administration (FDA) and the European Medicines Agency (EMA) have provided guidance to industry on incorporating the patient voice in clinical trials. In 2006, the FDA published a draft guidance to industry on the use of patient-reported outcomes (PRO) in clinical trials; after public comment, this document became a final guidance in 2009 [4], and while a series of new guidances are under development [5], the 2009 guidance, at the time of writing, remains the reference document to industry for the FDA. It is important to note that the FDA guidance documents are not regulations and are therefore nonbinding recommendations; however, these documents do describe the current thinking at the FDA on that particular topic. They also provide a road map to help drug developers navigate a particular topic to ultimately gain licensure for their products. Around the same time as the FDA draft PRO guidance was published, the EMA published a reflection paper on the regulatory guidelines for use of health-related quality of life (HRQL) measures in the evaluation of medical products [6]. Subsequently, the EMA published an appendix to the Guidelines on the

Evaluation of Anti-Cancer Medicinal Products in Man to address the use of PROs specifically in cancer clinical trials in 2013 [7]. The FDA 2009 PRO guidance focuses primarily on assessing the measurement properties of PRO instruments. Sponsors can use this guidance to develop their PRO strategy and provide appropriate evidence to regulators that the instrument(s) included in their clinical trial is reliable, valid, and sensitive to change over time for the target population. The EMA guideline appendix for anti-cancer medicinal products, on the other hand, focuses on endpoints and considerations related to PROs. For example, the guideline cautions "careful thought must go into designing and implementing PRO measures in the oncology clinical trial setting in order to investigate a well-formulated predefined hypothesis" and notes that there is no standard approach. Despite the different focuses, this EMA advice is, for example, in line with the FDA's frequent comment to come and discuss PRO endpoints with the Agency early.

In the regulatory context, the broad umbrella term of PROs is used to describe "a measurement that comes directly from the patient about the status of their health condition without amendment or interpretation of the response by a clinician or anyone else" [1]. While PROs and the concepts of QoL and HRQL are terms that are sometimes used interchangeably, the terms describe different concepts from a regulatory perspective. Broadly speaking, both HRQL and QoL are multidimensional concepts that aim to capture a person's assessment of their well-being, though HRQL dimensions are focused on a person's QoL using a health lens. In the EMA 2005 reflection paper, HRQL, within the drug development paradigm, is defined as "patient's subjective perception of the impact of his disease and its treatment(s) on his daily life, physical, psychological and social functioning and well-being" [6]. The FDA defines HRQL as "a multidomain concept that represents the patient's general perception of the effect of illness and treatment on physical, psychological, and social aspects of life" [4]. Using an example, a patient who reports how bad their pain is on a 0–10 numerical rating scale is providing a response on a PRO measure.

If pain severity on this 11-point numerical rating scale is the only PRO assessed in the clinical trial, this would be insufficient to understand patients' HRQL because multiple domains related to HRQL must be measured in order to report on how a treatment might have influenced patients' HRQL.

Regulatory advice from the FDA, EMA, and groups such as SPIRIT-PRO [8], the PROTEUS consortium [9], and SISAQOL [10] have provided recommendations and clear guidance that PROs should be treated similarly to other outcomes of interest in clinical trials. In this chapter, we aim to bring these resources all together to describe how PRO and HRQL data can be used to inform regulatory assessment of new therapies. This will include the considerations that go into clearly defined endpoints that could be used to assess efficacy or safety and ultimately end up in the product label. We will describe how the use and applicability of these data may vary with respect to disease setting. We will review commonly drawn conclusions with respect to HRQL-related endpoints in cancer clinical trials literature and discuss why some of these conclusions are problematic. We provide both a patient and a clinician perspective and discuss how real-word data might help fill a gap of efficacy and effectiveness, as well as safety.

This chapter will enable the reader to (a) identify key guidance and guideline documents for use of PRO data in cancer clinical trials; (b) know what are key concepts of interest in drug development; (c) recognize differences in how PRO data are used by different regulatory agencies; (d) understand how missing PRO data can influence the interpretation of PRO results from cancer clinical trials; and (e) hear both a patient and a clinician perspective in relation to PRO measures and the use of the data captured.

21.2 PRO Measures in Drug Labeling

Historically, the FDA and EMA have used different criteria to determine what patient-reported data will be included in their drug label. As there

are multiple factors that can affect a person's conception of HQRL, the FDA asks that sponsors focus on concepts that are proximal to the drug effects, specifically of the drugs' ability to control disease as well as the adverse effects. For the FDA Oncology Center of Excellence (OCE), the concepts that are considered most proximal to the drug effect and that are broadly applicable across all types of cancers and therapies include (1) physical function, (2) disease symptoms, and (3) side effects and the impact of side effects (e.g., bother) (Fig. 21.1). It is recognized by the FDA OCE that distal concepts like social functioning and emotional well-being are important to patients, and possibly other stakeholders. However, when assessing the benefit-risk profile of an investigational therapy, there are non-therapy factors (e.g., satisfaction with care, family relationships) that contribute to these more distal concepts, which is why the results regarding these concepts are given less weight in the overall regulatory assessment [11, 12]. The notion of proximal and distal concepts was initially illustrated in the Wilson and Cleary model. This conceptual model of patient outcomes integrates both bio-medical and HRQL outcomes by describing five levels containing specific health concepts: (1) biological/physiological factors, (2) symptoms, (3) functional status, (4) general health perceptions, and (5) HRQL [13]. Health concepts 2 and 3 reflect where the OCE places their focus for PRO data. This is because the concepts falling under these broad headings have greater proximity to the disease and treatment of that disease. This is then ultimately reflected in what PRO label claims have been included by the FDA in the US prescribing information (i.e., the drug label). The EMA, on the other hand, has included the more distal and broader concept of HRQL in their drug labels for certain products (i.e., summary of product characteristics (SmPC)). The EMA has suggested that where the treatment is intended to be palliative as opposed to curative, the "focus of care is on promoting and preserving quality of life" [12]. The EMA advises that "in order to approve a global claim that a product 'improves HRQL,' it would be necessary to demonstrate robust improvement in
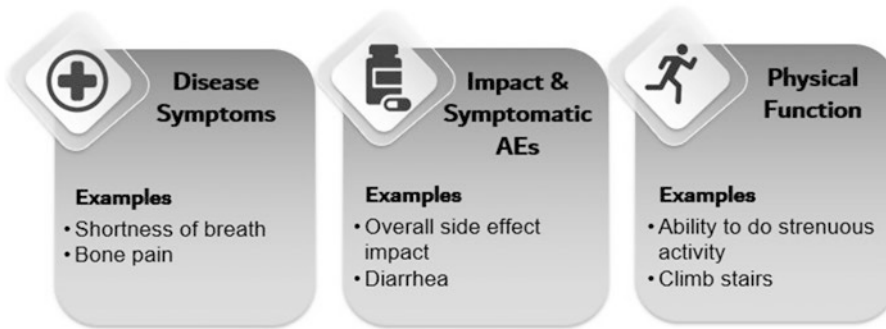
Fig. 21.1 Core Concepts of Interest to the US FDA Oncology Center of Excellence in Assessment of the Benefit-Risk of Investigational Therapies [15]

all or most of these domains" [6]. In line with this, in the new PFDD discussion document for guidance 3, the FDA wrote "For example, if improvement in a score for a multi-domain concept (e.g., symptoms associated with a certain condition) is driven by a single responsive item (e.g., pain intensity improvement) whereas other important items (e.g., other symptoms) did not show a response, a general claim about the multi-domain concept (e.g., improvements in symptoms associated with the condition) cannot be supported" [14].

More recently, the FDA has been encouraged via legislation (the 2012 Safety and Innovation Act [16] and in 2016 the twenty-first Century Cures Act [17]) to build on patient-focused drug development and include the patient experience in the benefit-risk assessment of new therapies when it has been collected, even when the data informs only exploratory endpoints. The FDA Office of Oncologic Diseases (OOD) has been successful in incorporating the patient experience into their reviews. As presented by Gnanasakthy, when there was patient experience data submitted as part of a New Drug Application (NDA) or a Biologics License Application (BLA), it was incorporated into the OOD's reviews 100% of the time since the twenty-first Century Cures Act was enacted [18]. However, there has been no change in the number of labeling claims based on PRO data since the introduction of the Cures Act. This is mainly because the trials that have read out their results since the Cures Act went into effect

were designed at least 3–5 years prior to the legislation. This meant the PRO strategy was not prioritized, e.g., not included in the statistical hierarchy, for achieving a labeling claim.

In a published review of the inclusion of PRO claims in oncology drug labels, it was reported that of the 45 indications that included PRO data in the clinical trials, there were no oncology drugs that included PRO data in the US prescribing information between 2012 and 2106. This review, however, overlooked the approval of certinib [19] in 2014 and did not review label updates, which lead to exclusion of crizotinib, which received regular approval in 2013 without PRO data included in the label. However, an efficacy labeling change in 2015 lead to the inclusion of PRO data [20], highlighting how challenging it can be to track this information. The current US prescribing information includes PRO results for both these drugs. On the other hand, for the EMA it was found that 21 (47%) SmPCs where results from the analysis of the PRO data were included. As evidenced from the respective agencies' guidance documents this is to be expected as there are differences in the focus on how PRO data is incorporated into the benefit-risk assessment by the FDA and the EMA [21].

An example of the differences in how the FDA and EMA use PRO data in the label can be seen with the drug, ceritinib (Zykadia), approved for patients with metastatic ALK-positive non-small-cell lung cancer. In Table 21.1 the language from the FDA and EMA labels is presented

Table 21.1 Labeling Claim Language for Ceritinib (Zykadia)

Regulatory body	Year approved	Labeling language
US FDA [19]	2014	Exploratory analyses of patient-reported outcome measures suggested a delay in time to development of or worsening of "shortness of breath" in patients treated with ZYKADIA as compared to chemotherapy. The patient-reported delay in onset or worsening of "shortness of breath" may be an overestimation because patients were not blinded to treatment assignment.
EMA [23]	2015	Patient-reported outcome questionnaires (Lung cancer symptom scale [LCSS], EORTC-QLQ-C30 [C30], EORTC QLQ-LC13 [LC13], and EQ-5D-5L) were completed by 80% or more of patients in the ceritinib and chemotherapy arms for all questionnaires at most of the time-points during the course of the study. Ceritinib significantly prolonged time to deterioration for the pre-specified lung cancer-specific symptoms of interest of cough, pain, and dyspnea (composite endpoint LCSS: HR = 0.61, 95% CI: 0.41, 0.90, median time to deterioration [TTD] NE [95% CI: 20.9, NE] in the ceritinib arm versus 18.4 months [13.9, NE] in the chemotherapy arm; LC13: HR = 0.48, 95% CI: 0.34, 0.69, median TTD 23.6 months [95% CI: 20.7, NE] in the ceritinib arm versus 12.6 months [95% CI: 8.9, 14.9] in the chemotherapy arm). Patients receiving ceritinib showed significant improvements over chemotherapy in general Quality of Life and global Health Status measures (LCSS [$p < 0.001$], QLQ-C30 [$p < 0.001$], and EQ-5D-5L index [$p < 0.001$]).

(Table 21.1). In the US prescribing information from the FDA, the description of the results is limited in detail (e.g., no primary measures of interest such as point estimates, confidence intervals, or p-values). The FDA label also highlights that the analyses conducted were exploratory and may even be biased because of the trial design. The results presented focus on delay of onset or worsening of the symptom "shortness of breath," fitting with the use of concepts that are proximal to the drug effect. The description is also consistent with the advice provided by the FDA regarding inclusion of multiple endpoints, such that no point estimates are provided from *exploratory analyses*. Broadly speaking, the FDA, in their multiple endpoints' guidance, suggests that for an endpoint to be considered for inclusion in the drug label, the endpoint needs to be included in the endpoint hierarchy (i.e., prespecified and with multiplicity adjusted for). This is to overcome Type 1 errors, or in other words, false-positive findings [22]. Exceptions have been made to include exploratory analyses such as the current example for ceritinib, but the details presented in the drug label are generally limited. In the case of ceritinib, the information provided on "shortness of breath" comes from two randomized clinical

trials. In both trials, the same conclusion regarding "shortness of breath" was drawn and the results were considered not to be a false-positive finding and therefore included descriptively in the US prescribing information.

On the other hand, the EMA included in their SmPC the point estimates, confidence intervals, and p-values. These results came from the delay of onset analyses, where the dependent variables were worsening of the symptom composite score from the Lung Cancer Symptom Scale as well as a composite score from the European Organisation for Research and Treatment of Cancer, lung module (EORTC QLQ-LC-13). In addition, in the EMA SmPC, improved QoL was reported for patients treated in the ceritinib arm.

The results presented in the FDA and EMA ceritinib label are not even from the same models described differently; the results are from completely different analyses. In the SmPC, the results are from time to event models, where the dependent variables are composite scores. For example, SmPC include the concepts of cough, pain, and dyspnea, whereas the results presented in the FDA label only address the concept of "shortness of breath." Though the names of the questionnaires are not provided in the FDA label,

both the LC13 and LCSS questionnaires include items that measure "shortness of breath"; therefore, the results could be either from instrument or from both with the same trend in the results. The EMA labeling text does not specifically address time to deterioration in the concept of "shortness of breath." The results are for composite scores, and from the SmPC alone, it is not possible to know whether cough, pain, and dyspnea were all improved in similar magnitude the treatment arm, as is suggested in the EMAs reflection paper on HRQL [6].

There is no single way to approach the inclusion of PRO results in a drug label though it could be argued that neither of these examples for ceritinib are ideal for health care providers and patients. While there are a few reasons for this, an important one is the result of there being limited standardization for PRO endpoints; with standardization comes the ability to summarize findings briefly. It is hard to imagine how this PRO information would be conveyed by a clinician to a patient. In the US prescribing information, there is no information on how long shortness of breath was delayed. In the SmPC, there is no information on whether all the symptoms in the composite were delayed or whether one or two of the symptoms led to increased delay. Later in the chapter we present a template for thinking about a standardized presentation of patient-reported symptom data and discuss the FDA OCEs pilot Project Patient Voice [24].

Examples of PRO Data Supporting Approval There are two examples in the US where patient-reported information was considered a marker of how patients feel, function and survive, and were part of the primary support for regulatory approval. In 1996, gemcitabine (Gemzar) was approved for "the first-line treatment of patients with advanced (nonresectable Stage II or Stage III) or metastatic (Stage IV) adenocarcinoma of the pancreas." In the pivotal trial, the primary endpoint was "clinical benefit response," a composite endpoint, which was defined by the trial sponsors as "based on analgesic consumption, *pain intensity,* performance status and weight change." More specifically,

patients were considered to have a response if they "showed a ≥50% reduction in pain intensity (Memorial Pain Assessment Card) or analgesic consumption, or a 20-point or greater improvement in performance status (Karnofsky Performance Scale) for a period of at least 4 consecutive weeks, without showing any sustained worsening in any of the other parameters OR the patient was stable on all of the aforementioned parameters and showed a marked, sustained weight gain (≥7% increase maintained for ≥4 weeks) not due to fluid accumulation." The FDA reviewers acknowledged that "the clinical benefit endpoint measured in this study are "published and recognized as valid, reproducible, and reliable…"" [25]. However, this was the only time this novel endpoint was used for regulatory decision making.

The other example is for ruxolitinib (Jakafi), which was approved for the treatment of patients with intermediate- or high-risk myelofibrosis, including primary myelofibrosis, post-polycythemia vera myelofibrosis, and post-essential thrombocythemia myelofibrosis. The FDA decision was based on the reduction of both spleen volume and the six-item PRO measure total score of disease-related symptoms. The endpoint was defined as "The proportion of subjects who have a 50% reduction from baseline to Week 24 in the total symptom score" using the Myelofibrosis Symptom Assessment Form version 2 (MFSAF v2.0). The FDA noted in their review summary that this improvement is "potentially a direct measure of clinical benefit" and concluded that "These endpoints provide evidence of both a biologic effect of ruxolitinib and a direct patient benefit" [26].

Each of these clinical trials illustrate that there is potential for patient-reported information to support regular approval of new anti-cancer therapies. Use of PRO data was planned during the design and development of both studies. In the case of ruxolitinib, the sponsors requested a special protocol assessment, which led to the FDA agreement that the novel endpoint proposed in the protocol would be acceptable for consideration of approval. For PRO data, and really any

data collected during a clinical trial to be meaningful in the benefit-risk assessment of a new therapy, careful forethought is required to ensure that the design will answer the intended research question.

21.3 Efficacy Vs Safety/Tolerability

The benefit-risk assessment of any new therapy recognizes there is, at times, a tradeoff between increased therapeutic benefit and increased risk of adverse events/toxicity, which is especially true in the evaluation of new oncology therapies. If the risk is acceptable given the benefit (i.e., the primary endpoint was met and the safety profile acceptable) of a new therapy, the therapy is approved. Data capturing the patient experience while on the clinical trial can be used in cancer drug development to answer questions about therapeutic benefit by way of efficacy hypotheses (e.g., ruxolitinib (Jakafi). The results are then presented in Sect. 14 Clinical Studies of US prescribing information) or questions about risk with respect to symptomatic adverse events (e.g., crizotinib (Xalkori), results presented in Sect. 6 Adverse Reactions of the US prescribing information) and tolerability.

In all advanced oncology trials, there is a place for the use of PROs to assess tolerability of the new therapy from the patient perspective because many common adverse events are unobservable (e.g., fatigue, nausea), making patient report a reliable means to understand these symptomatic effects [27]. The analysis of this data will likely be descriptive in nature, and care should be taken in the selection of an appropriate number of items. For example, while the National Cancer Institute's PRO Common Terminology Criteria for Adverse Events (PRO-CTCAE) [28] measurement system includes 124 items representing 78 symptomatic toxicities, the inclusion of all these items in a single trial is neither necessary nor good practice. As not all these items are needed in a single clinical trial, sponsors can work to identify a set of items that strike a balance between capturing relevant symptoms,

avoiding ascertainment bias, and not over burdening trial participants. This can be achieved by using the free text option, and software is available where dropdown options populate with terms from the PRO-CTCAE library as well as MedDRA Lowest Level Terms [29]. The FDA OCE Excellence launched in 2020 a pilot project, Project Patient Voice, to provide a Web-based platform for healthcare providers to look at patient-reported symptom data collected from cancer clinical trials in order to discuss them at the point of care with patients and their caregivers [24]. The plan is to make this an option to cancer clinical trial sponsors to present their trial data when they have rigorously collected patient-reported symptom data. Efficacy endpoints, on the other hand, must be included in the endpoint hierarchy to be fully described in the US prescribing information. In a review of 25 lung cancer clinical trials used to support FDA drug approval between January 2008 and December 2017, no PRO endpoints were included in the efficacy hierarchy where type I error is controlled for [30].

Whether assessing an efficacy or safety research question, the objective and endpoint should be clearly described in the study protocol [31]. Also, the assessment frequency of a valid and reliable PRO measure should be appropriate for the endpoint. For example, if the treatment administration is intravenous infusion once every 28 days, asking patients to report their side effects over the past 7 days on day 1 of a cycle (i.e., 28 days after their last infusion) is unlikely to provide a realistic snapshot of the acute side effects that were experienced by patients. By day 1 of a new cycle, most side effects will have resolved. The most relevant time to ask may be around 5–7 days post-infusion, which would provide the most information for a safety/tolerability endpoint. However, typically the capture of PRO measures is tied to clinic visits, primarily to improve completion rates. This tradeoff between completion and optimal timing of the concept must be weighed, though electronic PRO measurement could in theory overcome the tying of assessments to clinic visits and can be done well, it is not without its own set of challenges [32, 33].

For example, if using the patient's own device, sometimes referred to as "bring your own device," there may be storage issues or updates to the operating system that can impact how PRO data is collected on the patients' own device that will require careful planning in the protocol.

21.4 What QoL Results Are Reported in the Literature

Primary clinical trial manuscripts describing the results of cancer clinical trials rarely include PRO results; however, there may be another manuscript published to describe the findings from the PRO data. In a literature review of PRO-focused manuscripts published between January 1, 2017, and December 31, 2018, it was found that while 93% of the papers reviewed included a PRO-related endpoint, only 33% tested a specific directional hypothesis [34]. In a systematic review of breast cancer clinical trial manuscripts published between January 2001 and October 2017 reporting PRO data, the majority of papers reviewed included a PRO endpoint. However, only 12% of these papers reported testing a directional hypothesis. The authors make an important point that the lack of a clear hypothesis can lead to the use of different analytic techniques that have the potential to lead to different conclusions. A clear research hypothesis helps in all stages from trial design to data analysis and finally to interpretation and translation of the results [8].

The results of PRO/HRQL analyses are often translated to a broad conclusion of no or small differences in HRQL or functioning between the clinical trial arms despite observing notable differential toxicity. An example of such a conclusion from a phase III randomized clinical trial of men with metastatic castration-resistant prostate cancer stated "mean changes from baseline in the FACT-P subscales were similar in both treatment groups, indicating that the addition of apalutamide to androgen deprivation therapy did not result in a decrease in HRQOL" [35]. This example is not intended to call out these particular authors, as Merzoug et al. found that 73% of the papers they reviewed came to the conclusion that the HRQL concepts assessed in the investigational arm were either better or the same as in the control arm [34]. In other words, the majority of the published conclusions reviewed had similar statements that study results favored the treatment arm or suggested equivalence between the control and treatment arms.

These findings could be related to a publication bias where only positive findings are accepted for publication. But there is also a methodological challenge here. Specifically, the challenge with conclusions indicating no difference or similar scores is that most clinical trials are not designed to test what is more formally referred to as an equivalence or non-inferiority hypothesis with respect to the PRO data [36]. What the authors are actually reporting is the absence of an effect or that the null hypothesis cannot be rejected. However, in trials that aimed to test superiority hypotheses (i.e., the investigational treatment is significantly and clinically better than the control arm treatment), we can only say that there *may be* no difference between the arms or that we did not have sufficient evidence to detect the difference when the test does not indicate superiority. There are several issues that arise in cancer clinical trials that must be considered and factored into the analysis and interpretation of absence of effect findings.

Two serious issues affecting the analysis and interpretation of PRO data are missing data and asymptomatic withdrawal. Missing data in cancer clinical trials is common. There can be missing items (i.e., items that a patient skipped) or missing assessments (i.e., the patient did not complete the PRO assessment and therefore no items were completed). Missing assessments are important to assessing data quality, and if not presented in the clinical study report, the FDA will likely send an information request to obtain the completion rates. Completion, in most trials, is defined as the proportion of on-study participants who were scheduled to complete a PRO assessment and filled in at least one question. While prevention of missing data is the best strategy, two low-burden actions that can be taken to improve interpretation in the face of missing data

were suggested in 1998 by Bernhard et al. [37]. First, collection of the reason for missing data helps researchers determine the mechanism of the missing data. For example, the EORTC uses the following reasons for missing assessments: patient felt too ill; clinician or nurse felt the patient was too ill; patient felt it was inconvenient or took too much time; patient felt it was a violation of privacy; patient did not understand the actual language or was illiterate; administrative failure to distribute the questionnaire; not required at this time point; other, specify; and unknown [38]. The other issue is that all clinical study reports could include the answers to the following three questions:

1. How many missing data were there?
2. Why were the data missing?
3. How might the missing data affect the interpretation of the results? [37]

Answering these three questions helps contextualize the PRO data findings. For example, if by month 3, only 60% of trial participants on either arm completed their PRO assessment, the generalizability of the results is limited. When the driver for missing assessments is sicker patients, this will likely lead to an overestimation of HRQL. Understanding why data are missing would further help regulators incorporate PRO findings into their benefit-risk assessment.

With asymptomatic withdrawal, it could be that in both arms 95% of participants who were scheduled to complete a PRO assessment did so, but that by month 6, only 30% of those randomized to the control arm remained on-treatment, whereas 70% of those in the treatment arm were on-treatment. This is problematic because in many trials PRO data collection stops when treatment ends. If PRO data collection continues post-treatment, it is often collected at less frequent intervals than while on study treatment and the quality of the data may be low (e.g., low completion rates). Asymptomatic withdrawal can introduce bias because there is only PRO data from the patients who were able to tolerate the control arm treatment and they remained on trial and the patients who experienced side effects or whose

disease progressed withdrew earlier, and therefore, no PRO data was collected in the post-treatment epoch. This means that the PRO data is not missing at random [39]. One way to potentially mitigate this bias would be to pick a relevant time point in the treatment course where all patients complete a PRO assessment regardless of whether they remain on treatment or not and prioritize collection of that data.

Another important element for overcoming interpretation issues is pre-specification of well-defined PRO endpoints. In trials where PRO data is collected, the associated endpoint is not often detailed, for example, a frequently used endpoint is that *PRO data will be examined between the arms* [8, 40]. The Standard Protocol Items: Recommendations for Interventional Trials in Patient Reported Outcomes (SPIRIT-PRO) recommends that "Primary, secondary, and other outcomes, include the specific measurement variable (e.g., systolic blood pressure), analysis metric (e.g., change from baseline, final value, time to event), method of aggregation (e.g., median, proportion), and time point for each outcome" are included in the study protocol [31]. The largest barrier to this recommendation is that, as mentioned earlier, there are no standardized PRO endpoints for all cancer clinical trials. However, applying the estimand framework can help trial sponsors to structure their endpoints, including their PRO-specific endpoints. The estimand framework has been proposed by the International Council for Harmonisation and outlined in the E9(R1) addendum [41]. A detailed description of this framework is beyond the scope of this book chapter; however, the broad goal of the E9(R1) addendum is to align trial objectives, design, analysis, and interpretation. Finally, there is an ongoing multi-stakeholder project, Setting International Standards in Analyzing Patient-Reported Outcomes and Quality of Life Endpoints Data (SISAQOL) Consortium, that is aiming "to develop recommendations for standardizing the analysis and interpretation of patient reported outcomes and quality of life data in cancer randomized trials" [42]. This initiative includes regulatory agencies, payers, trialists, industry, academia, and most importantly patients, with the intended result

of standards developed using existing guidances and guidelines to help with the design of appropriate patient-centric endpoints as well as help to translate findings so that clinicians and patients can make sense of the results and use the results in shared decision making.

21.5 Disease and Treatment Context Matters

Most of the examples provided thus far have been trials that have supported approval of new treatments in the advanced stages of cancer. For many patients with early-stage cancer, there are few noticeable symptoms and diagnosis is made via screening efforts or due to clinical investigations related to another medical issue. On the other hand, patients with advanced disease may experience a greater number of disease-related symptoms. Therefore, just as we see disease-free survival, and not overall survival, used as a primary clinical endpoint in adjuvant trials, the PRO endpoints need to be different. For example, it may be reasonable in a trial investigating a new treatment for metastatic castrate-resistant prostate cancer to use a PRO endpoint where time to pain palliation is investigated [43]. This is because for there to be pain palliation, patients must start the trial with a certain degree of pain (usually >3 points on a 0–10 numerical pain rating scale) [44] and therefore baseline pain should be included in the inclusion criteria. In the adjuvant setting where patients are unlikely experiencing pain before treatment, it would not be possible to recruit patients into the trial. Patient-centric endpoints in the early-stage setting are an area that is continuing to develop. What remains the same though for both settings is understanding safety and tolerability of the investigational treatment.

21.6 Patient Perspective – Lee Jones

PROs are becoming more expected to be measured and reported in the clinical trial component of drug development. This is due on part to the requirements for "beneficence" in clinical trials, but also due to the importance of QoL considerations for patients on clinical trials as well as in post-approval clinical care.

The relationship between PROs and QoL is not always easy to determine. QoL is totally patient-centric, no two patients will consider the exact same experiences when asked to rate their QoL. This is because every patient is different in terms of sex at birth, gender identity, age, body structure, racial and ethnic background, genetic profile, and economic background among others. As a result, they will react differently to drug treatments clinically, emotionally, and intellectually. Clinical side effects can range from inconvenience to death. Emotional side effects can range from calm acceptance to clinical depression. Intellectual side effects can range from stoic acceptance to obsession. These differing reactions can result in differing pain thresholds and ability to accept and withstand whatever side effects they may be experiencing and will have a major impact on patients' real experience of symptoms and side effects, and their perceived impact on QoL. For example, diarrhea might be an inconvenience for a retired patient, but for a stage performer, it could dramatically affect their ability to work and thus negatively impact their QoL.

Patients will also differ in their short- and long-term objectives regarding their treatment. One patient may want to experience no treatment side effects, another may be willing to do anything to be able to live until their son's or daughter's wedding, and another may be willing to suffer anything for the best chance of long-term survival.

As a result of these differences, defining "quality of life" in a way that would apply to all or even most patients is very difficult. Most of what is measured today and that affect treatment decisions are clinical outcomes (e.g., laboratory values) for which the healthcare establishment has determined thresholds that are used to define "tolerability." This is even less relevant to many patients since clinical trials do not enroll patients that represent every combination of these individual characteristics so only when the drug is

approved for use in the real world is the real "testing" conducted.

Despite these considerations, QoL is a critical endpoint in the drug development process. Though the results will not be definitive and applicable to every patient, giving patients the range and scope of the factors that affect QoL will offer some comfort if and when they experience any of these same effects. Ultimately, it may be possible to give patients a "Chinese menu" of treatment options, with varying efficacies and side effects, so each patient can, in a shared decision-making process with their doctor, choose the treatment that will best take into consideration both the clinical effects of the drug and the feelings, goals, and needs of the patient. We have fleshed out a hypothetical example at the end of this section.

It is also likely that different data presentations of PRO/QoL concepts could be used, one set as part of the regulatory process, to measure the statistical difference between study arms and another set for patient decision making, where a different focus might be important, and the presentation of the data quite different. The former is primarily quantitative, the latter primarily descriptive and much more effectively presented visually so that patients do not need to understand statistics, for example, hazard ratios and 95% confidence intervals. An example of this might be peripheral neuropathy. For regulatory purposes, the CTCAE grade is important and how the proportions between treatment arms differ. However, for patients the grade may be less important, but knowing the length of time they might experience the symptom may be more significant—an intense, short-term bout may be of less concern than a milder but longer-term experience which might have a greater impact on their QoL.

One initiative underway that leads in the direction of presenting descriptive information is being undertaken by the US FDA. This initiative, called "Project Patient Voice," will show, using easy-to-understand graphics, the side effects reported by participants in clinical trials in terms of both timing and intensity of the effect [24]. Though currently limited to a demonstration of

the approach, this initiative has the promise of offering patients the most realistic picture of what they might expect to experience when treated with the drug. In this way, each patient, in consultation with their oncologist, will be able to determine what combination of factors can result in the best (or least bad) side effects based on their unique set of attributes and perspectives. The process is still overly complicated to be able to be used by most patients and to be most useful to patients it would need to include information about patient characteristics, such as age, race, comorbidities, and tumor mutations as well as drug data related to efficacy, physical function, and PROs, so that a patient could better assess the effects of a drug on a "patient like me." This would become a massive database management and data collection, retrieval, and presentation issue that might be best handled with an artificial intelligence application.

Cancer patients need a better way to understand how the drugs available to treat their cancer will affect them, their cancer, and their QoL. Capturing PROs is a critical first step but the massive amount of data that is collected needs to be effectively managed and reported in a form that patients can understand and use in consultation with their oncologist to determine the best course of treatment for them. This would indeed make the promise of personalized medicine a reality.

21.6.1 Menu Presentation

In the face of a changing treatment landscape that has potential for multiple treatment options, understanding the tradeoffs between different side-effect profiles in light of efficacy findings would be useful for patients and healthcare providers. One could imagine a guide outlining benefits and risks of the approved treatment options next to each other for review as a shared decision-making tool (Fig. 21.2). Information regarding the patient's disease, including actionable mutations and biomarker information, could be fed in via a series of questions and this would pull from a database the relevant treatment options based
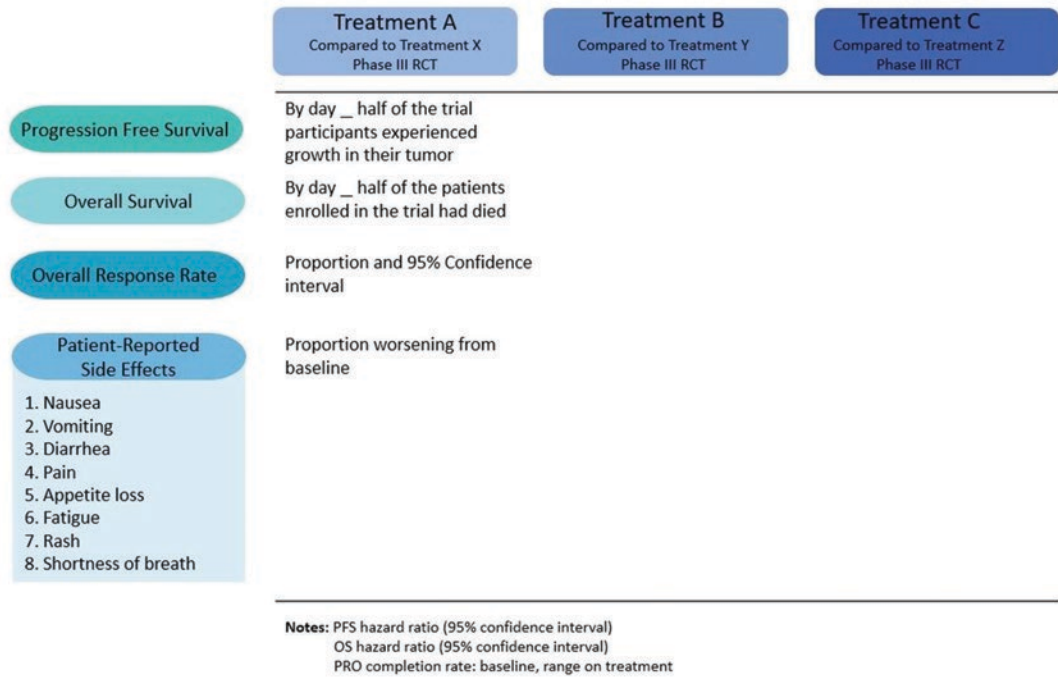
Fig. 21.2 Aspirational Menu Presentation of Clinical Trial Information

on the National Comprehensive Cancer Network Clinical Practice Guidelines.

The figure and description of our hypothetical shared decision-making tool is aspirational, and not currently possible to populate. Before such a tool can be developed, there are many challenges to overcome. However, one possible starting point is to leverage the data presented on the FDA's Project Patient Voice website once more trials are added. The symptom summary information presented in the table (worsening in symptoms from baseline assessment), as well as information on overall survival, PFS, and overall response rate (ORR) from the clinical trial, could be used to populate a tool like that presented in Fig. 21.2.

There are several limitations in relying solely on clinical trial data that need to be considered. For example, not all trials collect the same side-effect data, and this would leave gaps in the table because it might not be relevant to ask about hair loss in a trial comparing two tyrosine kinase inhibitors which are not known to cause hair loss. There is, however, a core set of side effects (anorexia, anxiety, cognitive disturbance, constipation, depres-

sion, diarrhea, dyspnea, fatigue, insomnia, nausea, neuropathy, and pain) that was arrived at via an NCI-supported consensus that could be routinely captured [45] but requires guidance from the regulatory agencies to be used more extensively. There are also challenges in comparisons of trial data. This is because the trial data can differ due to differences in trial inclusion and exclusion criteria. How these limitations would be incorporated as well as differences in the length of follow-up or missing PRO data need to be considered and a balance struck between sufficient description and too much description that could lead to difficulty to understand the important take away points. Clinical trial data is also not necessarily representative of the wider range of patients receiving treatment in the community. To overcome this, the table could be augmented with real-world data; however, at this time, PROs systematically capturing side effects are not commonplace in healthcare systems. Finally, the hosting and maintenance of such a tool is critical, and who should take on this role and how any related costs should be allocated are not clear.

But what is clear is that patients would benefit significantly by having a full range of efficacy and side-effect information so that together with their healthcare providers they could choose a treatment that best accords with their personal QoL and healthcare preferences.

21.7 Clinician Perspective – Lynn Howie

Patient-reported outcome measures can improve the data needed for clinicians and their patients to decide between therapies when disease-related outcomes are similar and there is no clear therapy that is substantially superior with respect to disease-related outcomes. Currently, we have very limited patient-reported data in FDA labels; however, as noted earlier, there are some key examples where this data has helped to inform the severity and duration of symptoms. Ruxolitinib, an agent for patients with myelofibrosis, was approved using a composite endpoint that included a radiographic endpoint of reduction in spleen size along with a reduction in patient-reported assessment of symptom burden as the primary efficacy endpoint for approval. Figure 21.3 is from the label describing the symptom reduction observed at week 24 [26] (Fig. 21.3). From these results, clinicians can advise patients that about half of the patients who receive ruxolitinib report that their symptoms are reduced by about one half after being on therapy for approximately 6 months. Crizotinib, an oral tyrosine kinase therapy for those patients with advanced lung cancer which has an ALK or ROS-1 mutation, is associated with ocular toxicities which can have a significant impact on patient function and QoL. In both examples, PRO data were used to characterize the frequency, duration, and impact of symptoms on patients' daily lives which can then be used to communicate benefit as with ruxolitinib and risk with crizotinib.

In choosing a therapy, patients and clinicians are interested in the side effects of treatment and how these will impact daily life. As we know, daily persistent symptoms can be more aggravating than more severe symptoms that are shorter in duration [46]. For patients who are continuing to work during treatment, it will be important to understand the impact of therapies on this aspect of their lives, as well as the impact on other daily activities such as exercise, ability to perform household tasks such as cooking and eating meals, and patient-reported experiences with symptomatic adverse events. So, questions that assess the impact on these areas will be most useful as patients and clinicians work to identify the best treatment for that patient when several options are reasonable.

Currently, we do not fully understand the patient experience of side effects and we even less so understand the impact on physical function and role function. We need to encourage drug manufacturers to include assessment of symptomatic adverse events and assessment of treatment impact on physical and role function in order to better understand the effect of therapy on patients' lives. This will help to provide patients and clinicians the data needed to make treatment decisions. In the current landscape of global clinical trials, it will also be important to understand how patient responses may be affected by the social and economic structures of the place where the patient lives. In geographical locations where there are robust social insurance programs that allow for the person to have job and/or economic security despite being unable to perform their job due to illness, the impact of side effects may be reported differently than in those places where the inability to perform job and other functions can have a more significant impact on patients' experiences.

21.8 The Future – What Role Can Real-World Data Play in Closing the Efficacy/Effectiveness Gap?

Both patients and clinicians are looking for representative data to help their patients make informed treatment choices. One path to that is
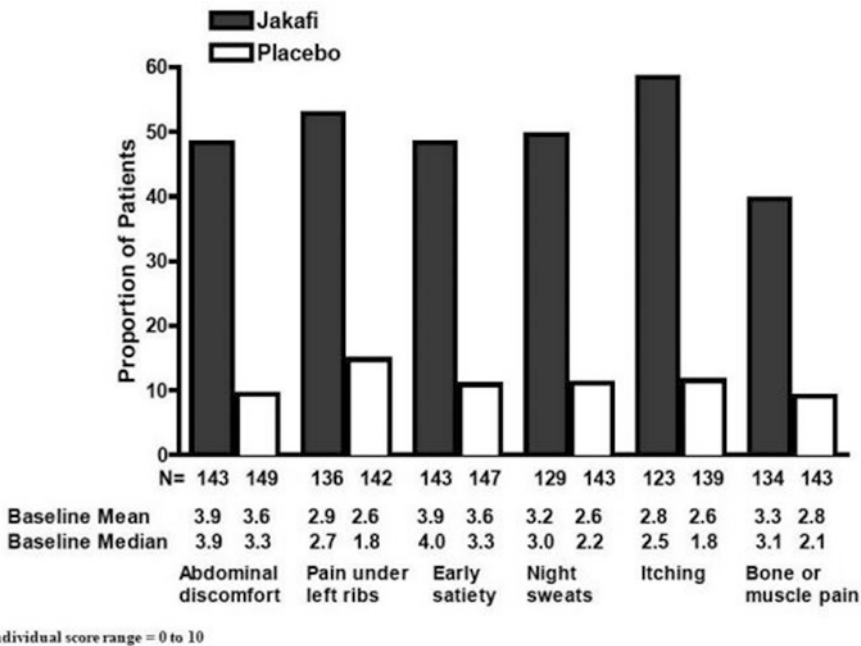
Individual score range = 0 to 10

Fig. 21.3 Proportion of Patients with Myelofibrosis Achieving 50% or Greater Reduction in Individual Symptom Scores at Week 24

via the use of real-world data (RWD). This has been defined as "the data relating to patient health status and/or the delivery of health care routinely collected from a variety of sources. RWD can come from a number of sources, for example: electronic health records, claims and billing activities, product and disease registries, patient-generated data including in home-use settings, data gathered from other sources that can inform on health status, such as mobile devices" [47].

We are currently sitting at the forefront of the possibilities of real-world PRO data. This is because, at the moment, widespread implementation of routine collection of PRO measures in clinical practice is limited, which in turn limits the use of RWD for PROs. In a systematic review of the literature, the authors found that only 3 of 36 articles reviewed reported on implementation of PRO measures in clinical practice with the goal of managing patient care; the majority of papers reviewed were interven-

tions that were carried out in clinical practice and used PROs to assess the success of the intervention [48]. This review may not reflect the true situation, as it is likely that more data is being collected than is reported in the academic literature. However, the collection of RWD that can be converted into real-world evidence (RWE) to support regulatory decision making and possibly close the efficacy/effectiveness gap starts with high-quality data collected in the clinic. Assessing the quality of that data and sharing of best practices is critical. The International Society of Quality of Life (ISOQOL) guidelines present some of the barriers to implementation into the clinic. These include resources, both procurement of equipment (e.g., tablet for electronic capture) and person power (e.g., establishing and sustaining the program). Beyond these challenges, other difficulties include standardization of collection of data and lack of best practices around the analysis and interpretation of the data.

To gain traction with RWD for PROs, straightforward questions and hypotheses are needed. RWD that describes the safety/tolerability of a new therapy may have the most immediate benefit, as these data can be used to better describe patient-reported side-effect experiences by subgroups (e.g., older age) of patients that look more like the patients regularly seen in the clinic. Also, many of the PRO projects currently center around symptom monitoring [48], meaning that there is existing infrastructure in place to capture this data. One of the issues that will need to be reconciled around symptom data collection for drug development is real-time monitoring versus passive data capture. Currently in industry-sponsored clinical trials, almost all PRO data collection is passively collected and not actively reviewed by the care team in real time. This is not always clear to patients enrolled in clinical trials [49]. However, PRO data captured to actively monitor and manage symptoms during routine cancer treatment has been shown to improve overall survival [50, 51]. Acknowledging the impact active monitoring may have will be an important consideration in the use of RWD that may be used to generate RWE.

21.9 Conclusion

In this chapter, we have touched upon many important issues for the inclusion of PRO measures to represent the patient's perspective in drug development and how that data can be applied in clinical practice. Many of the guidelines outlined within this chapter should not be taken to be prescriptive. Each study requires consideration of the specific treatment or study population and what research questions help inform the benefit-risk assessment of a new therapy. However, with careful planning of PRO endpoints, the results are interpretable and meaningful to all stakeholders, but especially to those who have been diagnosed with cancer and want to make informed choices.

21.10 Questions That Can Be Used for Learning/Testing

- When planning a trial that will be part of a licensing application, what patient-reported concepts are most relevant and why?
- What are the key considerations for timing of patient-reported assessments when planning the schedule of assessments?
- If planning to include a PRO label claim, what are the key considerations for the inclusion of PRO data in the drug label?

21.11 A Topic for Discussion That can Be Used for Teaching

- What are the possible implications for reporting different PRO results in the US prescribing information and the European summary of product characteristics?

21.12 Further Reading List

The following list presents literature that extends the contents of this chapter. Readers looking for in-depth information and further material are advised to consult the following sources.

1. US Food and Drug Administration. Guidance for industry use in medical product development to support labeling claims guidance for industry. Clin Fed Regist. 2009;(12):1–39.
2. European Medicines Agency. Appendix 2 to the guideline on the evaluation of anticancer medicinal products in man. 2014;44(4):1–18. Available from: www.ema.europa.eu/contact
3. Kluetz PG, O'Connor DJ, Soltys K. Incorporating the patient experience into regulatory decision making in the USA, Europe, and Canada. Lancet Oncol. 2018;19(5):e267–74.
4. Calvert M, Kyte D, Mercieca-Bebber R, Slade A, Chan AW, King MT. Guidelines for inclusion of patient-reported outcomes in clinical trial protocols the spirit-pro extension. JAMA. 2018;319(5):483–94.

21.13 Research in Context

An important manuscript where the authors reviewed the advanced breast cancer randomized clinical trial literature between January 2001 and October 2017, with the aim to examine the types of analyses that were used for the PRO data collected and published in this peer-reviewed literature. The authors' search led them to review 66 papers. From these papers, it was determined that only 12% of papers presented a predefined directional hypothesis that they set out to test with the analyses conducted. Over half of the papers (58%) investigated multiple domains from the questionnaires used, though only 16% used a statistical adjustment to correct for multiple testing. Nearly a quarter (23%) of papers presented a p-value, indicating some types of comparative analyses were conducted, but did not report the type of analyses that were used to obtain the p-value(s). Most papers (73%) did not report how missing data were handled, which is critical as missing data is a key issue when analyzing PRO data from randomized clinical trials. Completion rates at baseline were presented for 47% of papers, and for the period where patients were on study, only 29% of papers included completion rates. Pe et al. provide the following example of how missing a hypothesis, one of the most fundamental steps of conducting a clinical trial, can impact the results: "if a study aimed to measure HRQL changes over a 6-week period, a cross-sectional HRQL analysis at 6 weeks is not equivalent to an area under the curve analysis within the same timeframe; in fact, these two analytical techniques could yield different results." Because there are no standards with how PRO data are analyzed and reported from clinical trials, the results from this study are not surprising. However, this work was carried out as a part of the

Setting International Standards in Analyzing Patient-Reported Outcomes and Quality of Life Endpoints Data for Cancer Clinical Trials (SISAQOL) consortium, which will address this exact problem over the coming years [8].

References

1. FDA-NIH Biomarker Working Group. BEST (Biomarkers, EndpointS, and other Tools) Resource. Silver Spring, MD; 2016.
2. Amir E, Seruga B, Kwong R, Tannock IF, Ocaña A. Poor correlation between progression-free and overall survival in modern clinical trials: are composite endpoints the answer? Eur J Cancer. 2012;48(3):385–8.
3. Gyawali B, Hey SP, Kesselheim AS. Assessment of the clinical benefit of cancer drugs receiving accelerated approval. JAMA Intern Med. 2019;179(7):906–13.
4. US Food and Drug Administration. Guidance for industry use in medical product development to support labeling claims guidance for industry. Clin Fed Regist. 2009;(12):1–39.
5. US Food and Drug Administration. FDA patient-focused drug development guidance series for enhancing the incorporation of the patient's voice in medical product development and regulatory decision making [Internet]. [cited 2021 Feb 28]. Available from: https://www.fda.gov/drugs/development-approval-process-drugs/fda-patient-focused-drug-development-guidance-series-enhancing-incorporation-patients-voice-medical.
6. European Medicines Agency. Reflection paper on the regulatory guidance for the use of health- related quality of life (HRQL) measures in the evaluation of medicinal products. Reproduction. 2005;(1):1–5.
7. European Medicines Agency. Appendix 2 to the guideline on the evaluation of anticancer medicinal products in man. 2014;44(4):1–18. Available from: www.ema.europa.eu/contact.
8. Pe M, Dorme L, Coens C, Basch E, Calvert M, Campbell A, et al. Statistical analysis of patient-reported outcome data in randomised controlled trials of locally advanced and metastatic breast cancer: a systematic review. Lancet Oncol. 2018;19(9):e459–69.
9. The PROTEUS Consortium [Internet]. 2020 [cited 2020 Dec 15]. Available from: www.theproteusconsortium.org.
10. European Organisation for Research and Treatment of Cancer. Setting International Standard in Analyzing Patient-Reported Outcomes and Quality of Life Endpoints [Internet]. [cited 2021 Mar 26]. Available from: https://event.eortc.org/sisaqol/.

11. Kluetz PG, Slagle A, Papadopoulos EJ, Johnson LL, Donoghue M, Kwitkowski VE, et al. Focusing on core patient-reported outcomes in cancer clinical trials: symptomatic adverse events, physical function, and disease-related symptoms. Clin Cancer Res. 2016;22(7):1553–8.

12. Kluetz PG, O'Connor DJ, Soltys K. Incorporating the patient experience into regulatory decision making in the USA, Europe, and Canada. Lancet Oncol. 2018;19(5):e267–74.

13. Wilson IB, Paul D. Linking clinical variables with health-related quality of life. A conceptual model of patient outcomes. JAMA. 1995;273:59–65.

14. US Food and Drug Administration. Methods to identify what is important to patients select, develop or modify fit-for-purpose clinical outcomes assessments. Patient-Focused Drug Dev Guid Public Work Oct 15–16 [Internet]. 2018; Available from: https://www.fda.gov/downloads/Drugs/NewsEvents/UCM620708.pdf.

15. King-Kallimanis BL. Incorportaitng relevant and meaningful patient reported outcomes. In: 17th international kidney cancer symposium. Miami, FL; 2018.

16. Food and Drug Administration Safety and Innovation Act, S 3187, 112th Congress [Internet]. 2012 [cited 2020 Dec 20]. Available from: https://www.gpo.gov/fdsys/pkg/PLAW-112publ144/pdf%0A/PLAW-112publ144.pdf.

17. 21st Century Cures Act.114th Congress [Internet]. 2016 [cited 2020 Dec 20]. Available from: https://www.congress.gov/114/plaws/publ255/PLAW-114publ255.

18. Gnanasakthy A. Multi-stakeholder perspectives on the patient experience data summary section in new product approvals: current state & future considerations. In: International society for pharmacoeconomics and outcomes research. Virtual; 2020.

19. Novartis Pharmaceuticals Corporation. Zykadia® (ceritinib) [package insert] [Internet]. US Food and Drug Administration website. 2019 [cited 2021 Dec 20]. Available from: https://www.accessdata.fda.gov/drugsatfda_docs/label/2017/205755s010lbl.pdf.

20. Pfizer Inc. Xalkori® (crizotinib) [package insert] [Internet]. U.S Food and Drug Administration website. [cited 2020 Dec 20]. Available from: https://www.accessdata.fda.gov/drugsatfda_docs/label/2021/202570s030lbl.pdf.

21. Gnanasakthy A, Barrett A, Evans E, D'Alessio D, Romano C. A review of patient-reported outcomes labeling for oncology drugs approved by the FDA and the EMA (2012–2016). Value Health. 2019;22(2):203–9.

22. US Food and Drug Administration. Draft guidance to industry: multiple endpoints in clinical trials. US Food and Drug Administration; 2017.

23. Novartis Pharmaceuticals Corporation. Summary of product characteristics: Zykadia® (ceritinib) [Internet]. 2012. Available from: https://www.ema.europa.eu/en/documents/product-information/zykadia-epar-product-information_en.pdf.

24. US Food and Drug Administration. Project patient voice [Internet]. 2020 [cited 2021 Mar 3]. Available from: https://www.fda.gov/about-fda/oncology-center-excellence/project-patient-voice.

25. Center for Drug Evaluation and Research. Gemzar® (gemcitibine hydrochloride) [orginal review] [Internet]. 1996 [cited 2021 Jan 10]. Available from: https://www.accessdata.fda.gov/drugsatfda_docs/nda/pre96/020509Orig1s000rev.pdf.

26. Deisseroth A, Kaminskas E, Grillo J, Chen W, Saber H, Lu HL, et al. US Food and Drug Administration approval: ruxolitinib for the treatment of patients with intermediate and high-risk myelofibrosis. Clin Cancer Res. 2012;18(12):3212–7.

27. Basch E, Jia X, Heller G, Barz A, Sit L, Fruscione M, et al. Adverse symptom event reporting by patients vs clinicians: relationships with clinical outcomes. J Natl Cancer Inst. 2009;101(23):1624–32.

28. United States National Cancer Institute. Patient-reported outcomes version of the Common Terminology Criteria for Adverse Events (PRO-CTCAE™). [Internet]. [cited 2021 Feb 14]. Available from: https://healthcaredelivery.cancer.gov/pro-ctcae/.

29. Chung AE, Shoenbill K, Mitchell SA, Dueck AC, Schrag D, Bruner DW, et al. Patient free text reporting of symptomatic adverse events in cancer clinical research using the National Cancer Institute's Patient-Reported Outcomes version of the Common Terminology Criteria for Adverse Events (PRO-CTCAE). J Am Med Inform Assoc. 2019;26(4):276–85.

30. Fiero MH, Roydhouse JK, Vallejo J, King-Kallimanis BL, Kluetz PG, Sridhara R. US Food and Drug Administration review of statistical analysis of patient-reported outcomes in lung cancer clinical trials approved between January, 2008, and December, 2017. Lancet Oncol. 2019;20(10):E582–9.

31. Calvert M, Kyte D, Mercieca-Bebber R, Slade A, Chan AW, King MT. Guidelines for inclusion of patient-reported outcomes in clinical trial protocols the spirit-pro extension. J Am Med Assoc. 2018;319(5):483–94.

32. Austin E, LeRouge C, Hartzler AL, Chung AE, Segal C, Lavallee DC. Opportunities and challenges to advance the use of electronic patient-reported outcomes in clinical care: a report from AMIA workshop proceedings. JAMIA Open. 2019;2(4):407–10.

33. Coons SJ, Eremenco S, Lundy JJ, O'Donohoe P, O'Gorman H, Malizia W. Capturing patient-reported outcome (PRO) data electronically: the past, present, and promise of ePRO measurement in clinical trials. Patient. 2015;8(4):301–9.

34. Merzoug L, King-Kallimanis B, Bhatnagar V, Kanapuru B, Kim J, Li X, et al. Reporting of patient-reported outcomes and quality of life data in oncology drug development trials: a literature review. In: Quality of life research; 2019. p. S111–2.

35. Saad F, Cella D, Basch E, Hadaschik BA, Mainwaring PN, Oudard S, et al. Effect of apalutamide on health-related quality of life in patients with non-metastatic castration-resistant prostate cancer: an analysis of the SPARTAN randomised, placebo-controlled, phase 3 trial. Lancet Oncol. 2018;19(10):1404–16.

36. US Food and Drug Administration. Guidance for industry: non-inferiority clinical trials to establish effectiveness [Internet]. November, 2016. [cited 2021 Mar 4]. Available from: https://www.fda.gov/media/78504/download.

37. Bernhard J, Cella DF, Coates AS, Fallowfield L, Ganz PA, Moinpour CM, et al. Missing data workshop, 1–3 July, 1996: missing quality of life data in cancer clinical trials: serious problems and challenges. Stat Med. 1998;17(5–7):517–32.

38. European Organisation for Research and Treatment of Cancer. EORTC Quality of Life FAQ [Internet]. [cited 2021 Mar 3]. Available from: https://qol.eortc.org/faq-category/scoring/.

39. Fairclough DL, Peterson HF, Chang V. Why are missing quality of life data a problem in clinical trials of cancer therapy? Stat Med. 1998;17(5–7):667–77.

40. King-Kallimanis BL, Johnson LL. Systematically defining research objectives and framing questions using the estimand framework. In: FDA-ASCO Public Workshop: 2019 Clinical Outcome Assessments in Cancer Clinical Trials Fourth Annual Workshop. Silver Spring, Maryland; 2019.

41. International Concil for Harmonisation (ICH). Addendum on estimands and sensitivity analysis in clinical trials to the guideline on statistical principles for clinical trials E9(R1). Fed Regist. 2019;9(9)

42. European Organisation for Research and Treatment of Cancer. Setting International Standards in Analyzing Patient-Reported Outcomes and Quality of Life Endpoints Data (SISAQOL) Consortium [Internet]. Available from: https://event.eortc.org/sisaqol/.

43. Basch E, Trentacosti AM, Burke LB, Kwitkowski V, Kane RC, Autio KA, et al. Pain palliation measurement in cancer clinical trials: the US Food and Drug Administration perspective. Cancer. 2014;120(5):761–7.

44. Scher HI, Morris MJ, Stadler WM, Higano C, Basch E, Fizazi K, et al. Trial design and objectives for castration-resistant prostate cancer: updated recommendations from the prostate cancer clinical trials working group 3. J Clin Oncol. 2016;34(12):1402–18.

45. Reeve BB, Mitchell SA, Dueck AC, Basch E, Cella D, Reilly CM, et al. Recommended patient-reported core set of symptoms to measure in adult cancer treatment trials. J Natl Cancer Inst. 2014;106(7):dju129.

46. Basch EM, Campbell A, Hudgens S, Jones L, King-Kallimanis BL, Kluetz P, et al. Broadening the definition of tolerability in cancer clinical trials to better measure the patient experience. Friends Cancer Res. 2018;10

47. US Food and Drug Administration. Real-World Evidence [Internet]. [cited 2021 Mar 15]. Available from: https://www.fda.gov/science-research/science-and-research-special-topics/real-world-evidence.

48. Anatchkova M, Donelson SM, Skalicky AM, McHorney CA, Jagun D, Whiteley J. Exploring the implementation of patient-reported outcome measures in cancer care: need for more real-world evidence results in the peer reviewed literature. J Patient Rep Outcomes. 2018;2(1):64.

49. Basu Roy U, King-Kallimanis BL, Kluetz PG, Selig W, Ferris A. Learning from patients: reflections on use of patient-reported outcomes in lung cancer trials. J Thorac Oncol. 2018;13(12):1815–7.

50. Basch E, Deal AM, Dueck AC, Scher HI, Kris MG, Hudis C, et al. Overall survival results of a trial assessing patient-reported outcomes for symptom monitoring during routine cancer treatment. J Am Med Assoc. 2017;318(2):197–8.

51. Denis F, Lethrosne C, Pourel N, Molinier O, Pointreau Y, Domont J, et al. Randomized trial comparing a web-mediated follow-up with routine surveillance in lung cancer patients. J Natl Cancer Inst. 2017;109(9):21–2.