Patient-Reported Outcomes in Clinical Trials Leading to Cancer Immunotherapy Drug Approvals From 2011 to 2018: A Systematic Review

Houssein Safa, MD,[1,†] Monica Tamil, MD,[2,†] Philippe E. Spiess, MD,[3] Brandon Manley, MD,[3] Julio Pow-Sang (iD), MD,[3] Scott M. Gilbert, MD,[3] Firas Safa, MD,[4] Brian D. Gonzalez (iD), PhD,[5] Laura B. Oswald (iD), PhD,[5] Adele Semaan, PhD,[6] Adi Diab, MD,[1] Jad Chahoud, MD, MPH[3,*]

[1]Department of Melanoma Medical Oncology, The University of Texas MD Anderson Cancer Center, Houston, TX, USA; [2]Department of Internal Medicine, The University of Texas Health Science Center at Houston, Houston, TX, USA; [3]Department of Genitourinary Oncology, H. Lee Moffitt Cancer Center and Research Institute, Tampa, FL, USA; [4]John W. Deming Department of Medicine, Section of Hematology and Medical Oncology, Tulane University, New Orleans, LA, USA; [5]Department of Health Outcomes and Behavior, H. Lee Moffitt Cancer Center and Research Institute, Tampa, FL, USA and [6]Participant Research, Interventions, and Measurements Core, H. Lee Moffitt Cancer Center and Research Institute, Tampa, FL, USA

†Authors contributed equally.

*Correspondence to: Jad Chahoud, MD, MPH, Department of Genitourinary Oncology, H. Lee Moffitt Cancer Center, 12902 Magnolia Dr, Tampa, FL 33612, USA (e-mail: jad.chahoud@moffitt.org).

Abstract

Background: Patient-reported outcomes (PROs) promote patient centeredness in clinical trials; however, in the field of rapidly emerging and clinically impressive immunotherapy, data on PROs are limited. **Methods:** We systematically identified all immunotherapy approvals from 2011 through 2018 and assessed the analytic tools and reporting quality of associated PRO reports. For randomized clinical trials (RCTs), we developed a novel 24-point scoring scale: the PRO Endpoints Analysis Score based on 24 criteria derived from the recommendations of the Setting International Standards in Analyzing Patient-Reported Outcomes and Quality of Life Endpoints Data Consortium. **Results:** We assessed 44 trial publications supporting 42 immunotherapy approvals. PROs were published for 21 of the 44 (47.7%) trial publications. Twenty-three trials (52.3%) were RCTs and 21 (47.7%) pertained to single-arm trials. The median time between primary clinical outcomes publications and their corresponding secondary PRO publications was 19 months (interquartile range = 9-29 months). Of the 21 PRO reports, 4 (19.0%) reported a specific hypothesis, and most (85.7%) used descriptive statistics. Three (3 of 21 [14.3%]) studies performed a control for type I error. As for RCTs, 14 of 23 (60.9%) published PRO data, including 13 (56.5%) that published a secondary dedicated manuscript. One-half of these 14 trials scored less than 13 points on the 24-point PRO Endpoints Analysis Score. The mean score was 12.71 (range = 5-17, SD = 3.71), and none met all the recommendations of the Setting International Standards in Analyzing Patient-Reported Outcomes and Quality of Life Endpoints Data Consortium. **Conclusions:** Suboptimal reporting of PROs occurs regularly in cancer immunotherapy trials. Increased efforts are needed to maximize the value of these data in cancer immunotherapy development and approval.

Immunotherapies, specifically immune checkpoint inhibitors and chimeric antigen receptor T-cell therapy, have received broad attention, with multiple Food and Drug Administration (FDA) approvals in recent years (1-4). Interest in these therapies derives from their ability to generate durable and complete responses and to improve overall survival across multiple types of cancers (5,6). Given their efficacy, immunotherapy drugs are being investigated in an increasing number of clinical trials. In fact, the number of immunotherapy drugs in active development increased from 2030 in 2017 to 3876 in 2019, which is a 91.0% increase over 2 years (7).

Patient-reported outcomes (PROs) are defined by the FDA as patient-centered, self-reported questionnaires completed by the patients without any subsequent interpretations by clinicians or other providers (8). These tools aim to involve the patients in the treatment process by allowing them to express

their perceptions of the value of the treatment regimen (8-10). Following a medical intervention, PROs measure outcomes and effects of the intervention on 1 or more pertinent clinical measures, including health-related quality of life (HRQoL) and symptoms (8-10). These tools augment the physician's ability to measure the impact of a drug beyond its survival benefit (9).

Despite their clinical promise, immunotherapy drugs are associated with serious adverse events (AEs) (11,12). Several reports have suggested that many AEs are best reported by patients themselves rather than by clinicians and that including PRO data in clinical trials is a fundamental step in ensuring accurate reporting of AEs (13-15). Given the important role of PROs in treatment decisions, efforts are needed to maximize the clinical utility of PRO data by standardizing the design and the analytic and reporting methods used to deliver these data (16-18).

Systematic reviews have identified heterogeneity and several areas of deficiency in the reporting of PROs across different cancer treatment types (19-22). Some reviews highlighted the deficiencies in standards of PRO protocol content, revealing the common omission of rudimentary PRO design elements such as "rationale for PRO assessment," "PRO hypothesis and objectives," "PRO analysis plan," and "approaches to handle missing PRO data" (23,24). Many efforts have been under way to improve the state of PROs in cancer clinical trials and to offer additional tools to accurately detect these data. A leading example of these efforts is the expansion of the scope of the Common Terminology of Criteria for Adverse Events (CTCAE) to include measures for PROs (25,26). The review of the 790 AEs listed in the CTCAE identified 78 items that are amenable to patient self-reporting and were deemed eligible for the development of the PRO-CTCAE (25). These tools were introduced with the purpose of improving the precision and the comprehensiveness of PRO data and are being increasingly used in cancer treatment clinical trials (27,28). Recently, the Standard Protocol Items: Recommendations for Interventional Trials (SPIRIT-PRO) released recommendations aiming to standardize the inclusion of items that are relevant to PROs in the protocols of clinical trials in which PROs are primary or key secondary endpoints (29). Similarly, the Consolidated Standards of Reporting Trials (CONSORT-PRO) (30) recommendations focused on standardizing high-quality reporting of PROs in the final study reports. Most recently, the Standards in Analyzing Patient-Reported Outcomes and Quality of Life (SISAQOL) Consortium published recommendations in 2020 that focus on guiding and standardizing the methodology used for the design and analysis of PRO endpoints in cancer randomized clinical trials (RCTs) (31). Some studies have focused on evaluating separate specific characteristics of PRO data in oncology trials submitted to the US FDA, such as statistical analysis, completion rates, and PRO assessment frequency (32-35). However, to the best of our knowledge, no study has been conducted so far to assess the state of PROs in FDA-approved immunotherapy drugs specifically since the first approval in 2011. With the aim of providing insight into the quality of PROs in immunotherapy era, we conducted this study to assess the trends in the reporting, standardization, and quality of PRO endpoints for immunotherapy RCTs and single-arm clinical trials that led to corresponding FDA approvals between 2011 and 2018. We hope to provide an objective review identifying the strengths and limitations of current PRO data in immunotherapy, further enabling decision makers and trialists to better interpret and design PROs in the future.

Methods

Search Strategy and Selection Criteria

We conducted a systematic review of the FDA Office of Hematology and Oncology Products archives and created a list of hematology and oncology drug approvals issued between January 2011 and December 2018. Two authors (H.S. and M.T.) independently identified all immunotherapy drugs approved for cancer treatment. Immunotherapy drugs were defined as immune checkpoint inhibitors, chimeric antigen receptor T-cell therapy, and nonspecific immunotherapies (ie, interferons). For each immunotherapy drug, we identified the pivotal clinical trial(s) used as the primary basis for FDA approval and retrieved them from PubMed and ClinicalTrials.gov. We attempted to source each trial's protocol (final approved version), its primary manuscript reporting the clinical results, and any secondary manuscript reporting PROs. We identified PRO primary or secondary publications published on or before December 31, 2019, to allow trials supporting FDA approvals announced in 2018, a minimum timeframe of 1 year to publish PRO results. Exclusion criteria included FDA immunotherapy drug label modifications and trial results published in abstracts only. Examples of drug label modifications include approval of a new companion diagnostic test to determine programmed cell death protein 1 (PD-L1) levels in tumor tissues, dosage modification following population pharmacokinetics analyses and dose- or exposure-response analyses, or revision of drug label to limit the indication of a given drug after evidence of limited survival benefit on long-term follow-up. Figure 1 shows the search strategy flowchart and the inclusion and exclusion criteria used. Further details regarding our search strategy are described in the Supplementary Methods (available online). We adhered to the Cochrane Handbook for Systematic Reviews of Interventions, and the study is reported in accordance with the Preferred Reporting Items for Systematic Reviews and Meta-Analyses (PRISMA) guidelines.

Two authors (H.S. and M.T.) independently collected study characteristics and PRO information from both published journal articles and trial protocols (when available). If an FDA approval was supported by more than 1 trial, we collected information from all relevant trials. A third researcher (J.C. or A.S.) was involved when no consensus was reached in any of the data collection steps. This study did not use patient data, because it only included publicly available data; therefore, institutional review board approval was not required.

Data Collection

The data extraction sheet was derived and modified from a previous review (19) and included 47 predefined evaluation criteria, which were divided into 6 categories: general trial description, reporting of research hypothesis and objectives, PRO instruments, clinical relevance, statistical analysis, and handling of missing data (Supplementary Methods and Supplementary Table 1, available online).

Quantitative Scoring System for the Quality of PRO Reporting and Endpoint Analysis Methods

We developed a quantitative scoring scale to evaluate the quality of the methods used to report and analyze PROs in the RCTs that were included in our data set: the PRO Endpoints Analysis
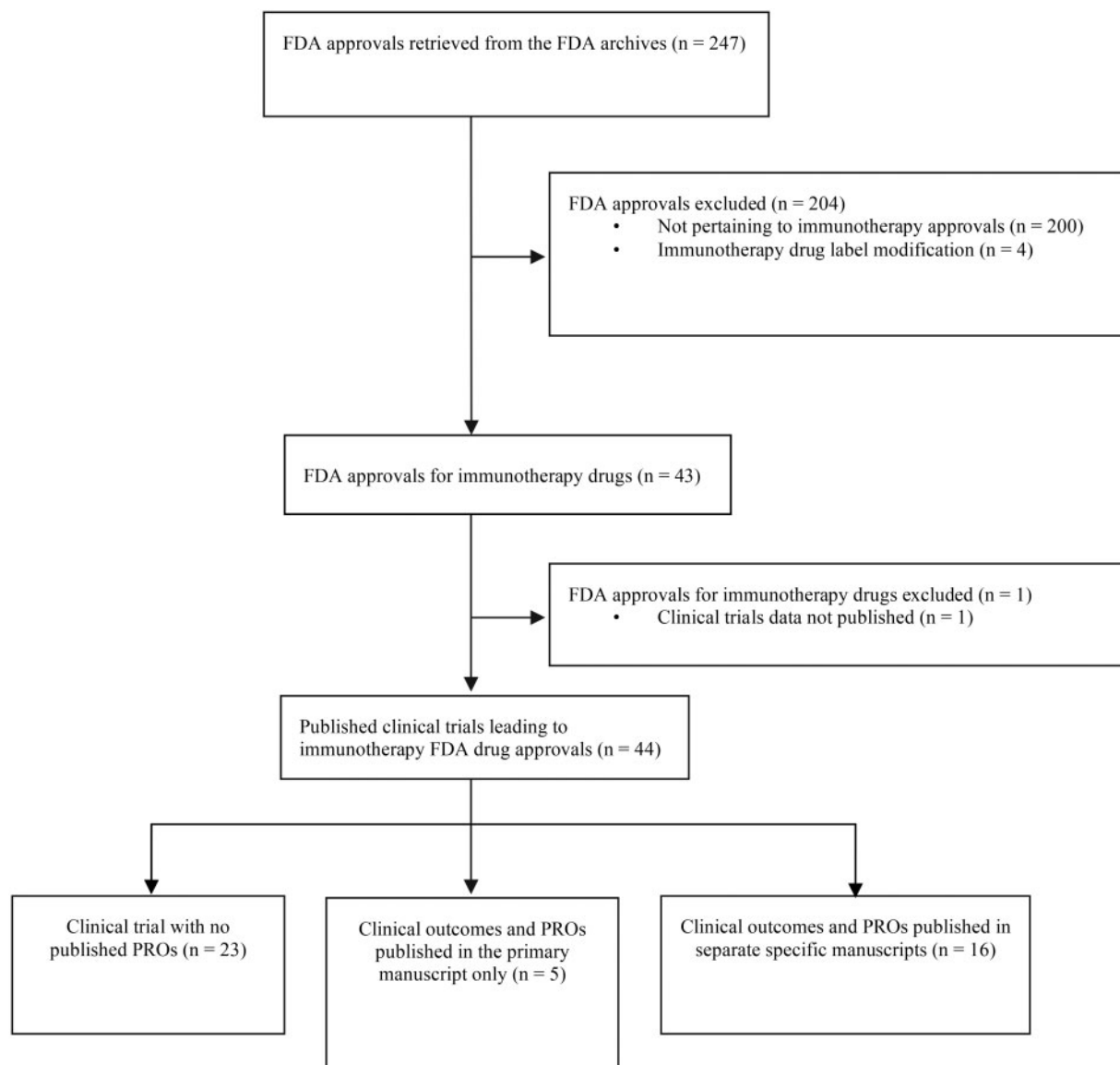
Figure 1. Study selection flowchart. FDA = Food and Drug Administration; PRO = patient-reported outcome.

Score (PROEAS), which is based on the 2020 recommendations of the SISAQOL Consortium (31). Three authors (H.S, M.T., and J.C.) refined the criteria used for the PROEAS and developed the initial scoring sheet. Two authors (H.S and M.T.) independently pilot tested the scoring sheet on 4 initial studies. Following the pilot testing, a reevaluation was conducted to further refine and troubleshoot the scoring scale. When opinions differed on how a criterion should be defined, the variable was further clarified. Final agreements were then reviewed and approved by a third author (J.C.). Two authors (H.S and M.T.) then independently collected data from all eligible studies using the final version of the PROEAS scoring sheet. When disagreements arose, the articles were reviewed by both authors. If no consensus was reached, J.C. was the third reviewer and served as a mediator to resolve disagreements. The final criteria included in the scoring sheet were based on 24 items derived from the recommendations of the 2020 SISAQOL consortium and resulted in a 24-point scoring scale. Each item was scored "1" if it was adequately reported or "0" if it was not clearly reported or not reported at all; items were weighted equally. The criteria included in the PROEAS were classified into 4 categories as detailed below (Table 1).

Taxonomy of research objectives

This section aimed at evaluating the extent to which the investigators defined a priori meaning in the trial's protocol, the PRO endpoint, the direction of the hypothesis, the PRO objectives, and the clinical relevance for PRO score changes. According to the SISAQOL recommendations, a hypothesis was required for trials in which PROs were a primary or secondary endpoint but not for exploratory endpoints (31). Additionally, we assessed whether comparative conclusions made regarding PRO data were appropriate based on the type of PRO endpoint. Comparative conclusions were considered appropriate if PROs were a primary or secondary endpoint only and inappropriate if PROs were exploratory endpoints.

Table 1. PROEAS items and the SISAQOL recommended statements they are derived from

SISAQOL statement	PROEAS item	Points
	Taxonomy of research objectives (7 points)	
RS 1	PRO endpoints were specified in the protocol	1
	Hypothesis requirement was met as needed[a]	1
	Endpoints were used to make appropriate conclusions[b]	1
RS 2	Direction of the hypothesis was prespecified in the protocol if required	1
	Clinical relevance for between-group differences was prespecified in the protocol	1
RS 3–8	Within-treatment group objective stated in the protocol	1
	Clinical relevance for within-patient or within–treatment group change was predefined in the protocol	1
	Recommending statistical methods (6 points)	
RS 10	Statistical test comparing 2 groups done when appropriate	1
	Provided P values for statistical significance	1
	Tests used adjusted for baseline covariates	1
	Tests used handled clustered data (repeated assessments)	1
	Correction for multiple testing done appropriately	1
RS 11–15	Used at least 1 appropriate statistical test to evaluate the tested outcome[c]	1
	Standardizing statistical terms related to missing data (6 points)	
RS 16	A definition for missing data was reported	1
RS 18	Study did not consider PRO assessments for deceased patients as missing data	1
RS 21–22	Variable denominator rate reported	1
RS 19–20	Fixed denominator rate reported	1
RS 23	Absolute number for both numerators and denominators were reported	1
	A CONSORT diagram or table reporting reasons for treatment discontinuation was provided	1
	General handling of missing data (5 points)	
RS 27	Study documented a priori the approach for handling missing data	1
RS 28	Item-level missing data handled according to the scoring algorithm of the instrument	1
RS 30	A method that allows the use of all available data was used to approach missing data	1
RS 31	Study did not use explicit imputation methods unless justified	1
RS 32	At least 2 different approaches to handle missing data were used	1

[a]A hypothesis statement was required if PROs were a secondary endpoint and not required if PROs were an exploratory endpoint. CONSORT = Consolidated Standards of Reporting Trials; PRO = patient-reported outcome; PROEAS = Patient Reported Outcomes Endpoints Analysis Score; RS = recommended statement; SISAQOL = Standards in Analyzing Patient-Reported Outcomes and Quality of Life.
[b]Comparative conclusions were considered appropriate if PROs were a primary or secondary endpoint only and inappropriate if PROs were exploratory endpoints.
[c]Statistical tests were considered appropriate if they met the recommendations of the SISAQOL.

Recommending Statistical Methods

The main focus of this subcategory score was assessing whether the statistical techniques used for the data analysis were appropriate and met the recommendations of the SISAQOL. The SISAQOL recommended a list of statistical techniques that should be used for each prespecified objective (31). The recommended techniques met certain criteria deemed necessary for a transparent data analysis.

Standardizing statistical terms related to missing data

In this section, the focus was to determine whether the clinical trials defined and presented the missing data in accordance with the recommendations of the SISAQOL. We evaluated whether a definition for missing data was provided a priori and whether the authors provided absolute and relative numbers for the rates of missing data at each time point. Per SISAQOL recommendations, deceased patients should be excluded from the denominator of the completion rate at assessment points after death. We also identified whether a CONSORT diagram reporting reasons for treatment discontinuation was provided.

General handling of missing data

In this subcategory score, we evaluated whether the methods used to handle missing data met the recommendations of the SISAQOL. Most importantly, we identified whether the authors documented a priori the approach for handling missing data.

Statistical Analysis

We used descriptive statistics with means, medians, and interquartile ranges for continuous variables and frequencies and percentages for categorical variables. The χ^2 and Fisher exact tests were used to compare categorical variables. All descriptive statistics for PRO publications characteristics were prespecified. Additionally, we conducted an exploratory analysis investigating the correlation between the likelihood of publishing PROs according to the following variables: type of immunotherapy, clinical trial phase, number of study arms, type of FDA approval, type of cancer, approval indication, and type of primary endpoint of the trial. All P values were 2-sided, and a P value of less than .05 was considered statistically significant. All statistical analyses were conducted using SPSS (Version 25.0, IBM Corp, Armonk, NY).

Results

PRO Publications

Our search identified 43 FDA approvals of immunotherapy drugs. One approval was excluded because it was based on pooled data from several trials and no published manuscripts were identified. Two approvals were based on 2 different trials each. Thus, the total number of FDA approvals included in our dataset was 42 and the total number of trial publications

Table 2. Characteristics of FDA approvals (n = 42) and clinical trial publications leading to these approvals (n = 44)[a]

Trial characteristic	No. (%)
Approved immunotherapy type (n = 42)	
Anti–PD-1	
Pembrolizumab	11 (26.2)
Nivolumab	11 (26.2)
Cemiplimab	1 (2.4)
Anti–PD-L1	
Atezolizumab	3 (7.1)
Avelumab	2 (4.8)
Durvalumab	2 (4.8)
Anti–PD-1 + Anti–CTLA-4	
Nivolumab + ipilimumab	3 (7.1)
Anti–PD-1 + chemotherapy	
Pembrolizumab + carboplatin + paclitaxel/nab-paclitaxel	1 (2.4)
Pembrolizumab + carboplatin + pemetrexed	1 (2.4)
Pembrolizumab + pemetrexed + platinum	1 (2.4)
CAR-T cell therapy	
Tisagenlecleucel	2 (4.8)
Axicabtagene ciloleucel	1 (2.4)
Anti–CTLA-4	
Ipilimumab	2 (4.8)
PegInterferon	1 (2.4)
Approval type (n = 42)	
Regular	22 (52.4)
Accelerated	20 (47.6)
Approval indication (n = 42)	
First line	8 (19)
Second line and beyond	30 (71.4)
Adjuvant	3 (7.1)
Maintenance	1 (2.4)
Immunotherapy drug approved alone (n = 42)	35 (83.3)
Immunotherapy drug approved in combination (n = 42)	7 (16.7)
Year of FDA drug approval (n = 42)	
2018	13 (31.0)
2017	13 (31.0)
2016	6 (14.3)
2015	6 (14.3)
2014	2 (4.8)
2011	2 (4.8)
Tumor type (n = 42)	
Non-small cell lung cancer	12 (28.6)
Melanoma	7 (16.7)
Urothelial cancer	5 (11.9)
Large B-cell lymphoma	3 (7.1)
Renal cell carcinoma	2 (4.8)
Colorectal cancer	2 (4.8)
Hepatocellular carcinoma	2 (4.8)
Squamous cell carcinoma of the head and neck	2 (4.8)
Hodgkin lymphoma	2 (4.8)
Merkel cell carcinoma	2 (4.8)
Cervical cancer	1 (2.4)
Cutaneous squamous cell carcinoma	1 (2.4)
Gastric or gastroesophageal junction adenocarcinoma	1 (2.4)
Small cell lung cancer	1 (2.4)
Acute lymphoblastic leukemia	1 (2.4)
Phase of published trial supporting the approval (n = 44)	
Phase 1	4 (9.1)
Phase 1/2	2 (4.5)

(continued)

Table 2. (continued)

Trial characteristic	No. (%)
Phase 2	21 (47.7)
Phase 3	17 (38.6)
Randomization status of published trial supporting the approval (n = 44)	
Randomized clinical trial	23 (52.3)
Single-arm clinical trial	21 (47.7)
Supporting trial published data on PROs (n = 44)	21 (47.7)

[a]CAR = chimeric antigen receptor; CTLA-4 = cytotoxic T lymphocyte-associated protein 4; FDA = Food and Drug Administration; PD-1 = programmed cell death protein 1; PD-L1 = programmed cell death ligand 1; PRO = patient reported outcome.

supporting these FDA approvals was 44. Information on each trial publication is shown in Supplementary Table 2 (available online). Almost one-half (20 of 42 [47.6%]) of the approvals were accelerated (Table 2). Twenty-three trials (23 of 44 [52.3%]) pertained to RCTs and 21 (21 of 44 [47.7%]) pertained to single-arm trials. Data on PROs were published for less than one-half (21 of 44 [47.7%]) of the trial publications supporting the FDA approvals, including 5 studies (5 of 44 [11.4%]) that reported preliminary PRO results in the primary clinical outcomes publication only and 16 studies (16 of 44 [36.4%]) that reported PROs in a secondary dedicated publication, 4 of which also published preliminary PRO data in the primary clinical outcomes publication. Among the 44 trials included in our cohort, 36 (81.8%) planned to collect PRO data (as reported in the trial's protocol or in the methods section of the primary clinical outcomes manuscript or on clinicaltrials.gov) and 8 (18.1%) did not. Of the 36 that planned to collect PRO data, 21 (58.3%) reported PRO results and 15 (41.7%) failed to report. Three FDA approvals were based on interim efficacy analysis; however, it is important to note that for all 3, PRO data were subsequently published in a secondary dedicated manuscript after proper follow-up periods.

The median time between primary clinical outcomes publications and their corresponding secondary PRO publication was 19 months (interquartile range = 9-29 months). Trial publications supporting regular FDA approvals were more likely to have published PROs than were those supporting accelerated FDA approvals (62.5% vs 30.0%, respectively; P = .03) (Supplementary Table 3, available online). Similarly, among different trial phases, phase III and phase IV trials were the most likely to publish PRO results (P = .012). Further analysis showed that clinical trials that had overall survival as a primary outcome were more likely to publish PRO results than those that did not (66.7% vs 37.9%); however, the difference was not statistically significant (P = .7). Supplementary Table 4 (available online) shows a side-by-side comparison of reporting characteristics between PROs published in a secondary dedicated manuscript and those published only in the primary manuscript.

PRO Characteristics

Among the 21 studies that published PRO data, PROs were defined as exploratory endpoints in 10 (10 of 21 [47.6%]); as secondary endpoints in 9 (9 of 21; [42.9%]); and the type of PRO endpoint was unclear, meaning not specified, in 2 (2 of 21 [9.5%]) (Table 3). Only 4 studies (4 of 21 [19.0%]) reported a specific PRO hypothesis, including 2 that defined the PRO

SYSTEMATIC REVIEW

Table 3. Characteristics of included PRO publications and their PRO data collection methods (n = 21)[a]

PRO characteristic	No. (%)
PRO publication year	
2019	3 (14.3)
2018	4 (19.0)
2017	10 (47.6)
2016	2 (9.5)
2012	1 (4.8)
2009	1 (4.8)
Journal impact factor at time of publication	
<10	5 (23.8)
10-20	3 (14.3)
>20	13 (61.9)
PRO stated as an endpoint	
Primary endpoint	0
Secondary endpoint	9 (42.9)
Exploratory endpoint	10 (47.6)
Unclear	2 (9.5)
PRO hypothesis	
Specific	4 (19.0)
Broad	12 (57.1)
Not reported	5 (23.8)
PRO instruments	
PRO instruments used	
EQ-5D	17 (81.0)
EORTC QLQ-C30	14 (66.7)
EORTC QLQ-LC13	4 (19.0)
LCSS	2 (9.5)
IFN-specific symptom checklist	1 (4.8)
FKSI-DRS	1 (4.8)
FACT-G	1 (4.8)
FACT-M	1 (4.8)
FKSI-19	1 (4.8)
PedsQL	1 (4.8)
QLQ-H&N35	1 (4.8)
Site-specific PRO instrument	9 (42.9)
Reference of the PRO instrument provided	18 (85.7)
Data collection	
PRO collection method	
Electronic	10 (47.6)
Paper	6 (28.6)
Not reported	5 (23.8)
Time point assessment	
Baseline time point collected	21 (100)
Two or more follow-up time points collected	21 (100)

[a]EORTC QLQ-C30 = European Organisation for Research and Treatment of Cancer Quality of Life Questionnaire-core questionnaire; EORTC QLQ-LC13 = EORTC QLQ-Lung Cancer Module; EQ-5D = EuroQol-5D; FACT-G = Functional Assessment of Cancer Therapy: General; FACT-M = FACT-Melanoma; FKSI-19 = Functional Assessment of Cancer Therapy – Kidney Symptom Index; FKSI-DRS = Functional Assessment of Cancer Therapy – Kidney Symptom Index—Disease Related Symptoms; IFN = interferon; LCSS = Lung Cancer Symptom Scale; PedsQL = Pediatric Quality of Life Inventory; PRO = patient-reported outcome; QLQ-H&N35 = Quality of Life Questionnaire Head and Neck Cancer Module.

hypothesis in the trial protocol only and not in the published manuscript.

PRO Instruments

The most commonly used PRO instrument was the generic standardized EuroQol-5D questionnaire (17 of 21 [81.0%]). Less than one-half of the studies that reported PRO data (9 of 21 [42.9%]) used a disease site-specific instrument, for example the Quality of Life Questionnaire and Lung Cancer Module, a lung cancer-specific questionnaire.

Clinical Relevance, Statistical Analysis, and Handling of Missing Data

There was heterogeneity in the definitions of the main data set used for the analysis of PRO data, with the most commonly used being "patients with a baseline assessment and at least 1 postbaseline assessment" (8 of 21 [38.1%]) (Table 4). Most studies (17 of 21 [81.0%]) reported questionnaire completion, and/or compliance rates, at baseline and at subsequent time points. Clinical relevance was addressed using several definitions; the most commonly used was "a change of X points from baseline within a patient or within a treatment group" (12 of 21 [57.1%]).

Among the 21 PRO studies, 18 (18 of 21 [85.7%]) used descriptive summaries either on their own or alongside other, more advanced statistical techniques for data analysis. We observed heterogeneity in the strategies used to conduct the main analysis of the PROs. For example, we identified 9 different statistical tests used to assess the magnitude of change from baseline. Although the PRO studies included in our review tested for multidimensional endpoints and assessed multiple time points, only 3 (3 of 21 [14.3%]) studies performed a correction for multiple testing (control for type I error). The strategy to handle missing data was stated in 17 (17 of 21 [81.0%]) studies; however, 2 studies stated that the approach to deal with missing data was "no imputations" and 1 stated that it was "left as missing."

PRO Conclusions and Limitations

Sixteen of the 21 (76.2%) published PRO studies discussed the limitations facing the interpretation of the data (Table 4). As for the conclusion, single-arm studies concluded either improvement in key PROs (6 of 21 [28.6%]) or stable PROs (1 of 21 [4.8%]) after initiation of treatment. The majority of RCTs concluded that PRO findings in the experimental arm were superior to those in the control arm (9 of 21 [42.9%]).

PROEAS Ratings

Of the 23 RCTs in our dataset, 14 (60.9%) published PRO data, including 13 that published a secondary dedicated PRO manuscript. The PROEAS was used to assess data from these 14 RCTs, and all single-arm studies were excluded from this analysis. The mean PROEAS was 12.71 (range = 5-17, SD= 3.71) on a 24-point scale. One-half of the studies had a score of 13 or less (7 of 14 [50.0%]). The most appropriately reported item was specifying PRO endpoints in the clinical trial protocol (14 of 14 [100%]) (Table 5). In contrast, none of the studies provided a definition for missing data in the trial protocol or in the methods section of the published manuscript. The most deficient subcategory score was "general handling of missing data," with a mean score of 1.71 on a 5-point subscale (range = 0-4, SD = 1.27); only 4 studies (4 of 14 [28.6%]) documented a priori, meaning in the trial protocol, the approach for handling missing data. The second-most deficient subcategory score was "standardizing statistical terms related to missing data" with a mean score of 2.5 on a 6-point subscale (range = 0-4, SD = 1.16). The "taxonomy of research objectives" subcategory mean score was 3.86 (range = 2-6, SD = 1.29) on a 7-point subscale. In this

Table 4. Data reporting, clinical relevance, and statistical analysis methods used for PRO data (n = 21)

Variable	No. (%)
Data reporting	
Definition of study population	
Patients with a baseline assessment and at least 1 postbaseline assessment	8 (38.1)
Intent to treat	5 (23.8)
Patients who received at least 1 dose of study medication and completed at least 1 assessment	4 (19.0)
Modified intent-to-treat population	1 (4.8)
Intent-to-treat population with nonmissing baseline measurements	1 (4.8)
Not clearly defined	2 (9.5)
PRO completion rate	
Completion/compliance rate table included in the manuscript	17 (81.0)
Clinical relevance threshold was prespecified as	
Change of X points from baseline within-patient or treatment group	12 (57.1)
Difference of X points between arms at a certain timepoint	2 (9.5)
Both	3 (14.3)
Not reported	4 (19.0)
Clinical relevance was justified and cited[a]	17 (81.0)
Reporting of descriptive data	18 (85.7)
Data analysis	
Primary statistical technique	
Time-to-event analysis	
Cox proportional hazard/Cox regression model	11 (52.4)
Log rank test	6 (28.6)
Brookmeyer and Crowley	1 (4.8)
Magnitude of change from baseline analysis	
Mixed-model for repeated measures	6 (28.6)
cLDA/LDA	4 (19.0)
Mixed effects model	3 (14.3)
ANCOVA	2 (9.5)
t tests	3 (14.3)
Chi-square	2 (9.5)
Conventional Wald method	1 (4.8)
ANOVA	1 (4.8)
Linear regression	1 (4.8)
Response trajectory over time	
Linear mixed effects model	4 (19.0)
Symptom improvement rate at time t analysis	
Clopper Pearson	2 (9.5)
Logistic regression	2 (9.5)
Not reported or unclear	2 (9.5)
PRO scores were compared at baseline between 2 arms	13 (61.9)
Control for type I error	3 (14.3)
PRO data analysis stratified by ethnicity/race	0
Handling of missing data	
Strategy to deal with missing data is defined[b]	17 (81.0)
Detailed reasons for missing data by timepoint reported	0
PRO specific limitations stated in the discussion section	16 (76.2)
PRO conclusion	
Improvement in key PROs[c]	6 (28.6)
Stable PROs[c]	1 (4.8)
Experimental arm is superior to control arm	9 (42.9)
Similar outcomes between experimental arm and control arm	3 (14.3)
Experimental arm is inferior to control arm	2 (9.5)

[a]Articles that did not justify clinical relevance where those that did not specify the clinical relevance. Numbers are rounded to the nearest whole number. ANCOVA = analysis of covariance; ANOVA = analysis of variance; cLDA = constrained LDA; LDA = longitudinal data analysis; PRO = patient-reported outcome.
[b]Two manuscripts stated that the approach to deal with missing data was "no imputations" and 1 stated that the approach to deal with missing data was "left as missing."
[c]Single-arm studies.

SYSTEMATIC REVIEW

subcategory score, the clinical relevance for between-group differences was prespecified in the trial protocol in only 1 of the RCTs (1 of 14 [7.1%]). The "recommending statistical methods" subcategory mean score was 4.64 (range = 0-6, SD = 1.49) on a 6-point subscale. In this subcategory score, 12 RCTs (12 of 14 [85.7%]) used at least 1 appropriate statistical test to evaluate the tested PRO endpoint. On further analysis, we found that PRO studies of drugs approved during or after the year 2016 had

Table 5. PROEAS item reporting for randomized clinical trials (n = 14)[a]

Variable	No. (%)
Taxonomy of research objectives	
PRO endpoints were specified in the trial protocol	14 (100)
Hypothesis requirement was met as needed	9 (64.3)
Endpoints were used to make appropriate conclusions	9 (64.3)
Direction of the hypothesis was prespecified in the trial protocol if required	9 (64.3)
Clinical relevance for between-group differences was prespecified in the trial protocol	1 (7.1)
Within-treatment group objective stated in the trial protocol	5 (35.7)
Clinical relevance for within-patient or within–treatment group change was predefined in the trial protocol	7 (50.0)
Recommending statistical methods	
Statistical test comparing 2 groups done when appropriate	13 (92.9)
Provided P values for statistical significance	12 (85.7)
Tests used adjusted for baseline covariates	13 (92.9)
Tests used handled clustered data (repeated assessments)	12 (85.7)
Correction for multiple testing done appropriately	3 (21.4)
Used at least 1 appropriate statistical test to evaluate the tested outcome	12 (85.7)
Standardizing statistical terms related to missing data	
A definition for missing data was reported	0
Study did not consider PRO assessments for deceased patients as missing data	1 (7.1)
Variable denominator rate reported	9 (64.3)
Fixed denominator rate reported	6 (42.9)
Absolute numbers for both numerators and denominators were reported	11 (78.6)
A CONSORT diagram or table reporting reasons for treatment discontinuation was provided	9 (64.3)
General handling of missing data	
Study documented a priori the approach for handling missing data	4 (28.6)
Item-level missing data handled according to the scoring algorithm of the instrument	3 (21.4)
A method that allows the use of all available data was used to approach missing data	1 (7.1)
Study did not use explicit imputation methods unless justified	7 (50.0)
At least 2 different approaches to handle missing data were used	5 (35.7)

[a]CONSORT = Consolidated Standards of Reporting Trials; PRO = patient-reported outcome; PROEAS = Patient Reported Outcomes Endpoints Analysis Score; SISAQOL = Setting International Standards in Analyzing Patient-Reported Outcomes and Quality of Life Endpoints Data.

a numerically higher mean score than those published between 2011 and 2015 (13.29 vs 12.29, respectively). Similarly, PRO studies published during or after 2018 had a numerically higher mean score than those published between 2011 and 2017 (12.68 vs 11.78, respectively). Additionally, PRO articles published in journals with an impact factor (IF) of at least 11 had a higher mean score than those published in journals with an IF less than 11 (13.33 vs 11.6, respectively).

Discussion

Our study summarizes the state of PROs associated with immunotherapy drugs for cancer treatment that led to FDA drug approvals. Our data showed a considerable gap in the reporting of PROs. In fact, PROs were published for less than one-half (21 of 44 [47.7%]) of the trial publications supporting the FDA approvals, taking note that 5 (5 of 44 [11.4%]) studies published preliminary PRO results in the primary clinical outcomes manuscript only. The latter fact is important to note because the publication of PROs in primary clinical outcomes manuscripts limits the space allocated to PRO results and leads to the inclusion of only a small part of the data (22). Although the CONSORT-PRO recommends reporting key primary and secondary PRO outcomes in the main trial's manuscript, the consortium encourages publishing other PRO endpoints (such as exploratory) and components of composite PRO scores in the online supplements of the journal alongside the primary clinical outcomes manuscript (30). It is important to note that some of the trials that did not publish PRO data did not plan to include

PROs as endpoints; nonetheless, the vast majority of immunotherapy trials did plan to collect data on PROs but failed to report them. The relative recency of immunotherapy drugs and their accelerated approvals have led to short-term, real-world experience with these drugs. Therefore, providing comprehensive data about these drugs, including PROs, allows treating physicians to adopt a patient-centered approach.

The absence of a specific hypothesis is a well-recognized weakness in the design of PRO studies, as shown in previous reviews (20,21,23), and our study further emphasizes this weakness. The timeline and characteristics of AEs associated with immunotherapy are different from those associated with chemotherapy; therefore, stating a specific hypothesis specifying prespecified time points and PRO aspects pertinent to immunotherapy is crucial for synthesizing relevant data (36,37) because different analytic strategies may lead to different conclusions (38-40). Additionally, we identified several instruments used to collect PRO data, but the majority of studies did not use disease site-specific PRO tools, and none used an immunotherapy-specific PRO tool. This latter fact is largely due to the lack of instruments developed and validated to assess HRQoL changes and symptom patterns specific to patients receiving immunotherapy drugs. Traditional PRO instruments measure common symptoms and effects that historically have been associated with chemotherapy; however, these instruments may fail to detect changes in HRQoL or symptoms that are specific to the use of immunotherapy drugs (41,42).

Our review showed substantial variability among studies in the types of statistical tests and approaches used. This heterogeneity poses challenges when attempting to summarize the

evidence regarding PROs associated with the use of immunotherapy drugs; it also limits the medical community's ability to perform comparisons between trials. Our data also showed an additional deficiency in the analysis of PRO data regarding the correction for multiple testing (control for type I error). Adjusting *P* values to compensate for type I error is a fundamental step in preventing overreaching conclusions regarding safety data (43). This gap in reporting was also identified in other reviews examining the state of statistical reporting in clinical trials of other types of cancer treatments (19,32). Most importantly, none of the single-arm studies reported a deterioration in PROs after administration of the drug, and very few of the RCTs in our cohort reported inferior PROs compared with the control arm (2/21 [9.5%]), which raises the question of whether negative PRO results go unpublished because they are thought to be uninteresting data.

Efforts to standardize the methods to handle and analyze PRO data resulted in the recent publication of guidelines and recommendations by the SISAQOL Consortium regarding the existing gap in this area of RCTs (31). Using the PROEAS, our data showed that none of the RCTs met all of the recommendations of the SISAQOL Consortium, and one-half met 13 or fewer of 24 recommendations (7 of 14 [50.0%]). The most deficient subcategory scores were related to the definition and handling of missing data. According to the SISAQOL Consortium recommendations, missing data should only include data that are meaningful for testing a certain prespecified hypothesis but were not collected (31). The abundance of missing data, the absence of a clear definition for what are considered missing data, and the poor handling of missing data weaken the statistical significance of PRO findings and may discredit the stated conclusions (44,45). The majority of RCTs (85.7%) used at least 1 appropriate statistical test to evaluate the tested outcome. Although encouraging, these data should be interpreted with caution, because most studies used more than 1 statistical test for multiple outcomes, and not all of these tests were necessarily adhered to the SISAQOL Consortium recommendations. Several RCTs defined PROs as exploratory endpoints and stated a comparative conclusion for the 2 arms of the study (5 of 14 [35.7%]). Per SISAQOL recommendations, exploratory endpoints do not require a predefined hypothesis; however, the recommendations also hold that comparative analysis of exploratory endpoints should be avoided (31). Authors should be discouraged from stating definitive interpretations and would preferably use descriptive statistics for this type of endpoint. Similarly, clinical relevance thresholds for within-group or between-group changes were rarely predefined in the trials' protocols we reviewed. Without predefining clinical relevance thresholds, researchers risk performing statistical analyses without a prespecified and justified threshold for statistically significant differences, which may lead to false conclusions.

PROEAS mean scores suggested that the quality of PRO design and analysis in RCTs may have improved over time. This improvement might be due to previous efforts aimed at standardizing PROs, such as the SPIRIT-PRO (29) and the CONSORT-PRO (30), which have items that might overlap with those suggested by the SISAQOL Consortium. Similarly, PROEAS mean scores were higher when the PRO study was published in journals with an IF of at least 11. This finding might be explained by the strict requirements imposed by these journals, which commonly require more robust methods and reporting and may be particularly critical in their review of the manuscripts and trials' protocols before publication.

Our study has several limitations. First, we included only clinical trials related to immunotherapy drugs that led to FDA approvals. We believe this was a limitation given that our data and conclusions cannot be extrapolated to all immunotherapy trials in the mentioned timeframe, including those that did not lead to FDA approvals. Second, the PROEAS is based on recommendations that are pertinent to RCTs only, and therefore we were unable to provide an objective scale to measure the quality of the methods in single-arm clinical trials. Furthermore, the SISAQOL Consortium is still developing additional standards regarding PRO methods and designs; thus, our PROEAS evaluated only items that are included in the current recommendations. In contrast, the strength of this study emerges from its robust data collection strategy, which used both the trials' protocols and the published manuscripts to acquire the information required. Additionally, we used the most up-to-date recommendations from the SISAQOL Consortium; to our knowledge, this is the first study to use this approach for PRO quality assessment in RCTs.

The fast pace at which immunotherapy is evolving comes at the cost of suboptimal reporting and delays in publishing PROs. Much is still unknown about the PROs of immunotherapy drugs that have received FDA approval, because PRO data are missing for most trials. A standardized design and methodology as well as timely reporting of PROs are crucial for anticipating the real-world patient experience associated with these drugs. Collective efforts have been under way to overcome the absence of clear guidelines on designing, analyzing, and reporting PRO data, and the field has made tremendous progress exhibited by the recommendations of CONSORT-PRO, SPIRIT-PRO, and SISAQOL (29-31). Additionally, multiple new tools are emerging in the field of PROs to optimize the feasibility and usability of PROs in clinical trials, including the PRO-CTCAE and FACT-GP5 question. The PRO-CTCAE offers an opportunity to generate PRO data based on a standardized measurement system providing a flexible fit-for-purpose approach that is relevant across a broad range of cancer therapies (26,27). The FACT-GP5 is an increasingly popular tool consisting of 1 question with a promise of providing an overall summary measure of the burden of AEs on the patient's quality of life (46). This question measures the extent to which patients are bothered by side effects of treatments. It has been shown to strongly correlate with the ability of patients to enjoy life and with the grade of clinician-reported AEs (46). The PRO-CTCAE and the FACT-GP5 are not widely used in immunotherapy clinical trials as shown by our data. However, these tools have the potential to maximize the relevance of PRO data in clinical decision making and could offer trialists a valuable addition in the design of future immunotherapy trials. A next step is to move toward proper implementation of these recommendations and tools to maximize the value of PRO data. Efforts towards this next step are being led by initiatives such as the Patient-Reported Outcomes Tools: Engaging Users and Stakeholders consortium involving key patient, clinician, research, and regulatory groups working together to coordinate and promote the uptake of the most up-to-date practices in PRO collection and handling (47).

Our study provides insight into the status of PROs in the immunotherapy era, presents an objective measure of the quality of PROs using the PROEAS, and serves to identify the deficiencies and limitations in their design, methodology, and reporting. Ultimately, more inclusive reviews should be performed to evaluate the extent to which all oncology clinical trials align with the recent recommendations of the SISAQOL Consortium.

Funding

Notes

Role of the funder: The funders had no role in the study design, collection, analysis, or interpretation of data; in the writing of the report; or in the decision to submit this paper for publication.

Disclosures: BG reports personal fees from SureMed Compliance and Elly Health, Inc outside the submitted work. AD reports personal fees from Nektar therapeutics, Idera pharmaceuticals and Apexigen outside the submitted work. All other authors declare no competing interests.

Author contributions: JC had full access to all of the data in the study and takes responsibility for the integrity and the accuracy of the data analysis. Study concept and design: HS, MT, AS, JC. Acquisition of data: HS, MT. Analysis and interpretation of data: HS, MT, FS, BG, LO, AS. Drafting of the manuscript: HS, JC. Critical revision of the manuscript for important intellectual content: HS, MT, PS, BM, JPS, SG, FS, BG, LO, AS, AD, JC. Administrative, technical, or material support: JC. Study supervision: JC. All authors read and approved the final manuscript.

Data Availability

The datasets supporting the conclusions of this article are included within the article and its additional files.

References

1. Hargadon KM, Johnson CE, Williams CJ. Immune checkpoint blockade therapy for cancer: an overview of FDA-approved immune checkpoint inhibitors. *Int Immunopharmacol.* 2018;62:29-39.
2. Riley RS, June CH, Langer R, et al. Delivery technologies for cancer immunotherapy. *Nat Rev Drug Discov.* 2019;18(3):175-196.
3. Boyiadzis MM, Dhodapkar MV, Brentjens RJ, et al. Chimeric antigen receptor (CAR) T therapies for the treatment of hematologic malignancies: clinical perspective and significance. *J Immunother Cancer.* 2018;6(1):137.
4. Han X, Wang Y, Wei J, et al. Multi-antigen-targeted chimeric antigen receptor T cells for cancer therapy. *J Hematol Oncol.* 2019;12(1):128.
5. West HJ. JAMA oncology patient page. Immune checkpoint inhibitors. *JAMA Oncol.* 2015;1(1):115.
6. Fukumura D, Kloepper J, Amoozgar Z, et al. Enhancing cancer immunotherapy using antiangiogenics: opportunities and challenges. *Nat Rev Clin Oncol.* 2018;15(5):325-340.
7. Xin Yu J, Hubbard-Lucey VM, Tang J. Immuno-oncology drug development goes global. *Nat Rev Drug Discov.* 2019;18(12):899-900.
8. U.S. Department of Health and Human Services Food and Drug Administration. Guidance for Industry Patient-Reported Outcome Measures: Use in Medical Product Development to Support Labeling Claims. https://www.fda.gov/media/77832/download. Accessed September, 2020.
9. LeBlanc TW, Abernethy AP. Patient-reported outcomes in cancer care - hearing the patient voice at greater volume. *Nat Rev Clin Oncol.* 2017;14(12):763-772.
10. Bottomley A. The cancer patient and quality of life. *Oncologist.* 2002;7(2):120-125.
11. Brudno JN, Kochenderfer JN. Toxicities of chimeric antigen receptor T cells: recognition and management. *Blood.* 2016;127(26):3321-3330.
12. Martins F, Sofiya L, Sykiotis GP, et al. Adverse effects of immune-checkpoint inhibitors: epidemiology, management and surveillance. *Nat Rev Clin Oncol.* 2019;16(9):563-580.
13. Di Maio M, Gallo C, Leighl NB, et al. Symptomatic toxicities experienced during anticancer treatment: agreement between patient and physician reporting in three randomized trials. *J Clin Oncol.* 2015;33(8):910-915.
14. Atkinson TM, Ryan SJ, Bennett AV, et al. The association between clinician-based common terminology criteria for adverse events (CTCAE) and patient-reported outcomes (PRO): a systematic review. *Support Care Cancer.* 2016;24(8):3669-3676.
15. Atkinson TM, Dueck AC, Satele DV, et al. Clinician vs patient reporting of baseline and postbaseline symptoms for adverse event assessment in cancer clinical trials. *JAMA Oncol.* 2020;6(3):437-439.
16. The Lancet Oncology. Immunotherapy: hype and hope. *Lancet Oncol.* 2018;19(7):845.
17. The Lancet Oncology. Calling time on the immunotherapy gold rush. *Lancet Oncol.* 2017;18(8):981.
18. Calvert MJ, O'Connor DJ, Basch EM. Harnessing the patient voice in real-world evidence: the essential role of patient-reported outcomes. *Nat Rev Drug Discov.* 2019;18(10):731-732.
19. Pe M, Dorme L, Coens C, 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-e469.
20. Fiteni F, Anota A, Westeel V, et al. Methodology of health-related quality of life analysis in phase III advanced non-small-cell lung cancer clinical trials: a critical review. *BMC Cancer.* 2016;16(1):122.
21. Brundage M, Bass B, Davidson J, et al. Patterns of reporting health-related quality of life outcomes in randomized clinical trials: implications for clinicians and quality of life researchers. *Qual Life Res.* 2011;20(5):653-664.
22. Bylicki O, Gan HK, Joly F, et al. Poor patient-reported outcomes reporting according to CONSORT guidelines in randomized clinical trials evaluating systemic cancer therapy. *Ann Oncol.* 2015;26(1):231-237.
23. Kyte D, Retzer A, Ahmed K, et al. Systematic evaluation of patient-reported outcome protocol content and reporting in cancer trials. *J Natl Cancer Inst.* 2019;111(11):1170-1178.
24. Mercieca-Bebber R, Friedlander M, Kok PS, et al. The patient-reported outcome content of international ovarian cancer randomised controlled trial protocols. *Qual Life Res.* 2016;25(10):2457-2465.
25. Basch E, Reeve BB, Mitchell SA, 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)dju244.
26. Kluetz PG, Chingos DT, Basch EM, et al. Patient-reported outcomes in cancer clinical trials: measuring symptomatic adverse events with the National Cancer Institute's Patient-Reported Outcomes Version of the Common Terminology Criteria for Adverse Events (PRO-CTCAE). *Am Soc Clin Oncol Educ Book.* 2016;35(36):67-73.
27. Dueck AC, Mendoza TR, Mitchell SA, 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):1051-1059.
28. Basch E, Pugh SL, Dueck AC, 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):409-418.
29. Calvert M, Kyte D, Mercieca-Bebber R, et al.; the SPIRIT-PRO Group. Guidelines for inclusion of patient-reported outcomes in clinical trial protocols: the SPIRIT-PRO extension. *JAMA.* 2018;319(5):483-494.
30. Calvert M, Blazeby J, Altman DG, et al. Reporting of patient-reported outcomes in randomized trials: the CONSORT PRO extension. *JAMA.* 2013;309(8):814-822.
31. Coens C, Pe M, Dueck AC, et al. International standards for the analysis of quality-of-life and patient-reported outcome endpoints in cancer randomised controlled trials: recommendations of the SISAQOL Consortium. *Lancet Oncol.* 2020;21(2):e83-e96.
32. Fiero MH, Roydhouse JK, Vallejo J, et al. 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-e589.
33. Gnanasakthy A, Barrett A, Evans E, et al. A review of patient-reported outcomes labeling for oncology drugs approved by the FDA and the EMA (2012-2016). *Value Health.* 2019;22(2):203-209.
34. King-Kallimanis BL, Howie LJ, Roydhouse JK, et al. Patient reported outcomes in anti-PD-1/PD-L1 inhibitor immunotherapy registration trials: FDA analysis of data submitted and future directions. *Clin Trials.* 2019;16(3):322-326.

35. Roydhouse JK, King-Kallimanis BL, Howie LJ, et al. Blinding and patient-reported outcome completion rates in US Food and Drug Administration Cancer Trial submissions, 2007-2017. *J Natl Cancer Inst.* 2019;111(5):459-464.

36. Tolstrup LK, Bastholt L, Zwisler AD, et al. Selection of patient reported outcomes questions reflecting symptoms for patients with metastatic melanoma receiving immunotherapy. *J Patient Rep Outcomes.* 2019;3(1):19.

37. Anagnostou V, Yarchoan M, Hansen AR, et al. Immuno-oncology trial endpoints: capturing clinically meaningful activity. *Clin Cancer Res.* 2017;23(17):4959-4969.

38. Field KM, Jordan JT, Wen PY, et al. Bevacizumab and glioblastoma: scientific review, newly reported updates, and ongoing controversies. *Cancer.* 2015;121(7):997-1007.

39. Burris HA, Lebrun F, Rugo HS, et al. Health-related quality of life of patients with advanced breast cancer treated with everolimus plus exemestane versus placebo plus exemestane in the phase 3, randomized, controlled, BOLERO-2 trial. *Cancer.* 2013;119(10):1908-1915.

40. Campone M, Beck JT, Gnant M, et al. Health-related quality of life and disease symptoms in postmenopausal women with HR(+), HER2(-) advanced breast cancer treated with everolimus plus exemestane versus exemestane monotherapy. *Cur Med Res Opin.* 2013;29(11):1463-1473.

41. Chakraborty R, Sidana S, Shah GL, et al. Patient-reported outcomes with chimeric antigen receptor T cell therapy: challenges and opportunities. *Biol Blood Marrow Transplant.* 2019;25(5):e155-e162.

42. Hall ET, Singhal S, Dickerson J, et al. Patient-reported outcomes for cancer patients receiving checkpoint inhibitors: opportunities for palliative care-a systematic review. *J Pain Symptom Manag.* 2019;58(1):137-156.e1.

43. The European Agency for the Evaluation of Medicinal Products. Points to consider on multiplicity issues in clinical trials. https://www.ema.europa.eu/en/documents/scientific-guideline/points-consider-multiplicity-issues-clinical-trials_en.pdf . Accessed September, 2020.

44. Machin D, Weeden S. Suggestions for the presentation of quality of life data from clinical trials. *Stat Med.* 1998;17(5-7):711-724.

45. Bell ML, Fairclough DL. Practical and statistical issues in missing data for longitudinal patient-reported outcomes. *Stat Methods Med Res.* 2014;23(5):440-459.

46. Pearman TP, Beaumont JL, Mroczek D, et al. Validity and usefulness of a single-item measure of patient-reported bother from side effects of cancer therapy. *Cancer.* 2018;124(5):991-997.

47. Patient-Centered Outcomes Research Institute. "PROTEUS" patient-reported outcomes tools: engaging users and stakeholders. https://www.pcori.org/research-results/2018/proteus-patient-reported-outcomes-tools-engaging-users-stakeholders. Accessed September, 2020.