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

Mallorie H Fiero, Jessica K Roydhouse, Jonathon Vallejo, Bellinda L King-Kallimanis, Paul G Kluetz, Rajeshwari Sridhara

Lancet Oncol 2019; 20: 582–89

Office of Biostatistics (M H Fiero PhD, J Vallejo PhD, R Sridhara PhD) **and Office of Hematology and Oncology Products** (J K Roydhouse PhD, B L King-Kallimanis PhD)**, Center for Drug Evaluation and Research, US Food and Drug Administration, Silver Spring, MD, USA; and Oncology Center of Excellence, US Food and Drug Administration, Silver Spring, MD, USA** (P G Kluetz MD)

Correspondence to:
Dr Mallorie H Fiero, Office of Biostatistics, Center for Drug Evaluation and Research, US Food and Drug Administration, Silver Spring, MD 20993, USA
mallorie.fiero@fda.hhs.gov

With the advent of patient-focused drug development, the US Food and Drug Administration (FDA) has redoubled its efforts to review patient-reported outcome (PRO) data in cancer trials submitted as part of a drug's marketing application. This Review aims to characterise the statistical analysis of PRO data from pivotal lung cancer trials submitted to support FDA drug approval between January, 2008, and December, 2017. For each trial and PRO instrument identified, we evaluated prespecified PRO concepts, statistical analysis, missing data and sensitivity analysis, instrument completion, and clinical relevance. Of the 37 pivotal lung cancer trials used to support FDA drug approval, 25 (68%) trials included PRO measures. The most common prespecified PRO concepts were cough, dyspnoea, and chest pain. At the trial level, the most common statistical analyses were descriptive (24 trials [96%]), followed by time-to-event analyses (19 trials [76%]), longitudinal analyses (12 trials [48%]), and basic inferential tests or general linear models (10 trials [40%]). Our findings indicate a wide variation in the analytic techniques and data presentation methods used, with very few trials reporting clear PRO research objectives and sensitivity analyses for PRO results. Our work further supports the need for focused research objectives to justify and to guide the analytic strategy of PROs to facilitate the interpretation of patient experience.

Introduction

Over the past decade, the US Food and Drug Administration (FDA) has increased its focus on patient-centred care in drug development and regulatory decision making. For example, the patient-focused drug development initiative, led by the Center for Drug Evaluation and Research, has included FDA-led meetings on specific disease areas and the development of four methodological guidance documents, which seek to systematically gather and incorporate patient perspectives on drug development.[1] Additionally, the FDA is reviewing existing datasets to inform regulatory research and future directions for the systematic evaluation of patient experience data.

In alignment with the FDA's patient-focused drug development effort, and in accordance with the 21st Century Cures Act,[2] the FDA is required to assess patient experience data that is submitted as part of a drug's marketing application. FDA guidance[3] defines a patient-reported outcome (PRO) as a type of clinical outcome assessment that is a measurement of a patient's health condition reported by the patient without interpretation from the clinician or anyone else. PROs are becoming more common in cancer trials, supplementing the results of clinical endpoints, such as overall survival, progression-free survival, and objective response rate.[4,5] The FDA review seeks to isolate the effect of the intervention and minimise the risk of potential bias, and so it focuses the initial review of PRO data supporting cancer drugs on a core set of outcomes, which are important contributors to patient health-related quality of life, symptomatic adverse events, physical function, and disease symptoms.[6]

Patient experience data can provide useful additional information on the risk–benefit assessment of cancer therapies.[7] However, inappropriate statistical analyses can lead to a biased interpretation of patient experience.[8,9] Current challenges include appropriately defining a PRO research objective, handling of missing data, multiplicity (ie, testing of multiple outcomes), and clinical interpretability.[3,10,11] In addition, the absences of standardisation in PRO analysis and reporting of results makes interpretation and comparison of PROs between cancer trials difficult.[9,12] Continuing efforts to standardise and improve the collection, analysis, and reporting of PROs in trials have come from a number of initiatives, including the Setting International Standards in Analyzing Patient-Reported Outcomes and Quality of Life Endpoints Data for Cancer Clinical Trials (SISAQOL),[4,9] the Consolidated Standards of Reporting Trials Statement PRO extension,[13] and the Standard Protocol Items: Recommendations for Interventional Trials-PRO extension.[14]

Lung cancer remains a leading cause of cancer-related mortality in the USA and is a focus of newer therapeutic approaches, such as immunotherapy.[15–17] Understanding the symptoms and functional effects of patients exposed to these new therapies is of interest to the pharmaceutical industry, regulators, clinicians, and patients. Thus, with increased attention on appropriately evaluating and communicating the patient experience in oncology drug development, our objective was to evaluate the statistical analysis of PROs in lung cancer pivotal trials that were used as the basis for drug approval by the FDA Office of Hematology and Oncology Products between January, 2008, and December, 2017. More specifically, we examined five aspects related to the statistical analysis of PROs: prespecified PRO concepts; statistical analysis; missing data and sensitivity analysis; instrument completion; and clinical relevance.

See Online for appendix

	Example statistical methods
Time-to-event	Log-rank test and Cox proportional hazards model
Longitudinal analysis	Generalised linear mixed model
Basic inferential test or generalised linear model	
Continuous basic inferential test or generalised linear model	
By timepoint	*t* test evaluating mean change in PRO score from baseline at time t within the treatment group
Overall	Analysis of covariance, Mann–Whitney–Wilcoxon test evaluating maximum improvement or worsening from baseline
Responder analysis basic inferential test or generalised linear model	
By timepoint	Logistic regression, χ^2 test or Fisher's exact test evaluating proportion of patients who improved, remained stable, or deteriorated between treatment arms at time t
Overall	Logistic regression, χ^2 test or Fisher's exact test evaluating proportion of patients who improved, remained stable, or deteriorated between treatment arms over time
Descriptive summaries	
Summary statistics	Descriptive mean response by treatment group at time t
Change from baseline	Descriptive mean change from baseline by treatment group at time t
Responder analysis	Descriptive proportion of patients who improved, remained stable, or deteriorated between treatment arms by timepoint or over time

PRO=patient-reported outcome.

Table 1: **Classification of PRO statistical analyses**

Data collection

Search strategy and selection criteria

We included new drug applications and biologic licensing applications approved for new lung cancer indications by the FDA Office of Hematology and Oncology Products between Jan 1, 2008, and Dec 31, 2017. MHF, JKR, and JV identified new drug applications and biologic licensing applications using internal FDA databases. Similar to the work by Zhou and colleagues[18] and Sridhara and colleagues[19], we excluded new drug applications and biologic licensing applications involving medical imaging products, dosing and safety labelling revisions, and applications for topically applied dermatological oncology products, Center for Biologics Evaluation and Research products, and Center for Devices and Radiological Health products. For each application, we identified the pivotal trial(s) that were used as the primary basis for FDA approval. Generally, a pivotal trial is considered a trial that is completed immediately before submission of the new drug or biologic licensing application to the FDA and is used to support efficacy claims. We identified PRO instruments in each trial and excluded PROs that were reportedly used for economic models because these analyses are often used to support sponsor submissions to payers rather than regulators and might be reported separately from the study reports and documents sent to the FDA.

We included information regarding PRO reporting and statistical analysis found in the protocol, statistical analysis plan, or clinical study report of the reviewed trials. If discrepancies were found between the documents, information was taken first from the clinical study report and then the statistical analysis plan and then the trial protocol. Two reviewers (MHF and JKR) collected data using a standardised form (appendix pp 1–3), and any disagreements over data extraction were resolved by discussion or with the assistance of a third reviewer (JV).

We report the general characteristics of each trial, including whether it was randomised or single arm, blinding status (blinded or open-label), and the line of therapy (first-line, second-line and beyond, or maintenance). We identified the sample size of the intention-to-treat (ITT) population, defined as the total number of patients randomly assigned; additionally, each PRO instrument and PRO assessment schedule have been included.

Prespecified PRO concepts

For each trial and PRO instrument, we identified whether PRO concepts of interest were prespecified in the protocol or statistical analysis plan. The term PRO concept refers to the event being measured and represents how a patient functions or feels.[3] We evaluated whether a clear rationale was provided for the choice of PRO concept(s) used in each trial and whether the PRO hypotheses were specific enough to inform the statistical method used for the PRO concept.[14] Additionally, we identified whether the prespecified PRO concept was treated as a composite endpoint (when two or more components or concept scores are combined to create a general score), or not, in the analysis.[3,20]

Statistical analyses

We assessed the statistical analyses done for PROs in each trial (table 1). In addition, we evaluated the statistical analyses done for each prespecified PRO concept, domain, and single item by PRO instrument because statistical techniques often varied between instruments within a trial. Analyses also varied by prespecified PRO concept, domain, and single item within each instrument. Single items are defined as an individual question, statement, or task to address a particular PRO concept. Domains are made up of multiple items or domains to address a larger concept.[3]

We classified PRO analyses into four categories: time-to-event analyses (improvement, deterioration, or maintenance), longitudinal analyses, basic inferential tests or general linear models, and descriptive summaries (table 1). Basic inferential tests or general linear models were further classified into continuous or responder analyses. Continuous analyses were then further subcategorised into tests by timepoint or tests of overall score over time. Similarly, responder analyses were also

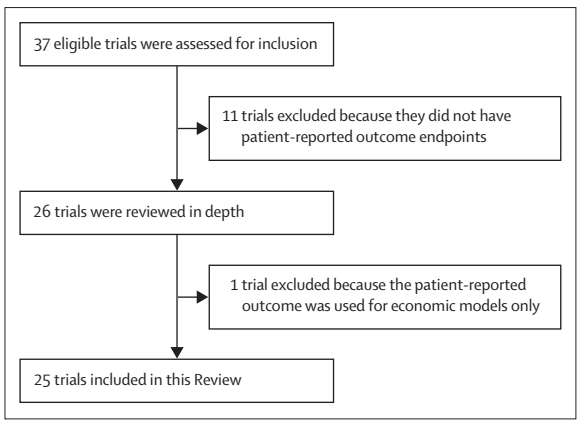

Figure: **Flowchart of selected studies**

subcategorised into tests of proportion of patients with an event by timepoint or overall tests of proportion of patients with an event over time. Summary statistics by timepoint, descriptive change from baseline by timepoint, and descriptive responder analyses were considered descriptive summaries.

In addition, we report whether a prespecified timepoint was used for analysis. We evaluated the PRO analysis populations as they might vary from the ITT population and whether a correction for multiple testing was done at the trial level and whether adjustments were made for testing multiple domains within an instrument.

Missing data and sensitivity analysis

We report whether a missing data method was used for each analysis. An analysis was said to make use of a missing data method if any imputation of missing data was specified. We assessed whether a sensitivity analysis for missing data was done and whether the sensitivity analysis results were reported in the clinical study report.

Instrument completion

Reporting the completion rate of each PRO instrument is part of the FDA's strategy to examine PRO data. In general, PRO completion is defined as the percentage of patients who were still considered on study who had an observed PRO at a specific timepoint.[21] We assessed whether a PRO instrument completion table was reported and whether the definitions of the numerator and the denominator used were reported to compute instrument completion.

Clinical relevance

The threshold for clinical relevance is defined as the clinically meaningful individual patient PRO score change from baseline over time, which should be determined empirically and might vary by study population.[3] We report whether there was a threshold specified for clinical relevance and whether a reference existed that justified the threshold that was used.

	Trials (n=25)
Randomised	
Randomised	21 (84%)
Single-arm	4 (16%)
Blinding status	
Blinded	4 (16%)
Open-label	21 (84%)
Line of therapy	
First	10 (40%)
Second-line and beyond	12 (48%)
Maintenance	3 (12%)
Number of patient-reported outcome instruments included	
1	6 (24%)
2	13 (52%)
3	3 (12%)
4	3 (12%)

Table 2: **General characteristics of pivotal lung cancer trials used to support US Food and Drug Administration drug approval that have reports of patient-reported outcomes between 2008 and 2017**

Findings

Of the 37 pivotal lung cancer trials used to support FDA approval of a product between January, 2008, and December, 2017, 25 (68%) trials reported collecting PROs (figure; table 2). 19 trials (76%) included one or two PRO instruments. The median sample size of the ITT population from the 25 studies was 376 patients (range 87–1253).

Across all of the 25 trials, 53 PRO instruments were included, with 16 studies (64%) including the European Organisation for Research and Treatment of Cancer Quality of Life Questionnaire-core questionnaire (EORTC QLQ-C30),[22] 15 (60%) including the EORTC QLQ-Lung Cancer Module (EORTC QLQ-LC13),[23] eight (32%) including the EuroQoL-5D,[24] seven (28%) including the Lung Cancer Symptom Scale (LCSS),[25] three (12%) including the Visual Symptom Assessment Questionnaire in anaplastic lymphoma kinase,[26] two (8%) including the Functional Assessment of Cancer Therapy-Lung (FACT-L),[27] and two (8%) including the PRO-Common Terminology Criteria for Adverse Events[28] in their analyses (table 2). All of the trials assessed PROs at multiple timepoints and assessments were generally based on the treatment cycle the patient was receiving.

Prespecified PRO concepts

21 (84%) of 25 trials prespecified a PRO concept. Of these 21 trials, six (29%) provided clear rationale for the choice of PRO concept(s) and five (24%) provided a prespecified PRO hypotheses that were specific enough to inform the statistical method for the chosen PRO concept. A total of 28 instruments had prespecified PRO concepts; the most common concepts were cough, dyspnoea, and chest pain from EORTC QLQ-LC13. Nine (32%) of the

	Instruments with prespecified PRO concepts (n=28)*
EORTC QLQ-C30	
Dyspnoea	3 (11%)
Pain	2 (7%)
Fatigue	2 (7%)
Appetite loss	1 (4%)
Cognitive function	1 (4%)
Global health status and quality of life	3 (11%)
EORTC QLQ-LC13	
Cough	5 (18%)
Chest pain	5 (18%)
Dyspnoea	5 (18%)
Pain	5 (18%)
Haemoptysis	2 (7%)
Arm and shoulder pain	4 (14%)
Composite of cough, chest pain, and dyspnoea	4 (14%)
Composite of cough, chest pain, dyspnoea, and arm and shoulder pain	3 (11%)
Composite of cough, dyspnoea, and pain	1 (4%)
LCSS	
Total score	1 (4%)
ASBI	4 (14%)
Each of six symptoms	1 (4%)
Each of nine items	2 (7%)
Global three-item index (symptom distress, interference with activities, and HRQOL)	1 (4%)
Composite of pain, cough, and dyspnoea	1 (4%)
FACT-L	
LCS score	2 (7%)

Data are number of instruments (%) with prespecified PRO concepts. PRO=patient reported outcome. 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. LCSS=Lung Cancer Symptom Scale. ASBI=Average symptom burden index. HRQOL=health-related quality of life. FACT-L=Functional Assessment of Cancer Therapy-Lung. LCS=Lung Cancer Subscale. *More than one PRO concept might have been specified in a single instrument.

Table 3: **Prespecified PRO concepts**

28 instruments with prespecified PRO concepts included a composite endpoint (table 3).

Statistical analyses

24 (96%) of 25 trials provided descriptive summaries, 19 (76%) carried out time-to-event analyses, 12 (48%) did longitudinal analyses, and ten (40%) did basic inferential tests or general linear models. Most of the trials did domain-level and item-level analyses.

Of the 53 instruments included in the 25 trials, 28 (53%) instruments had prespecified PRO concepts, 48 (91%) had domain-level analyses and 47 (89%) had item-level analyses (table 4). Seven (28%) of the 25 trials specified a timepoint for analysis: three (48%) specified 12 weeks, two (29%) specified 15 weeks, and two (29%) specified 6 months. All other analyses were assessed over time or at each PRO scheduled assessment.

Based on our categories of statistical analyses, trials most frequently did three (range one to seven) statistical analyses for a single PRO instrument. The analysis population commonly differed across statistical methods within a trial. Examples of analysis populations included the ITT population, patients with a baseline assessment and at least one post-baseline assessment, and patients who received at least one dose of study medication and completed at least one assessment. Descriptions of the analysis populations were generally unclear or not explicitly detailed for each statistical method.

None of the trials included PRO endpoints in the statistical analysis plan to control for type I error in the formal efficacy analyses. Three of the 25 trials (12%) adjusted for testing multiple domains within an instrument.

Missing data and sensitivity analyses

Two (8%) of the 25 trials used a missing data method in their primary PRO analyses. More specifically, four domain-level and two item-level analyses used multiple imputation for the descriptive responder analyses. It was assumed that all other analyses used all available data. Three (12%) trials reported sensitivity analyses for missing PRO data in the statistical analysis plan. Sensitivity analyses included pattern-mixture models, control-based multiple imputation, and multiple imputation based on reasons for missing data. No results of sensitivity analyses for missing data were reported in the clinical study report.

Instrument completion

20 (80%) of the 25 trials reported a completion rate table for at least one PRO instrument. The amount of detail in the presentation of PRO completion prevalence ranged widely across the trials. For example, one clinical study report included a brief mention of the number of completed PROs, while another included detailed tables of completion rates accompanied by reasons for missingness by instrument, treatment group, and treatment cycle.

The definitions of PRO instrument completion varied between trials, making comparison difficult across trials (table 5). The most common PRO instrument completion numerator definition used was whether a patient completed at least one item on the PRO instrument. The most common denominator definition used for PRO completion was the number of patients with an expected PRO assessment. Four (16%) of the 25 trials did not report an instrument completion definition.

Clinical relevance

A range of different thresholds were used to show clinical relevance by PRO instrument and domain (table 6). All reported thresholds were included in either

time-to-event or responder analyses. Of the 14 trials that reported 10 points for clinical relevance for EORTC QLQ-C30, 12 cited the work by Osoba and colleagues,[29] one of which also cited the further work by Osoba and colleagues,[21] with the other two trials not providing any justification. Of the 12 trials that reported 10 points for clinical relevance for EORTC QLQ-LC13, nine cited Osoba and colleagues,[29] with three not providing justification. Two trials specified a five point threshold and did not provide justification. Of the seven trials that included the LCSS, two trials reported 10 points, with no citation for reference; four reported 15 points, with two citing the work by de Marinis and colleagues[30]—the other two trials did not provide justification; and one reported a 25% change from baseline with no citation for reference. For the two studies investigating FACT-L PROs, both reported thresholds for the Lung Cancer Subscale and Trial Outcome Index and cited Cella and colleagues.[31]

Conclusion

This Review of PRO statistical analysis methods in pivotal lung cancer trials that have supported FDA drug approval between January, 2008, and December, 2017, shows that trials often do not provide a clear rationale for prespecified PRO concepts and that a wide range of statistical techniques were done, resulting from an absence of clear PRO research objectives. The wide-ranging types of PRO analyses make understanding how different therapies might affect clinical outcomes, which are important to clinicians and patients, difficult. Sensitivity analyses for missing data were usually not specified, leaving the potential effect of incomplete forms and missed visits unclear. Additionally, variability in the definitions of instrument completion rates existed, which makes determining the quality of the data, the effect of missing data, and trial comparisons difficult to discern for regulators. Efforts at the FDA are underway to standardise the reporting of completion rates for review of risk–benefit assessment.

Compared with a 2018 systematic review of PRO analyses in 66 advanced breast cancer randomised trials (reported from 2001 to 2017),[9] the trials included in this Review used time-to-event analyses (19 [76%] of 25 lung cancer trials *vs* 6 [9%] of the 66 breast cancer trials) and mixed-models (12 [48%] of the lung cancer trials *vs* 18 [27%] of the breast cancer trials) more frequently.[9] A lower percentage of lung cancer trials (four [16%]) did continuous basic inferential tests or general linear models compared with the breast cancer trials (20 [30%]).[9] As the focus has shifted to include trials that report PROs to support FDA drug approval since 2008, it is possible that our findings might reflect the growing recognition of the need to move beyond descriptive analytical approaches, particularly if labelling is sought. Moreover, a higher percentage of lung cancer trials (23 [92%] of 25) proposed a threshold for clinical relevance compared with the breast cancer trials (28 [42%] of 66), although justification for the specific study population was not always appropriate. For example, the Osoba and colleagues[29] definition was applied beyond the original concepts for which it was examined (physical functioning, social functioning, emotional functioning, and global quality of life on the EORTC QLQ-C30). The threshold for clinical relevance should be assessed in the context of the target population or other characteristics of the trial design.[3]

	Instruments with prespecified PRO concepts (n=28)	Instruments with domain-level analyses (n=48)	Instruments with item-level analyses (n=47)
Time-to-event	23 (82%)	13 (27%)	12 (26%)
Longitudinal analysis	10 (36%)	21 (44%)	17 (36%)
Basic inferential test or general linear model*	7 (25%)	19 (40%)	17 (36%)
Continuous	0	7 (15%)	6 (13%)
By timepoint	0	4 (8%)	4 (9%)
Overall	0	3 (6%)	2 (4%)
Responder analysis	7 (25%)	13 (27%)	12 (26%)
By timepoint	2 (7%)	1 (2%)	1 (2%)
Overall	5 (18%)	12 (25%)	11 (23%)
Descriptive summaries*	15 (54%)	44 (92%)	43 (92%)
Summary statistics	2 (7%)	34 (71%)	31 (66%)
Change from baseline	7 (25%)	31 (65%)	25 (53%)
Responder analysis	9 (32%)	22 (46%)	19 (40%)

Data are number of instruments (%) with analyses of prespecified PRO concepts, domains, and single items. PRO=patient-reported outcome. *Trials may have performed more than one type of analysis in this category.

Table 4: **Statistical analysis of prespecified PRO concepts, domains, and single items of the 25 trials reviewed**

	Trials (n=25)
Completed questionnaire/expected	4 (16%)
Completed questionnaire/did not fill out treatment completion page	1 (4%)
Completed 50% of questions/expected	1 (4%)
Non-missing subscale/expected	2 (8%)
Completed at least one item/expected	4 (16%)
Completed at least one item/all patients filled out instrument patient-reported outcome pages at each cycle	2 (8%)
Completed at least one item/unclear	2 (8%)
Received questionnaire/expected	2 (8%)
Unclear/expected	1 (4%)
Unclear/all patients with baseline and at least one post-baseline assessment	1 (4%)
Unclear/unclear	1 (4%)
Patient-reported outcome instrument completion was not reported	4 (16%)

Data are trials with patient-reported outcomes. The table shows numerator/denominator. Expected=all patients who were expected to have a patient-reported outcome assessment. Unclear=the definition of the numerator or denominator was not clear in the documentation.

Table 5: **Definitions of patient-reported outcome instrument completion**

	Instruments
EORTC QLQ-C30 (n=16)	
10 points	14 (88%)
Not defined	2 (13%)
EORTC QLQ-LC13 (n=15)	
10 points	12 (80%)
5 points	2 (13%)
Not defined	1 (7%)
LCSS (n=7)	
10 points	2 (29%)
15 points	4 (57%)
25%	1 (14%)
FACT-L (n=2)	
2 points (LCS), 6 points (TOI), 6 points (total score)	1 (50%)
3 points (LCS), 6 points (TOI), 6 points (total score)	1 (50%)

Data are n (%). 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. LCSS=Lung Cancer Symptom Scale. FACT-L=Functional Assessment of Cancer Therapy-Lung. LCS=Lung Cancer Subscale. TOI=Trial Outcome Index.

Table 6: **Specified thresholds of clinical relevance for patient-reported outcomes**

An additional concern is the small number of sensitivity analyses done for missing data to assess the effect of assumptions made in the primary analysis of PRO data. Missing data occur frequently in cancer studies, and the amount of missing data might be higher in cancer trials than in other therapeutic areas because of the serious toxicity and morbidity associated with cancer therapies.[11,32,33] Missing data decrease statistical power of a trial and can lead to biased results of patient experience.[10] Because patient experience endpoints are used to help inform decision making for regulators, patients, and clinicians, such bias is concerning because it might have substantial implications for patients. Although many statisticians have recommended doing sensitivity analyses for missing data,[34–36] most trials do not report such analyses for PROs. In addition to minimising missing PRO data and using appropriate methods that handle missing data in the PRO primary analysis, a few suitable sensitivity analyses for missing data should be prespecified to evaluate whether primary PRO results are robust to missing data assumptions.

Furthermore, open-label studies are common in cancer trials, as reflected by our findings and previous assessments of trials to support FDA product approval.[37] Concerns exist that open-label designs might contribute to bias;[3] thus, sensitivity analyses might be especially important in open-label trials. These challenges are increasingly recognised by international drug development stakeholders, and we support ongoing efforts—such as SISAQOL[4,9] and ICH E9 (R1)[20]—that promote the standardisation of analytical methods and greater use of prespecified sensitivity analyses.

Another component of addressing missing data is understanding the extent of missingness. While the description of attrition in the ITT population (missing data) and patients completing their assessments (completion rate) is an important step in a rigorous analysis, our study identified wide variability in the presentation of PRO instrument completion, as well as the definitions used for completion, across the reviewed trials. This variability presents a substantial challenge to cross-trial comparisons, but one that is perhaps easiest to standardise across all cancer trial settings. Indeed, using standardised definitions for and presentation of missing data and completion rates are key goals of initiatives such as SISAQOL. In addition, documenting reasons for missing PRO data can also be informative for assessment and sensitivity analysis.

Aside from completion rate and missing data, no single analysis or data visualisation technique exists that will address all PRO research objectives.[7,11] Thus, the objective of the PRO analysis should be carefully defined within the context of the trial to guide appropriate statistical methodology and interpretation of patient experience.[7,14] Clear definitions of the research objective as well as the analysis population are important for all endpoints, including PROs.[36] Clear specification of the PRO research objective includes providing a clear rationale for PRO concept(s) of interest in a trial, which we found to occur relatively infrequently in the studies under review. Prespecifying PRO concepts might help to address the challenge of multiplicity in hypothesis testing when PROs are included in the statistical hypotheses testing plan to control for type I error.[38] For example, although all the functional concepts of the EORTC QLQ-C30 might be assessed, if the research objective targets maintenance of physical function as the concept of interest it can be prespecified and elevated in the hierarchy of statistical hypotheses testing.

Prespecified PRO concepts might be composite endpoints, which is another strategy for alleviating the issue of multiplicity. However, composite endpoints might present interpretation challenges if the individual components have differences in importance, frequency, or treatment effects.[3,39] Therefore, justification of combining components and examination of the treatment effect on individual components are important to avoid misleading interpretation of patient experience. An example of this is provided in the FDA-approved drug ruxolitinib, where a total symptom score composite was statistically tested, but the component symptoms were also presented in the FDA product label to provide additional information on whether any one symptom was driving the result.[40]

This Review has several limitations. The findings are restricted to pivotal lung cancer trials submitted to the FDA and do not represent all clinical cancer trials. Since the focus on patient-focused drug development has grown, the emphasis of trial results has been placed on

providing more comprehensive feedback on endpoints related to patient experience in cancer trials submitted to the FDA. This subset of cancer trials included those that received clinical and statistical feedback from the FDA during protocol development and this might explain some of the differences found between this Review and previous studies.[9] In addition, we excluded dosing and safety revisions; thus, it is possible that some approvals associated with lung cancer were not included in our analysis. Therefore, the generalisability of these findings to other cancer trials that are not intended to support a marketing application, trials in other tumour types, and non-cancer trials is not clear. However, the issues we identified regarding missing data have been frequently reported in reviews of randomised clinical trials in other tumour types.[41–46]

Despite these limitations our Review did have several strengths. Most previous studies are based on publications[9,35,41–46] or protocols,[47,48] but the data resources available at the FDA allowed for the examination of protocols, statistical analysis plans, and clinical study reports. Furthermore, although PRO completion has been assessed in numerous studies,[48,49] to the best of our knowledge, this Review is the first to examine how PRO completion is defined and to show the wide variability in what could otherwise be considered a relatively standard definition of PRO completion. These findings suggest that publications should include a clear definition of the completion rate used as well as the completion rate, when reporting on this topic.

In conclusion, this Review has identified key issues in the statistical analysis of PRO endpoints in lung cancer trials supporting FDA drug approval. As more trials are likely to include PRO endpoints following the 21st Century Cures Act,[2] which encourages the assessment of patients' experiences when they receive therapy, and pharmaceutical investment in cancer research continues to grow, addressing these issues will be important considerations in future cancer trials.

Contributors

MHF, JKR, and RS conceptualised the search strategy and evaluation process. MHF, JKR, and JV collected the data. MHF led the drafting of the manuscript. JKR, BLK-K, and PGK provided clinical interpretation of the results. RS supervised the development of the entire Review work. All authors provided feedback and reviewed the manuscript.

Declaration of interests

BLK-K reports previous employment by Pharmerit International. All other authors declare no competing interests.

Acknowledgments

We would like to acknowledge Alexander Mackenzie-Siguel for his help extracting data from patient-reported outcome instrument completion tables.

References

1 US Food and Drug Administration. CDER patient-focused drug development. December, 2018. https://www.fda.gov/drugs/developmentapprovalprocess/ucm579400.htm (accessed Dec 1, 2018).

2 114th US Congress. 21st Century Cures Act. Dec 13, 2016. https://www.congress.gov/114/plaws/publ255/PLAW-114publ255.pdf (accessed Dec 20, 2018).

3 US Food and Drug Administration. Guidance for industry: patient-reported outcome measures: use in medical product development to support labeling claims. Oct 24, 2009. http://www.fda.gov/downloads/Drugs/Guidances/UCM193282.pdf (accessed 20 Dec 2018).

4 Bottomley A, Pe M, Sloan J, et al. Analysing data from patient-reported outcome and quality of life endpoints for cancer clinical trials: a start in setting international standards. *Lancet Oncol* 2016; **17:** e510–14.

5 Vodicka E, Kim K, Devine E, Gnanasakthy A, Scoggins J, Patrick D. Inclusion of patient-reported outcome measures in registered clinical trials: evidence from ClinicalTrials.gov (2007–2013). *Contemp Clin Trials* 2015; **43:** 1–9.

6 Kluetz PG, Slagle A, Papadopoulos E, 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:** 1553–58.

7 Kluetz PG, Kanapuru B, Lemery S, et al. Informing the tolerability of cancer treatments using patient-reported outcome measures: summary of an FDA and Critical Path Institute workshop. *Value Health* 2018; **21:** 742–47.

8 Fayers PM, Machin D. Quality of life: the assessment, analysis and interpretation of patient-reported outcomes. Chichester: John Wiley & Sons, 2013.

9 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:** e459–69.

10 Bell ML, Fairclough DL. Practical and statistical issues in missing data for longitudinal patient-reported outcomes. *Stat Methods Med Res* 2014; **23:** 440–59.

11 Fairclough DL. Patient reported outcomes as endpoints in medical research. *Stat Methods Med Res* 2004; **13:** 115–38.

12 Field KM, Jordan JT, Wen PY, Rosenthal MA, Reardon DA. Bevacizumab and glioblastoma: scientific review, newly reported updates, and ongoing controversies. *Cancer* 2015; **121:** 997–1007.

13 Calvert M, Blazeby J, Altman DG, et al. Reporting of patient-reported outcomes in randomized trials: the CONSORT PRO extension. *JAMA* 2013; **309:** 814–22.

14 Calvert M, Kyte D, Mercieca-Bebber R, et al. Guidelines for inclusion of patient-reported outcomes in clinical trial protocols: the SPIRIT-PRO extension. *JAMA* 2018; **319:** 483–94.

15 Ferlay J, Soerjomataram I, Dikshit R, et al. Cancer incidence and mortality worldwide: sources, methods and major patterns in GLOBOCAN 2012. *Int J Cancer* 2015; **136:** e359–86.

16 American Cancer Society. Key statistics for lung cancer. Feb 8, 2018. https://www.cancer.org/cancer/non-small-cell-lung-cancer/about/key-statistics.html (accessed Dec 20, 2018).

17 Torre LA, Bray F, Siegel RL, Ferlay J, Lortet-Tieulent J, Jemal A. Global cancer statistics, 2012. *CA Cancer J Clin* 2015; **65:** 87–108.

18 Zhou J, Vallejo J, Kluetz P, et al. Overview of oncology and hematology drug approvals at US Food and Drug administration between 2008 and 2016. *J Natl Cancer Inst* 2019; **111:** 449–58.

19 Sridhara R, Johnson JR, Justice R, Keegan P, Chakravarty A, Pazdur R. Review of oncology and hematology drug product approvals at the US Food and Drug Administration between July 2005 and December 2007. *J Natl Cancer Inst* 2010; **102:** 230–43.

20 ICH Steering Committee. ICH harmonised tripartite guideline: statistical principles for clinical trials (E9). Feb 5, 1998. https://www.ich.org/fileadmin/Public_Web_Site/ICH_Products/Guidelines/Efficacy/E9/Step4/E9_Guideline.pdf (accessed Dec 21, 2018).

21 Osoba D, Bezjak A, Brundage M, et al. Analysis and interpretation of health-related quality-of-life data from clinical trials: basic approach of the National Cancer Institute of Canada Clinical Trials Group. *Eur J Cancer* 2005; **41:** 280–87.

22 Aaronson NK, Ahmedzai S, Bergman B, et al. The European Organisation for Research and Treatment of Cancer QLQ-C30: a quality-of-life instrument for use in international clinical trials in oncology. *J Natl Cancer Inst* 1993; **85:** 365–76.

23 Bergman B, Aaronson N, Ahmedzai S, Kaasa S, Sullivan M. The EORTC QLQ-LC13: a modular supplement to the EORTC core quality of life questionnaire (QLQ-C30) for use in lung cancer clinical trials. *Eur J Cancer* 1994; **30:** 635–42.

24 EuroQol Group. EuroQol—a new facility for the measurement of health-related quality of life. *Health Policy* 1990; **16:** 199–208.

25 Hollen PJ, Gralla RJ, Kris MG, Potanovich LM. Quality of life assessment in individuals with lung cancer: testing the Lung Cancer Symptom Scale (LCSS). *Eur J Cancer* 1993; **29:** s51–S8.

26 Salgia R, Solomon BJ, Shaw AT, et al. Visual effects in anaplastic lymphoma kinase (ALK)-positive advanced non-small cell lung cancer (NSCLC) patients treated with crizotinib. *Proc Am Soc Clin Oncol* 2017; **30** (suppl_15)**:** 7596.

27 Cella DF, Bonomi AE, Lloyd SR, Tulsky DS, Kaplan E, Bonomi P. Reliability and validity of the functional assessment of cancer therapy—lung (FACT-L) quality of life instrument. *Lung Cancer* 1995; **12:** 199–220.

28 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). *Natl Cancer Inst* 2014; published online Sept 29. DOI:10.1093/jnci/dju244.

29 Osoba D, Rodrigues G, Myles J, Zee B, Pater J. Interpreting the significance of changes in health-related quality-of-life scores. *J Clin Oncol* 1998; **16:** 139–44.

30 De Marinis F, Pereira JR, Fossella F, et al. Lung cancer symptom scale outcomes in relation to standard efficacy measures: an analysis of the phase III study of pemetrexed versus docetaxel in advanced non-small cell lung cancer. *J Thorac Oncol* 2008; **3:** 30–36.

31 Cella D, Eton DT, Fairclough DL, et al. What is a clinically meaningful change on the functional assessment of cancer therapy–lung (FACT-L) questionnaire?: results from Eastern Cooperative Oncology Group (ECOG) Study 5592. *J Clin Epidemiol* 2002; **55:** 285–95

32 Ibrahim JG, Chu H, Chen M-H. Missing data in clinical studies: issues and methods. *J Clinl Oncol* 2012; **30:** 3297.

33 Simes RJ, Greatorex V, Gebski V. Practical approaches to minimize problems with missing quality of life data. *Stat Med* 1998; **17:** 725–37.

34 Little RJ, D'agostino R, et al. The prevention and treatment of missing data in clinical trials. *N Engl J Med* 2012; **367:** 1355–60.

35 Fielding S, Ogbuagu A, Sivasubramaniam S, MacLennan G, Ramsay C. Reporting and dealing with missing quality of life data in RCTs: has the picture changed in the last decade? *Qual Life Res* 2016; **25:** 2977–83.

36 International Council for Harmonisation of Technical Requirements for Pharmaceuticals for Human Use. ICH E9 (R1) technical document: estimands and sensitivity analysis in clinical trials. June 16, 2017. https://www.ich.org/fileadmin/Public_Web_Site/ICH_Products/Guidelines/Efficacy/E9/E9-R1EWG_Step2_Guideline_2017_0616.pdf. (accessed Dec 20, 2018).

37 Kanapuru B, Singh H, Kim J, Kluetz PG. Patient-reported outcomes (PRO) in cancer trials submitted to the FDA from 2012–2015. *Proc Am Soc Clin Oncol* 2017; **35** (suppl_15)**:** e14024.

38 Sloan JA, Dueck AC, Erickson PA, et al. Analysis and interpretation of results based on patient-reported outcomes. *Value Health* 2007; **10:** S106–15.

39 Cordoba G, Schwartz L, Woloshin S, Bae H, Gøtzsche PC. Definition, reporting, and interpretation of composite outcomes in clinical trials: systematic review. *BMJ* 2010; **341:** c3920.

40 US Food and Drug Administration. Jakafi (ruxolitinib) tablets, for oral use initial US approval: 2011. October, 2018. https://www.accessdata.fda.gov/drugsatfda_docs/label/2018/202192s016lbl.pdf (accessed Jan 18, 2019).

41 Dirven L, Taphoorn MJ, Reijneveld JC, et al. The level of patient-reported outcome reporting in randomised controlled trials of brain tumour patients: a systematic review. *Eur J Cancer* 2014; **50:** 2432–48.

42 Efficace F, Feuerstein M, Fayers P, et al. Patient-reported outcomes in randomised controlled trials of prostate cancer: methodological quality and impact on clinical decision making. *Eur Urol* 2014; **66:** 416–27.

43 Efficace F, Jacobs M, Pusic A, et al. Patient-reported outcomes in randomised controlled trials of gynaecological cancers: investigating methodological quality and impact on clinical decision-making. *Eur J Cancer* 2014; **50:** 1925–41.

44 Mercieca-Bebber RL, Perreca A, King M, et al. Patient-reported outcomes in head and neck and thyroid cancer randomised controlled trials: a systematic review of completeness of reporting and impact on interpretation. *Eur J Cancer* 2016; **56:** 144–61.

45 Rees JR, Whale K, Fish D, et al. Patient-reported outcomes in randomised controlled trials of colorectal cancer: an analysis determining the availability of robust data to inform clinical decision-making. *J Cancer Res Clin Oncol* 2015; **141:** 2181–92.

46 Wilson MK, Friedlander ML, Joly F, Oza AM. A systematic review of health-related quality of life reporting in ovarian cancer phase III clinical trials: room to improve. *Oncologist* 2018; **23:** 203–13.

47 Kyte D, Duffy H, Fletcher B, et al. Systematic evaluation of the patient-reported outcome (PRO) content of clinical trial protocols. *PloS One* 2014; **9:** e110229.

48 Mercieca-Bebber R, Friedlander M, Kok P-S, et al. The patient-reported outcome content of international ovarian cancer randomised controlled trial protocols. *Qual Life Res* 2016; **25:** 2457–65.

49 Atherton PJ, Burger KN, Pederson LD, Kaggal S, Sloan JA. Patient-reported outcomes questionnaire compliance in cancer cooperative group trials (Alliance N0992). *Clin Trials* 2016; **13:** 612–20.