OPEN ACCESS

UNIVERSITY OF BRISTOL

Peer reviewed version

University of Bristol - Explore Bristol Research
General rights

Differences in Patient-Reported Outcomes that are Most Frequently Detected in Randomized-Controlled Trials in Patients with Solid Tumours: a Pooled Analysis of 229 trials

Treatment differences in PROs in cancer trials

Johannes M. Giesinger[1], PhD, Jane Blazeby[2], MD, Neil K. Aaronson[3], PhD, Mirjam Sprangers[4], PhD, Peter Fayers[5], PhD, Francesco Sparano[6], MS, Jonathan Rees[2], MD, Amelie Anota[7,8], PhD, Chonghua Wan[9], MD, Mike Pezold[10], MD, Sumit Isharwal[11], MD, Francesco Cottone[6], PhD, Fabio Efficace[6], PhD*, On behalf of the EORTC Quality of Life Group

1. Medical University of Innsbruck, University Hospital of Psychiatry II, Innsbruck, Austria

2. Bristol Centre for Surgical Research and Bristol Biomedical Research Centre, Population Health Sciences, University of Bristol, Bristol, UK

3. Division of Psychosocial Research and Epidemiology, The Netherlands Cancer Institute, The Netherlands.

4. Department of Medical Psychology, Amsterdam University Medical Centers, Amsterdam, The Netherlands.

5. Institute of Applied Health Sciences, University of Aberdeen, Aberdeen, UK

6. Italian Group for Adult Hematologic Diseases (GIMEMA), Data Center and Health Outcomes Research Unit. Rome, Italy

7. Methodology and Quality of Life in Oncology Unit (INSERM UMR 1098), University Hospital of Besançon, Besançon, France.

8. French National Platform Quality of Life and Cancer, Besançon, France.

9. Guangdong Medical University, School of humanities and management, Research Center for Quality of Life and Applied Psychology, Dongguan, PR China.

10. Division of Vascular Surgery, Department of Surgery, New York University Langone Medical Center, New York, NY

11. Department of Urology, University of Virginia, Charlottesville, Virginia, USA.

Corresponding author:
Fabio Efficace, PhD
Head, Health Outcomes Research Unit
Italian Group for Adult Hematologic Diseases (GIMEMA)
GIMEMA Data Center
Via Benevento, 6
00161 - Rome, Italy
Phone: +39 06 441 639831; Fax: +39 06 4402516; E-mail: **f.efficace@gimema.it**

Précis

Our findings emphasize the importance of a multidimensional approach to PRO assessment to most comprehensively capture the overall burden of therapy from the patients' standpoint.

Word count: 3459

Number of pages: 18

Number of figures: 2

Number of tables: 4

Appendix:
Pages: 2
Figures:1
Tables: 1

Abstract

Objectives

Patient-reported Outcome (PRO) measures used in cancer research can assess a number of health domains. Our primary objective was to investigate which broad types of PRO domains, Functional health, Symptoms, and Global Quality of Life (QoL), most frequently yield significant differences between treatments in randomized controlled trials (RCTs).

Methods

Two-hundred-twenty-nine RCTs published between January 2004 and February 2019, conducted on patients diagnosed with most common solid malignancies, and using the EORTC QLQ-C30 questionnaire were considered. Studies were identified systematically using literature searches in key electronic databases. Unlike other PRO measures typically used in RCTs, the scoring algorithm of the multidimensional EORTC QLQ-C30 allows to clearly distinguish the three broad types of PRO domains.

Results

One hundred thirty-four RCTs (58·5%) reported statistically significant differences between treatment arms for at least one of the QLQ-C30 domains. Most frequently differences were reported for two or all three broad types of PRO domains (78 of 134 trials, 58·2%). In particular, 35 trials (26·1%), found significant differences for Symptoms, Functional health, and Global QoL, 24 trials (17·9%) for Symptoms and Functional health, 11 trials (8·2%) for Functional health and Global QoL, and 8 trials (6·0%) for Symptoms and Global QoL. The likelihood of finding a statistically significant difference between treatment arms was not associated with key study characteristics, such as study design (i.e., open-label vs. blinded trials) and industry support.

Conclusions

Our findings emphasize the importance of a multidimensional PRO assessment to most comprehensively capture the overall burden of therapy from the patients' standpoint.

Highlights

- Whilst there is convincing evidence of the value of assessing Patient-Reported Outcomes (PROs) in oncology, there is now ample discussion on "which" PROs should actually be measured in cancer clinical trials. Indeed, PRO measures can assess a number of health domains ranging from individual symptoms to broader health domains, such as functional status and quality of life (QoL).

- We analyzed 229 randomized controlled trials (RCTs) published over the last 15 years conducted in patients diagnosed with most common solid cancer malignancies, which have used the multidimensional EORTC QLQ-C30 questionnaire to assess treatment differences. This PRO measure allows to distinguish the following three broad types of PRO domains: Functional health, Symptoms, and Global QoL. We found that in most trials, differences are typically reported for combinations of functional health, symptoms, and Global QoL.

- Our findings emphasize the importance of a multidimensional approach to PRO assessment in cancer clinical trials to most comprehensively capture the overall burden of therapy from the patients' standpoint.

Background

Major progress has been made in the treatment of cancer, with a remarkable number of new drugs approved since 2012 by the European Medicines Agency (EMA) and the US Food and Drug Administration (FDA)[1]. Many of these clinical achievements stem from randomized controlled trials (RCTs) that are considered the gold standard with which healthcare professionals and policy-makers make decisions regarding treatment efficacy[2].

The number of RCTs that include patient-reported outcomes (PROs)[3] has increased substantially over the last 20 years[4], reflecting their general gain in importance in oncology, also from a regulatory standpoint[5-7]. However, PRO measures can assess a number of health domains ranging from individual or clusters of symptoms (e.g. pain, fatigue, nausea), to functional domains (e.g. physical or social function) and even broader concepts such as general health status and health-related quality of life (HRQoL)[8,9]. The latter is a multi-dimensional concept that the EMA has defined as the patient's subjective perception of the impact of disease and treatment on daily life, physical, psychological and social functioning and well-being[10].

Responding to the increased use of PROs as endpoints in cancer clinical trials, the FDA[3] and EMA[10] have published guidance documents providing a regulatory perspective on PRO measurement. This has also led to extensive discussion about what domains of PROs are the most relevant endpoints in cancer RCTs[1,11-13].

In 2014 a major international initiative led by the U.S. National Cancer Institute (NCI) emphasized the need to harmonize PRO outcome measurement across studies and proposed the use of a core set of symptoms across all trials[14]. Others, have suggested that, in the RCT setting, limiting PRO assessment to symptom endpoints "may provide outcome data sufficient to make decisions about the value of a therapy or to allow judgment about the relative value of one therapy contrasted with another"[15]. More recently, FDA representatives[12,16,17] have recommended to consider physical

function, symptomatic adverse events and disease-related symptoms as a starting point for defining patient-reported endpoints in cancer trials, while encouraging to measure also other aspects of patient experience[18] and to define core clinical outcome sets to inform regulatory authorities, health care providers and patients[17].

However, there is no empirical data with regards to PRO domains that most frequently differ between treatment arms in RCT settings. Such information could be critical in informing the ongoing discussion on the most relevant PRO domains to assess in clinical trials, and the design of future studies that include a PRO component.

Our primary objective was to investigate which *broad* types of PRO domains (i.e., functional health vs. symptoms vs. global QoL) most often yield statistically significant differences between treatment arms in cancer RCTs. Secondary objectives were to evaluate the frequency of such differences for *individual* PRO domains (e.g. physical function, pain, and fatigue) and to evaluate the relationship between frequency of differences and quality of PRO reporting.

Methods

We used the PROMOTION Registry[19] to identify all cancer RCTs published between January 2004 and February 2019 conducted in the most common cancer types[20], i.e. breast, lung, colorectal,

genitourinary, and gynaecological cancers. The PROMOTION Registry is a large database containing all RCTs published from 2004 across a wide range of cancer populations that included at least one PRO, either as a primary or secondary or exploratory endpoint. The trials were systematically identified, mainly through PubMed/Medline, and using ad-hoc key searching strategies for each cancer disease site. We included RCTs that enrolled a minimum of 50 patients and that reported at least one PRO endpoint. Trials comparing screening programs, complementary/alternative medicines and psychosocial interventions were excluded. Details on the search strategy and selection process were documented according to the Preferred Reporting Items for Systematic Reviews and Meta-analysis guidelines (PRISMA)[21].

For each trial selected, two trained reviewers independently extracted information. When disagreement about data extraction occurred, the two reviewers revisited the paper to reconcile any differences and, if necessary, a third reviewer was consulted to resolve any outstanding discrepancy and reach consensus. Each reviewer had a personal password to access the study website (REDCap[22]) and complete a pre-defined electronic data extraction form.

Data extracted from each RCT included: (1) basic trial information (e.g. study location, sample size, disease stage, type of treatments being compared); (2) quality of PRO reporting, based on the International Society for Quality of Life Research (ISOQOL) recommended standards[23]; and (3) PRO instruments used and type of PRO outcomes that yielded statistically significant differences between trial arms.

Outcomes definition and variables examined

Among the generic-cancer PRO instruments that are most frequently used for treatment outcome comparisons in RCT settings, only the multidimensional EORTC QLQ-C30 questionnaire[24] makes a clear distinction between three broad types of PRO domains: 1) Functional health; 2) Symptoms; and 3) Global health status/HRQoL. Other commonly used generic-cancer PRO questionnaires used

in RCTs do not cover all broad types of PRO domains. Therefore, for the purpose of this study, we only considered RCTs using the EORTC QLQ-C30 questionnaire. As our goal was not to evaluate the appropriateness of PRO domain selection in relation to the specific research questions posed in any given trial, we focused on examining the prevalence of statistically significant differences (where applicable) between treatment arms across all domains reported in the publication.

The EORTC QLQ-C30[24] comprises 30 items that cover 15 individual domains: Physical, Role, Social, Emotional and Cognitive functioning, Fatigue, Pain, Nausea/Vomiting, Appetite Loss, Dyspnoea, Sleep Disturbances, Diarrhoea, Constipation, Financial difficulties and Global Health Status/QoL (for the remaining of this article refereed to Global QoL). We have further grouped these individual domains into 3 broad categories: (1) Functional health (i.e., Physical, Role, Social, Emotional, and Cognitive functioning); (2) Symptoms (i.e., Fatigue, Pain, Nausea/Vomiting, Appetite Loss, Dyspnoea, Sleep Disturbances, Diarrhoea, and Constipation); and (3) Global QoL. The Financial difficulties domain was excluded from our analysis as it is distinct from these three broad types of PRO domains.

We also recorded information on *individual* PRO domains that yielded statistically significant differences between arms at any time point during the trial and then categorized the type of statistically significant outcomes observed as: 1) Functional health outcomes only; 2) Symptoms only; 3) Global QoL only; or 4) Mixed outcomes (i.e., including at least two of the previous categories).

Statistical analyses

We used absolute frequencies and proportions to describe the main characteristics of studies and to assess the prevalence of statistically significant differences between treatment arms in *broad* types of PRO domains, as defined above. A difference for a broad type of PRO domain was stated, if at

least one individual PRO domain within that broad type differed between treatment arms. For more detailed information, we also provide descriptive statistics at the level of individual domains.

To account for possible selective PRO outcome reporting, we also performed separate analyses for studies reporting an a priori PRO hypothesis for one or more specific domains versus those that did not. The same approach was adopted to investigate the prevalence of differences in *individual* PRO domains, overall, by cancer site (breast, lung, colorectal, gynaecological and genitourinary) and disease stage across cancer sites (non-metastatic vs metastatic). When assessing prevalence by disease stage, we excluded those studies whose samples included heterogeneous disease stage.

We assessed the quality of reporting of each RCT using the ISOQOL recommended criteria,[23] which are amongst the most comprehensive and highest quality criteria for the rigorous reporting of PROs in RCTs and were also the basis for the development of the CONSORT (CONsolidated Standards Of Reporting in Trials) PRO extension[25]. Based on previous work[26], each item of the ISOQOL criteria was scored "1" if reported and "0" if not, therefore, each trial received a maximum score of 18 (when the PRO was a secondary endpoint) and 29 (when the PRO was a primary endpoint). The ISOQOL checklist score was then divided by the maximum number of relevant items and multiplied by 100 to obtain a score ranging from 0 (worst quality) to 100 (best quality).

For each QLQ-C30 domain, we used multivariable logistic regression analysis, with the reported statistically significant difference between trial arms (yes vs. no) as the dependent variable and the ISOQOL checklist score (continuous) as the main independent variable. The analyses were adjusted for key study characteristics that we expected to possibly influence PRO outcomes. These included : international setting of the study (yes vs. no), being industry supported (yes vs. no), sample size ($n>200$ vs. $n\leq200$),whether PROs were the primary trial outcome (yes vs. no) and study design (open-label vs. blinded trials). We report the corresponding odds ratios and 95% confidence intervals per 10-point increase in the ISOQOL checklist score. All statistical tests were two-sided

and statistical significance was set as $\alpha=0.05$. All analyses were performed with SAS software v.9.4 (SAS Institute Inc., Cary, NC).

Results

Characteristics of clinical trials included in the analysis

Of the 649 RCTs included in the PROMOTION Registry published between January 2004 and February 2019, 262 RCTs used the EORTC QLQ-C30 questionnaire. Of these, we excluded 33 studies that did not report any information on statistical significance for any of the QLQ-C30 domains or did not perform any statistical significance testing. Therefore, 229 RCTs, including a total of 126,262 patients, were included in the current analysis. Details on the selection process are reported in Figure 1.

The majority of trials (n=170, 74.2%) enrolled more than 200 patients and nearly half of RCTs involved patients with metastatic disease (n=114, 49.8%). There were 182 open-label studies, 31 blinded and 16 studies that were considered as unclear. Further details of RCTs are provided in Table 1.

Insert Figure 1 and Table 1

Prevalence of differences by broad type of PRO domains

One hundred thirty-four trials (58.5%) reported statistically significant differences between treatment arms for at least one of the QLQ-C30 domains. The most commonly observed differences were reported for two or all three broad types of PRO domains (78 of 134 trials, 58.2%). In 35 trials (26.1%), significant differences were reported for Symptoms, Functional health, and Global QoL, in 24 trials (17.9%) for Symptoms and Functional health, in 11 trials (8.2%) for Functional health and global QoL, and in 8 trials (6.0%) for Symptoms and Global QoL.

If group differences were reported for only one broad type of PRO domain, these were most frequently reported for: Symptoms (33 trials, 24.6%), followed by Global QoL (14 trials, 10.5%)

and Functional health (9 trials, 6.7%) (Figure 2). Prevalence of statistically significant differences between treatment arms was similar when considering RCTs reporting an a priori PRO hypothesis (N=27) versus those that did not (N=107). For example, the prevalence of differences reported for more than one type of PRO domains was the same, that is 58% of studies (data are reported in online supplementary Figure 1).

Insert Figure 2

Prevalence of differences for individual PRO domains

Focusing on individual PRO domains, the five most frequently observed differences were for: Global QoL (29.7% of the trials), Physical Functioning (19.2%), Fatigue (18.8%), Nausea/Vomiting (18.3%), and Role Functioning (17.0%). In contrast, Constipation (8.7%) and Sleep Disturbances (10.5%) were least often observed as being significantly different between treatment arms (Table 2). Similar percentages were found by studies reporting an a priori hypothesis versus those that did not (data not shown).

Insert Table 2

In the 114 trials of patients with metastatic disease, the top three most frequent PRO differences were reported for: Global QoL (29.8%), Physical Functioning (23.7%) and Pain (19.3%). In the 68 trials of non-metastatic patients, these were found for: Global QoL (27.9%), Fatigue (25.0%) and Social Functioning (23.5%). The most pronounced differences in PROs between trials of metastatic versus non-metastatic patients were found for Pain (8.8% of trials with differences in non-metastatic

patients, vs 19.3% in metastatic patients), and Social Functioning (23.5% non-metastatic vs 13.2% metastatic) (Table 3).

For descriptive purposes a comparison of prevalence of PRO differences by diagnostic group was also performed and results are reported in online supplementary Table 1.

Insert Table 3

Differences in individual PRO domains by the quality of PRO reporting

In a multivariable analysis, we investigated the association between the presence of statistically significant PRO differences in the RCTs and the quality of PRO reporting and found a statistically significant association in 6 out of the 14 QLQ-C30 domains analyzed.

The top three strongest associations with quality of PRO reporting in terms of odds ratios (ORs) were observed for: Sleep Disturbances (OR 1.30, p=0.01), Nausea/Vomiting (OR=1.25, p=0.01) and Cognitive Functioning (OR 1.24, p=0.03). No significant association was found between Global QoL and the quality of PRO reporting (OR=1.00, p=0.98) (Table 4).

The likelihood of finding a statistically significant difference between treatment arms was not associated with RCT study design (i.e., open-label vs. blinded trials), industry support, type of PRO endpoint (primary vs. secondary) and study location (international vs. national) for any of the 14 individual PRO domains analyzed.

Insert Table 4

Discussion

In approximately 60% of the oncology RCTs reviewed, there was a significant PRO difference between treatment arms, and these differences were reported for two or three *broad* types of PRO domains, i.e. Functional health and/or Symptoms, and/or Global QoL. Also, the top three *individual*

PRO domains differing most frequently between treatment arms across all RCTs were: Global QoL, Physical Functioning and Fatigue.

Our results build on previous work concerned with the choice of PRO domains to be used as endpoints in cancer trials. As noted earlier, the NCI-driven initiative to harmonize trial outcomes[14] recommended a core set of 12 patient-reported symptoms for cancer trials based on a literature review, data from clinical trials and a consensus process. Following this, the FDA[12,16] has recommended that PROs focus on symptomatic adverse events, disease-related symptoms and measures of physical function. In contrast, the value added by broad concepts such as HRQoL and psychosocial domains, has been questioned as being too distal from the core business of cancer treatment, and thus of cancer clinical trials[12] (i.e., that these domains may be influenced by factors not directly related to cancer treatment, resulting in low sensitivity to treatment effects). However, we found that functional health domains and Global QoL were among the PROs that most often yielded statistically significant differences between treatment arms in oncology RCTs. Thus, our analysis provides empirical evidence for the sensitivity of functional health outcomes and Global QoL, and hence our findings challenge the assumption that such outcomes are too distal to be of sustained value in trial-based PRO assessments.

Our findings broadly support the approach taken by the EMA between 2012 and 2016 for approving PRO labelling in oncology treatments[1]. For about three-quarters of indications, such labelling mentioned symptom domains, while two-thirds referred to broad concepts such as global health status or QoL. Interestingly, no PRO labelling was granted by the FDA for cancer drugs in the same period, with FDA concerns mostly related to study design, magnitude of differences between treatment arms, validity of the measures, and missing data[1]. Across all medical fields, the FDA has approved PRO-related labelling claims less often than the EMA, and those claims granted were mostly related to symptoms. In contrast, the EMA has more frequently approved QoL claims[27]. The different regulatory policies of the EMA and FDA reflect the ongoing discussion of what PRO

domains to measure in cancer trials, and consequently what information should be made available to regulatory authorities, health care providers and patients.

Although we did not observe a significant association between the quality of PRO reporting and the likelihood that they would yield differences in Global QoL, this was the case for a number of functional health and symptom domains. Specifically, RCTs with lower quality of PRO reporting tended to yield fewer significant PRO outcomes. This suggests that in trials with low quality reporting, real differences in certain PRO domains may be underreported, resulting in an underestimation of the impact of trial-based treatments on patients' functioning and symptom burden (e.g., pain, fatigue and nausea). The finding that the prevalence of Global QoL differences was not associated with the quality of reporting may indicate that this domain is less prone to underreporting than the other domain, which may explain in part the higher frequency of observed differences for this domain. Underreporting of PRO data from oncology RCTs has been identified as a major problem in a previous study by Kyte et al.[28] who showed that about one-third of trials fail to publish their PRO results at all. Our analysis further extends this previous analysis by indicating that even if PRO results from a trial are published, the reporting of actual PRO results may be incomplete and selective.

We also found that some key RCT characteristics, such as industry support or open-label vs. blinded trials, were not associated with the likelihood finding a statistically significant difference between treatment arms. For example, despite some concerns expressed by regulators with regard to the potential bias of PRO results stemming from open-label trials[1], our findings do not seem to support this view. While there is paucity of empirical data in this area, and more in depth and comprehensive analyses are needed to fully consider the impact of blinding on PRO results, our current findings are in line with a recent structured review by Atkinson and colleagues[29], who concluded that inadvertent unblinding should not be considered as a meaningful source of bias.

Since the focus of our study was on the comparison of broad types of PRO domains, we restricted our analysis to those trials using the QLQ-C30, as this is the only measure that distinguishes clearly between functional health, symptoms and Global QoL, thus allowing such a thorough comparison of these three broad types of PRO domains. Due to the widespread use of the QLQ-C30, we could still include a large number of trials covering the main cancer diagnoses in our analysis. Assuming that the trials using the QLQ-C30 do not differ systematically from other trials, we think that our findings can be generalized and are not affected by a selection bias regarding trial setting.

Our analysis has several limitations. We relied on statistically significant differences between treatment arms and did not investigate the clinical relevance of the observed differences. However, we did not attempt such an evaluation as this would have been very difficult given the range of approaches and lack of consensus on how best to define and assess clinical significance[30]. Also, despite the fact that our analysis of prevalence of differences in studies reporting an a priori PRO hypothesis versus those that did not, confirmed that PRO differences were similar across multiple broad types of PRO domains, we cannot rule out the possibility that some trials used a selective approach by only reporting a subset of PRO domains, even though a priori hypotheses were posed in the manuscript.

Conclusions

In conclusion, we report the first evidence-based information regarding the frequency and the type of PRO differences that are observed between treatment arms in oncology RCTs. In most trials, such differences are typically reported for combinations of functional health, symptoms, and Global QoL, thereby highlighting the importance of a multidimensional approach to PRO assessment to most comprehensively capture the overall burden of therapy from the patients' standpoint.

References

1. Gnanasakthy A, Barrett A, Evans E, D'Alessio D, Romano CD. 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.
2. Downing NS, Aminawung JA, Shah ND, Krumholz HM, Ross JS. Clinical trial evidence supporting FDA approval of novel therapeutic agents, 2005-2012. *JAMA.* 2014;311(4):368-377.
3. US Department of Health and Human Services Food and Drug Administration. Guidance for industry: patient-reported outcome measures: use in medical product development to support labeling claims. 2009. Accessed May 6, 2019 http://www.fda.gov/downloads/Drugs/GuidanceComplianceRegulatoryInformation/Guidances/UCM193282.pdf
4. Efficace F, Fayers P, Pusic A, et al. Quality of patient-reported outcome reporting across cancer randomized controlled trials according to the CONSORT patient-reported outcome extension: A pooled analysis of 557 trials. *Cancer.* 2015;121(18):3335-3342.
5. 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.
6. Basch E. Toward patient-centered drug development in oncology. *N Engl J Med.* 2013;369(5):397-400.
7. Kluetz PG, O'Connor DJ, Soltys K. Incorporating the patient experience into regulatory decision making in the USA, Europe, and Canada. *Lancet Oncol.* 2018;19(5):e267-e274.
8. Velikova G, Stark D, Selby P. Quality of life instruments in oncology. *Eur J Cancer.* 1999;35(11):1571-1580.
9. Soni MK, Cella D. Quality of life and symptom measures in oncology: an overview. *Am J Manag Care.* 2002;8(18 Suppl):S560-573.
10. European Medicine Agency. The use of patient-reported outcome (PRO) measures in oncology studies. Appendix 2 to the guideline on the evaluation of anticancer medicinal products in man. 2016. http://www.ema.europa.eu/docs/en_GB/document_library/Other/2016/04/WC500205159.pdf Accessed July 9, 2017
11. Groenvold M, Aaronson NK, Darlington AE, et al. Focusing on Core Patient-Reported Outcomes in Cancer Clinical Trials-Letter. *Clin Cancer Res.* 2016;22(22):5617.
12. Kluetz PG, Slagle A, Papadopoulos EJ, et al. Focusing on Core Patient-Reported Outcomes in Cancer Clinical Trials: Symptomatic Adverse Events, Physical Function, and Disease-Related Symptoms. *Clin Cancer Res.* 2016;22(7):1553-1558.
13. Basch E, Geoghegan C, Coons SJ, et al. Patient-Reported Outcomes in Cancer Drug Development and US Regulatory Review: Perspectives From Industry, the Food and Drug Administration, and the Patient. *JAMA Oncol.* 2015;1(3):375-379.
14. Reeve BB, Mitchell SA, Dueck AC, et al. Recommended patient-reported core set of symptoms to measure in adult cancer treatment trials. *J Natl Cancer Inst.* 2014;106(7).
15. Cleeland CS. Symptom burden: multiple symptoms and their impact as patient-reported outcomes. *J Natl Cancer Inst Monogr.* 2007(37):16-21.
16. Kluetz PG, Papadopoulos EJ, Johnson LL, et al. Focusing on Core Patient-Reported Outcomes in Cancer Clinical Trials-Response. *Clin Cancer Res.* 2016;22(22):5618.
17. US Food and Drug Administration. FDA-ASCO Public Workshop: 2019 Clinical Outcome Assessments in Cancer Clinical Trials Fourth Annual Workshop. https://www.fda.gov/drugs/news-events-human-drugs/fda-asco-public-workshop-2019-clinical-outcome-assessments-cancer-clinical-trials-fourth-annual. Accessed January 20, 2020.

18. 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 in health : the journal of the International Society for Pharmacoeconomics and Outcomes Research.* 2018;21(6):742-747.

19. Efficace F, Rees J, Fayers P, et al. Overcoming barriers to the implementation of patient-reported outcomes in cancer clinical trials: the PROMOTION Registry. *Health Qual Life Outcomes.* 2014;12:86.

20. Ferlay J, Ervik M, Lam F, Colombet M, Mery L, Piñeros M, Znaor A, Soerjomataram I, Bray F (2018). Global Cancer Observatory: Cancer Today. Lyon, France: International Agency for Research on Cancer. Available from: https://gco.iarc.fr/today, accessed (June 24, 2019).

21. Moher D, Liberati A, Tetzlaff J, Altman DG. Preferred reporting items for systematic reviews and meta-analyses: the PRISMA statement. *Int J Surg.* 2010;8(5):336-341.

22. Harris PA, Taylor R, Thielke R, Payne J, Gonzalez N, Conde JG. Research electronic data capture (REDCap)--a metadata-driven methodology and workflow process for providing translational research informatics support. *J Biomed Inform.* 2009;42(2):377-381.

23. Brundage M, Blazeby J, Revicki D, et al. Patient-reported outcomes in randomized clinical trials: development of ISOQoL reporting standards. *Qual Life Res.* 2013;22(6):1161-1175.

24. Aaronson NK, Ahmedzai S, Bergman B, et al. The European Organization for Research and Treatment of Cancer QLQ-C30: a quality-of-life instrument for use in international clinical trials in oncology. *J Natl Cancer Inst.* 1993;85(5):365-376.

25. Calvert M, Blazeby J, Altman DG, Revicki DA, Moher D, Brundage MD. Reporting of patient-reported outcomes in randomized trials: the CONSORT PRO extension. *JAMA.* 2013;309(8):814-822.

26. 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(3):416-427.

27. DeMuro C, Clark M, Doward L, Evans E, Mordin M, Gnanasakthy A. Assessment of PRO label claims granted by the FDA as compared to the EMA (2006-2010). *Value Health.* 2013;16(8):1150-1155.

28. 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.

29. Atkinson TM, Wagner J-S, Basch E. Trustworthiness of Patient-Reported Outcomes in Unblinded Cancer Clinical Trials. *JAMA oncology.* 2017;3(6):738-739.

30. Ousmen A, Touraine C, Deliu N, et al. Distribution- and anchor-based methods to determine the minimally important difference on patient-reported outcome questionnaires in oncology: a structured review. *Health Qual Life Outcomes.* 2018;16(1):228.

Table 1: Characteristics of cancer clinical trials included in the analysis (N=229)

Trial characteristic		N	%
International (if more than one country)	No	134	58.5
	Yes	95	41.5
Industry supported (fully or in part)	No	99	43.2
	Yes	130	56.8
Disease stage	Metastatic	114	49.8
	Non-metastatic	68	29.7
	Both	37	16.2
	Unclear	10	4.3
Overall study sample size	≤200	59	25.8
	>200	170	74.2
Year of publication	2004-2010	127	55.5
	2011-2019	102	44.5
Secondary paper on PRO	No	155	67.7
	Yes	74	32.3
Type of treatment*	Radiotherapy	36	15.7
	Surgery	23	10.0
	Chemotherapy	148	64.6
	Targeted therapy	39	17.0
	Hormonal therapy	27	11.8
	Other	26	11.4
PRO endpoint	Primary	40	17.5
	Secondary or exploratory	189	82.5

*More than one option could be chosen

Table 2: Frequency of differences in individual PRO domains in cancer RCTs (N=229)

QLQ-C30 domains	Differences between treatment arms	
	N	%
Functioning domains		
Physical Functioning	44	19.2%
Role Functioning	39	17.0%
Social Functioning	37	16.2%
Emotional Functioning	31	13.5%
Cognitive Functioning	25	10.9%
Global Quality of Life	68	29.7%
Symptom domains		
Fatigue	43	18.8%
Pain	35	15.3%
Nausea/Vomiting	42	18.3%
Sleep Disturbances	24	10.5%
Dyspnoea	30	13.1%
Appetite Loss	33	14.4%
Constipation	20	8.7%
Diarrhoea	33	14.4%

RCT: randomized controlled trials.

Percentages refer to the total number of RCTs using the EORTC QLQ-C30 (N=229)

Table 3: Results from trials including EORTC QLQ-C30 domains as study endpoints in patients with non-metastatic (68 trials) and metastatic (114 trials) disease

QLQ-C30 domains	Differences between treatment arms			
	Non-metastatic		Metastatic	
	N	%	N	%
Functioning domains				
Physical Functioning	10	14.7%	27	23.7%
Role Functioning	13	19.1%	20	17.5%
Social Functioning	16	23.5%	15	13.2%
Emotional Functioning	8	11.8%	17	14.9%
Cognitive Functioning	8	11.8%	13	11.4%
Global Quality of Life	19	27.9%	34	29.8%
Symptom domains				
Fatigue	17	25.0%	20	17.5%
Pain	6	8.8%	22	19.3%
Nausea/Vomiting	12	17.7%	21	18.4%
Sleep Disturbances	11	16.2%	11	9.7%
Dyspnoea	9	13.2%	17	14.9%
Appetite Loss	11	16.2%	15	13.2%
Constipation	7	10.3%	11	9.7%
Diarrhoea	8	11.8%	16	14.0%

Table 4: Association between quality of PRO reporting and prevalence of differences for individual PRO domains

	Fatigue		Pain		Nausea/Vomiting		Sleep disturbances		Dyspnoea		Appetite loss		Constipation		Diarrhoea	
	OR (95% CI)	P value	OR (95% CI)	P value	OR (95% CI)	P value	OR (95% CI)	P value	OR (95% CI)	P value	OR (95% CI)	P value	OR (95% CI)	P value	OR (95% CI)	P value
Isoqol score[1]	1.23 (1.05-1.44)	0.01	1.21 (1.02-1.43)	0.03	1.25 (1.06-1.47)	0.01	1.30 (1.05-1.60)	0.01	1.15 (0.97-1.36)	0.12	1.17 (0.99-1.39)	0.06	1.06 (0.86-1.31)	0.59	1.12 (0.95-1.33)	0.17
International	1.73 (0.78-3.87)	0.18	1.45 (0.61-3.43)	0.40	1.58 (0.70-3.56)	0.27	1.13 (0.40-3.15)	0.82	1.64 (0.67-4.06)	0.28	1.70 (0.71-4.09)	0.24	1.08 (0.37-3.15)	0.89	1.11 (0.47-2.60)	0.81
Industry supported	0.78 (0.35-1.74)	0.55	1.51 (0.62-3.68)	0.37	1.64 (0.73-3.67)	0.23	1.34 (0.50-3.59)	0.56	0.56 (0.23-1.37)	0.20	0.78 (0.33-1.84)	0.57	0.87 (0.32-2.42)	0.79	1.23 (0.51-2.95)	0.64
Sample size >200	4.84 (1.32-17.66)	0.02	2.48 (0.74-8.28)	0.14	1.25 (0.45-3.51)	0.67	1.11 (0.34-3.57)	0.87	2.41 (0.65-8.96)	0.19	3.20 (0.86-11.89)	0.08	1.32 (0.38-4.58)	0.66	3.34 (0.91-12.33)	0.07
PRO primary	1.26 (0.45-3.53)	0.66	1.21 (0.40-3.62)	0.74	0.34 (0.09-1.27)	0.11	1.18 (0.36-3.83)	0.79	0.33 (0.07-1.56)	0.16	1.07 (0.34-3.35)	0.91	0.84 (0.21-3.36)	0.80	0.78 (0.23-2.60)	0.68
Blinded[2]	0.61 (0.19-2.0)	0.41	2.06 (0.77-5.49)	0.15	0.22 (0.05-1.03)	0.06	0.50 (0.11-2.34)	0.37	0.83 (0.22-3.10)	0.78	0.62 (0.17-2.30)	0.47	0.73 (0.16-3.45)	0.69	1.54 (0.54-4.37)	0.42

	Global QoL		Physical Functioning		Role Functioning		Social Functioning		Emotional Functioning		Cognitive Functioning	
	OR (95% CI)	P value	OR (95% CI)	P value	OR (95% CI)	P value	OR (95% CI)	P value	OR (95% CI)	P value	OR (95% CI)	P value
Isoqol score[1]	1.00 (0.88-1.14)	0.97	1.20 (1.03-1.40)	0.02	1.08 (0.93-1.26)	0.33	1.17 (0.99-1.38)	0.07	1.03 (0.87-1.23)	0.72	1.23 (1.02-1.50)	0.03
International	1.87 (0.96-3.65)	0.07	1.60 (0.73-3.51)	0.24	1.12 (0.50-2.50)	0.78	1.23 (0.53-2.86)	0.63	1.46 (0.59-3.59)	0.42	2.38 (0.85-6.67)	0.10
Industry supported	0.84 (0.44-1.61)	0.60	1.34 (0.62-2.90)	0.46	0.99 (0.45-2.15)	0.97	1.22 (0.53-2.79)	0.64	1.09 (0.45-2.65)	0.85	0.99 (0.37-2.64)	0.98
Sample size >200	2.13 (0.92-4.94)	0.08	1.48 (0.56-3.90)	0.43	1.82 (0.67-4.95)	0.24	2.69 (0.89-8.14)	0.08	1.76 (0.57-5.42)	0.32	1.19 (0.34-4.20)	0.79
PRO primary	0.64 (0.25-1.66)	0.36	1.01 (0.37-2.75)	0.99	0.74 (0.25-2.21)	0.59	2.15 (0.81-5.75)	0.13	1.55 (0.52-4.66)	0.44	1.00 (0.28-3.56)	1.00
Blinded[2]	0.96 (0.39-2.32)	0.92	0.69 (0.24-2.02)	0.50	0.71 (0.23-2.25)	0.56	0.91 (0.30-2.73)	0.87	1.60 (0.57-4.53)	0.37	0.78 (0.20-2.98)	0.72

[1]OR (odds ratios) are given per 10 points on the quality metric; values above 1 indicate a higher frequency of differences in high-quality trials (i.e. lower frequency in low quality trials). [2]For the purpose of this analysis, trials evaluated as "unclear" were grouped into one category along with open-label studies.

Figure 1: Schematic breakdown of literature search results of Randomised Controlled Trials (Preferred Reporting Items for Systematic Reviews and Meta-analysis).

Figure 2: Differences between treatment arms in cancer RCTs (N=134) by broad type of PRO domain

Legend: Percentages refer to the total number of RCTs using the EORTC QLQ-C30 with significant differences (N=134)