Statistical analysis of patient-reported outcome data in randomised controlled trials of locally advanced and metastatic breast cancer: a systematic review

Madeline Pe, Lien Dorme, Corneel Coens, Ethan Basch, Melanie Calvert, Alicyn Campbell, Charles Cleeland, Kim Cocks, Laurence Collette, Linda Dirven, Amylou C Dueck, Nancy Devlin, Hans-Henning Flechtner, Carolyn Gotay, Ingolf Griebsch, Mogens Groenvold, Madeleine King, Michael Koller, Daniel C Malone, Francesca Martinelli, Sandra A Mitchell, Jammbe Z Musoro, Kathy Oliver, Elisabeth Piault-Louis, Martine Piccart, Francisco L Pimentel, Chantal Quinten, Jaap C Reijneveld, Jeff Sloan, Galina Velikova, Andrew Bottomley, for the Setting International Standards in Analyzing Patient-Reported Outcomes and Quality of Life Endpoints Data Consortium (SISAQOL)

Although patient-reported outcomes (PROs), such as health-related quality of life, are important endpoints in randomised controlled trials (RCTs), there is little consensus about the analysis, interpretation, and reporting of these data. We did a systematic review to assess the variability, quality, and standards of PRO data analyses in advanced breast cancer RCTs. We searched PubMed for English language articles published in peer-reviewed journals between Jan 1, 2001, and Oct 30, 2017. Eligible articles were those that reported PRO results from RCTs of adult patients with advanced breast cancer receiving anti-cancer treatments with reported sample sizes of at least 50 patients—66 RCTs met the selection criteria. Only eight (12%) RCTs reported a specific PRO research hypothesis. Heterogeneity in the statistical methods used to assess PRO data was observed, with a mixture of longitudinal and cross-sectional techniques. Not all articles addressed the problem of multiple testing. Fewer than half of RCTs (28 [42%]) reported the clinical significance of their findings. 48 (73%) did not report how missing data were handled. Our systematic review shows a need to improve standards in the analysis, interpretation, and reporting of PRO data in cancer RCTs. Lack of standardisation makes it difficult to draw robust conclusions and compare findings across trials. The Setting International Standards in the Analyzing Patient-Reported Outcomes and Quality of Life Data Consortium was set up to address this need and develop recommendations on the analysis of PRO data in RCTs.

Lancet Oncol 2018; 19: e459–69

European Organisation for Research and Treatment of Cancer, Brussels, Belgium (M Pe PhD, L Dorme MSc, C Coens MSc, L Collette PhD, F Martinelli MSc, J Z Musoro PhD, A Bottomley PhD); **Lineberger Comprehensive Cancer Center, University of North Carolina, Chapel Hill, NC, USA** (E Basch MD); **Centre for Patient Reported Outcomes Research, Institute of Applied Health Research, College of Medical and Dental Sciences, University of Birmingham, Birmingham, UK** (Prof M Calvert PhD); **Genentech, a member of the Roche Group, San Francisco, CA, USA** (A Campbell MPh, E Piault-Louis PharmD); **Department of Symptom Research, The University of Texas MD Anderson Cancer Center, Houston, TX, USA** (Prof C Cleeland PhD); **Adelphi Values, Bollington, Cheshire, UK** (K Cocks PhD); **Leiden University Medical Center and Haaglanden Medical Center, Leiden and The Hague, Netherlands** (L Dirven PhD); **Alliance Statistics and Data Center, Mayo Clinic, Scottsdale, AZ, USA** (A C Dueck PhD); **Office of Health Economics, London, UK** (N Devlin PhD); **Clinic for Child and Adolescent Psychiatry and Psychotherapy, University of Magdeburg, Magdeburg, Germany** (Prof H-H Flechtner MD); **School of Population and Public Health, University of British Columbia, Vancouver, BC, Canada** (Prof C Gotay PhD); **Boehringer Ingelheim International GmbH, Ingelheim, Germany** (I Griebsch PhD); **Department of Public Health and Bispebjerg Hospital, University of Copenhagen, Copenhagen, Denmark** (Prof M Groenvold MD); **School of Psychology and Sydney**

Introduction

In a breakthrough report,[1] the US Institute of Medicine highlighted patient-centred care as a crucial component of quality health care. Patient-centred care is defined as "providing care that is respectful of and responsive to individual patient preferences, needs, and values and ensuring that patient values guide all clinical decisions".[1] The incorporation of patient-reported outcomes (PROs) in randomised controlled trials (RCTs) is one way of responding to this imperative. Increasingly, PRO endpoints are being included in RCTs to assess clinical benefit alongside overall and progression-free survival.[2] A PRO is any outcome that is reported directly by the patient.[3,4] By including PRO endpoints, such as health-related quality of life (HRQOL), the patient's perspective is included in clinical assessment, providing improved patient information and supporting shared decision making in the development of new therapies.[5,6]

However, the lack of standards and clear guidelines on how these patient-reported data should be analysed and interpreted in RCTs diminishes their recognised and important value by making it difficult to co mpare results across trials and draw conclusions about the patient experience of new types of cancer treatment.[7] Data generated from specific PROs such as HRQOL are complex because they are either multidimensional, with several subscales to characterise patients' symptoms and their effect on a spects o f p atient f unctioning; r equire repeated measurements to adequately capture changes in these outcomes; and are prone to missing data since obtaining complete PRO follow-up data from all randomised patients can be difficult.[8,9] Inappropriate handling of these crucial statistical issues could bias findings and lead to inaccurate conclusions drawn. Current guidelines do not provide concrete suggestions on how to deal with statistical issues concerning PROs and need to be supplemented with more detailed strategies on how to address these concerns.[3,10]

The Setting International Standards in Analyzing Patient-Reported Outcomes and Quality of Life Endpoints Data for Cancer Clinical Trials (SISAQOL) consortium was established to respond to a clear need to develop standards, guidelines, and recommendations for the analysis of PRO data in cancer RCTs. This consortium is made up of a wide range of international experts, including leading PRO researchers and statisticians, and key individuals from international oncological and medical societies, advisory and regulatory bodies, academic societies, the pharmaceutical industry, cancer institutes, and patient advocacy organisations.[11] A key task identified by the consortium was to undertake systematic literature reviews to describe the current state of PRO analyses in RCTs of cancer treatment. In this systematic review, we examine how analyses of PROs (such as HRQOL) are done in RCTs, using anti-cancer treatments for advanced breast cancer as an example set of trials that are commonly seen in the literature. Since maintaining HRQOL is important in the care of patients with advanced breast cancer, we expected a considerable number of advanced breast cancer RCTs to have PROs in their assessments.[12]

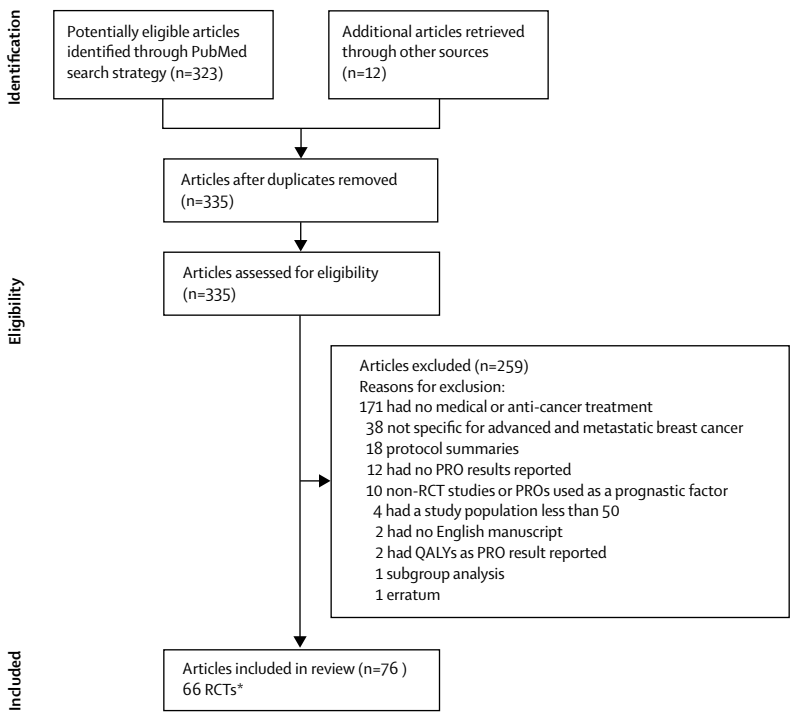

Figure: **Flowchart for the inclusion and exclusion of RCTs**
RCT=randomised controlled trial. PRO=patient-reported outcome. QALY=quality-adjusted life-year. *RCTs including patients with advanced and metastatic breast cancer older than 18 years, sample size more than 50, receiving regular oncology treatment, includes health-related quality of life data and PRO assessment, English manuscript, published between January, 2001, and October, 2017.

Medical School, University of Sydney, Sydney, NSW, Australia (Prof M King PhD); **Center for Clinical Studies, University Hospital Regensburg, Regensburg, Germany** (M Koller PhD); **College of Pharmacy, University of Arizona, Tucson, AZ, USA** (Prof D C Malone PhD); **Outcomes Research Branch, Healthcare Delivery Research Program, Division of Cancer Control and Population Sciences, National Cancer Institute, Bethesda, MD, USA** (S A Mitchell PhD); **International Brain Tumour Alliance, Surrey, UK** (K Oliver BA); **Institut Jules Bordet, Université Libre de Bruxelles, Brussels, Belgium** (M Piccart MD); **Blueclinical Phase I, Porto, Portugal** (F L Pimentel MD); **Centro de Estudos e Investigação em Saúde da Universidade de Coimbra, Coimbra, Portugal** (F L Pimentel); **European Centre for Disease Prevention and Control, Surveillance and Response Support Unit, Epidemiological Methods**

Data collection

Search strategy and selection criteria

We used the methodology detailed in the guidelines for the Cochrane Handbook for Systematic Reviews of Interventions[13] and the results of this systematic review are reported in accordance with PRISMA guidelines.[14] We did not publish a review protocol for this study. We did a literature search in PubMed on March 30, 2016 (updated on Feb 7, 2018) with the following keywords: (quality of life[MeSH Terms] OR quality of life[Text Word] OR patient reported outcomes[Text Word]) AND (advanced[All Fields] OR metastatic[All Fields]) AND breast cancer[Text Word] AND (Randomized Controlled Trial) AND (breast neoplasm[MeSH Terms]) AND (Clinical Trial[ptyp] AND ("2001/01/01"[PDat]: "2017/10/30"[PDat]) AND Humans[MeSH]). We identified 323 potentially eligible articles using this search strategy and checked the references of these publications for additional articles. Additionally, we did a Web of Science search on April 22, 2018 but no further articles were found.

The inclusion and exclusion criteria for the RCTs were similar to those of Ghislain and colleagues.[15] The inclusion criteria were articles that reported PRO findings from RCTs involving adult patients (≥18 years of age) with advanced breast cancer receiving anti-cancer treatments (chemotherapy, targeted therapy, endocrine therapy) with sample sizes of at least 50 patients. Advanced breast cancer refers to either metastatic breast cancer or locally advanced breast cancer (as defined in the European School of Oncology-European Society of Medical Oncology international consensus guidelines).[12] We only included articles published in peer-reviewed journals between January, 2001, and October, 2017, regardless of the start or completion date of the study. We originally considered including articles published from 1997 to review articles from a 20-year period; however, we decided to only include articles from 2001 because of the difficulty of retrieving electronic versions of all identified articles published before 2001.

Exclusion criteria were any RCTs that evaluated psychological, supportive, or supplementary interventions. Supplementary treatments were defined as any other interventions that did not include anti-cancer therapy. We also excluded purely methodological or review publications. We did not consider quality-adjusted life-years endpoints as PRO endpoints. We also excluded publications that reported interim analyses or analyses of subgroups of patients (ie, subgroups within the PRO cohort) since we wanted to limit our reporting to top-level PRO results of the RCTs. The figure shows the search strategy flowchart and the inclusion and exclusion criteria used. Two reviewers (MPe and LDo) received the initial list of the 323 potentially eligible articles and independently screened the articles on the basis of inclusion and exclusion criteria. 12 additional articles were found by manual ad-hoc checking of other publications. One reviewer (LDo) checked assessments by the two reviewers for any disagreements. Any disagreements were resolved through discussion. A third reviewer (CCo) was available when no consensus could be reached.

Evaluation process

Criteria to evaluate the assessment of statistical issues for PRO analysis were adapted from previous reviews[16,17] with adjustments made to enable a more in-depth assessment. The initial data extraction sheet was developed by MPe and CCo and pilot tested on three studies that had been selected for analysis. MPe and LDo independently evaluated the three studies on the basis of the initial data extraction sheet. When opions differed on how a variable should be coded for a study, the variable definition was further clarified (eg, differentiation between a broad and specific hypothesis). Final definitions were then reviewed and approved by CCo and the data sheet was refined. This refinement resulted in 23 evaluation criteria classified into five broad categories: general description of the article; reporting of research objectives; statistical analysis and clinical relevance; baseline assessment; and, assessing the amount of, and handling of missing data (appendix). Two reviewers (MPe and LDo) independently evaluated all identified studies on this predefined checklist of 23 criteria. One reviewer (LDo) checked the completed data extraction sheets for any disagreements. In case of disagreement, the

article was reassessed by both reviewers together. If no consensus could be reached, a third reviewer (CCo) served as a mediator to resolve disagreements. When we identified multiple publications of one RCT, the article with the most comprehensive PRO statistical reporting was included in the Review (appendix). Therefore, the findings we report in this systematic review are based on unique RCTs.

Findings

Table 1 summarises the main findings of this systematic review. To assess whether practices improved over time, results were grouped into three time periods—2001–06, 2007–12, and 2013–17 (table 2). Details about individual papers[18–93] can be found in the appendix.

Descriptive statistics

After duplicates were removed, we identified 335 eligible articles, of which 66 were eligible RCTs in advanced breast cancer involving 26 905 patients. No disagreements occurred between the two independent reviewers. The sample size ranged from 66 to 1102, with an average of 407. From the 66 RCTs, 12 (18%) were considered to be practice-changing trials. An RCT was considered practice changing if the trial led directly to new treatment options, drugs receiving approval from authorities, expanded indication, or new combinations or schedules of administration. Practice-changing trials were evaluated by two experts in the field of breast cancer (Mariana Brandao and Noam Pondé [Institut Jules Bordet, Brussels, Belgium]), with the supervision of MPi. The most commonly used PRO measures were two cancer-specific HRQOL questionnaires: the European Organisation for Research and Treatment of Cancer QLQ-C30 (35 [53%] of 66 RCTs) and the Functional Assessment of Cancer Therapy-Breast (22 [33%]). 27 (41%) of 66 RCTs used multiple assessment tools to measure PROs, of which six (22%) trials used an instrument that was not validated (eg, ad-hoc, trial-specific checklists) in addition to a validated questionnaire. Most of the PRO endpoints were reported as secondary endpoints (46 [70%]), with only three (5%) including a PRO as a primary endpoint. The remaining RCTs either reported PRO as an exploratory endpoint (three [5%]) or did not clearly report the PRO endpoint (14 [21%]).

Reporting of research objectives

Only eight (12%) of 66 RCTs reported a hypothesis specific enough to inform the analysis of the PRO endpoint (ie, the direction of hypothesis is stated with the domain of interest and specified time frame). Most articles either reported a broad hypothesis (25 [38%]; for example, "to evaluate HRQOL between treatment arms") or no hypothesis at all (33 [50%]). Most RCTs did not report a specific PRO hypothesis, and there was no consistent improvement in PRO hypothesis reporting over time (2001–06, 0 [0%] of 20; 2007–12, four [17%] of 24; and 2013–17, four [18%] of 22).

Statistical analysis and clinical relevance

Most trials (59 [89%]) reported analysing multivariate data with multiple PRO scales and domains, with or without repeated assessments to assess the PRO endpoint. Scales and domains refer to PRO variables that were analysed in the trial. 38 (58%) RCTs analysed multiple PRO scales and domains and 21 (32%) analysed a single PRO scale or

Section, Stockholm,Sweden (C Quinten MSc); **VU University Medical Center, Department of Neurology & Brain Tumor Center, Amsterdam, Netherlands** (J C Reijneveld MD); **Alliance Statistics and Data Center, Mayo Clinic, Rochester,**

	Yes	No	Not reported or unclear*
Reporting of research objectives			
Specific hypothesis	8 (12%)	25† (38%)	33† (50%)
Statistical significance and clinical relevance			
Multiple domains (more than one scale or domain included in analysis)	38 (58%)	21 (32%)	7 (11%)
If yes, was statistical correction used (multiple domains were independently tested)?	6 (16%) of 38	30 (79%) of 38	2 (5%) of 38
Repeated assessments (more than one follow-up assessment included in the analysis)	53 (80%)	8 (12%)	5 (8%)
If yes, was a statistical technique used that allowed inclusion of repeated assessment points, or was a statistical correction used (if repeated assessments were independently tested)?	33 (62%) of 53	12 (23%) of 53	8 (15%) of 53
Reporting of descriptive data	55 (83%)	11 (17%)	0 (0%)
Primary statistical technique			
Not reported or unclear	15 (23%)	NA	NA
(Generalised) linear mixed models, including pattern mixture models	18 (27%)	NA	NA
Wilcoxon rank-sums test or between subjects *t* test	11 (17%)	NA	NA
ANOVA or linear regression	9 (14%)	NA	NA
Time to event	6 (9%)	NA	NA
Repeated measures ANOVA	2 (3%)	NA	NA
Proportion of patients or responder analysis	2 (3%)	NA	NA
Others	3 (5%)	NA	NA
Reporting of clinical relevance	28 (42%)	38 (58%)	0 (0%)
Change of X points (from baseline)	18 (64%) of 28	NA	NA
X points difference (between arms)	9 (32%) of 28	NA	NA
Change of X points from baseline and X points differences (between arms)	1 (4%) of 28	NA	NA
Baseline assessment			
Assessed baseline	60 (91%)	6 (9%)	0 (0%)
Compared baseline scores between treatment arms	36 (60%) of 60	24 (40%) of 60	0 (0%) of 60
Included baseline as a covariate‡	13 (22%) of 60	35 (58%) of 60	12 (20%) of 60
Assessing the prevalence of, and handling of missing data			
Intention-to-treat population§	14 (21%)	28§ (42%)	24§ (36%)
Baseline compliance rates for each treatment arm¶	28 (47%) of 60	32 (53%) of 60	NA
Follow-up compliance rates for each treatment arm	19 (29%)	47 (71%)	NA
Strategy to handle missing data	18 (27%)	48 (73%)	NA

Data are n (%) and N is 66 unless otherwise stated. RCT=randomised controlled trial. *"Unclear" means that the article reported some information for the variable but not enough to give a clear yes or no response for the specific variable.†"No" means that a broad hypothesis was reported and "not reported or unclear" means no hypothesis was reported. ‡The remaining RCTs were coded as not applicable because the statistical method used did not allow for an inclusion of a covariate. §"No" means a modified intention-to-treat analysis was used and "not reported or unclear" means that the analysis population was not reported. ¶n is based on the number of studies that included a baseline assessment in their study design.

Table 1: **Key parameters for the analysis of patient-reported outcome data reported in 66 RCTs**

	2001–06 (n=20)			2007–12 (n=24)			2013–17 (n=22)		
	Yes	No	Not reported or unclear*	Yes	No	Not reported or unclear*	Yes	No	Not reported or unclear*
Reporting of research objectives									
Specific hypothesis	0 (0%)	6 (30%)†	14 (70%)†	4 (17%)	14 (58%)†	6 (25%)†	4 (18%)	5 (23%)†	13 (59%)†
Statistical significance and clinical relevance									
Multiple domains (more than one scale or domain included in analysis)	9 (45%)	8 (40%)	3 (15%)	18 (75%)	4 (17%)	2 (8%)	11 (50%)	9 (41%)	2 (9%)
If yes, was statistical correction used (multiple domains were independently tested)?	3 (33%) of 9	5 (56%) of 9	1 (11%) of 9	3 (17%) of 18	15 (83%) of 18	0 (0%) of 18	0 (0%) of 11	10 (91%) of 11	1 (9%) of 11
Repeated assessments (more than one follow-up assessment included in the analysis)	14 (70%)	3 (15%)	3 (15%)	19 (79%)	4 (17%)	1 (4%)	20 (91%)	1 (5%)	1 (5%)
If yes, was a statistical technique used that allowed inclusion of repeated assessment points or was a statistical correction used (if repeated assessments were independently tested)?	10 (71%) of 14	2 (14%) of 14	2 (14%) of 14	10 (53%) of 19	7 (37%) of 19	2 (11%) of 19	13 (65) of 20	3 (15%) of 20	4 (20%) of 20
Reporting of descriptive data	16 (80%)	4 (20%)	0 (0%)	19 (79%)	5 (21%)	0 (0%)	20 (91%)	2 (9%)	0 (0%)
Primary statistical technique									
Not reported or unclear	5 (25%)	NA	NA	6 (25%)	NA	NA	4 (18%)	NA	NA
(Generalised) linear mixed models, including pattern mixture models	8 (40%)	NA	NA	3 (13%)	NA	NA	7 (32%)	NA	NA
Wilcoxon rank-sums test or between subjects *t* test	5 (25%)	NA	NA	3 (13%)	NA	NA	3 (14%)	NA	NA
ANOVA or linear regression	1 (5%)	NA	NA	7 (29%)	NA	NA	1 (5%)	NA	NA
Time to event	1 (5%)	NA	NA	0 (0%)	NA	NA	5 (23%)	NA	NA
Repeated measures ANOVA	0 (0%)	NA	NA	2 (8%)	NA	NA	0 (0%)	NA	NA
Proportion of patients or responder analysis	0 (0%)	NA	NA	1 (4%)	NA	NA	1 (5%)	NA	NA
Others	0 (0%)	NA	NA	2 (8%)	NA	NA	1 (5%)	NA	NA
Reporting of clinical relevance	5 (25%)	15 (75%)	0 (0%)	11 (46%)	13 (54%)	0 (0%)	12 (55%)	10 (45%)	0 (0%)
Change of X points from baseline	5 (100%) of 5	NA	NA	5 (45%) of 11	NA	NA	8 (67%) of 12	NA	NA
X points difference (between arms)	0 (0%) of 5	NA	NA	6 (55%) of 11	NA	NA	3 (25%) of 12	NA	NA
Change of X points from baseline and X points differences (between arms)	0 (0%) of 5	NA	NA	0 (0%) 11	NA	NA	1 (8%) of 12	NA	NA
Baseline assessment									
Assessed baseline	18 (90%)	2 (10%)	0 (0%)	22 (92%)	2 (8%)	0 (0%)	20 (91%)	2 (9%)	0 (0%)
Compared baseline scores between treatment arms	13 (72%) of 18	5 (28%) of 18	0 (0%) of 18	14 (64%) of 22	8 (36%) of 22	0 (0%) of 22	9 (45%) of 20	11 (55%) of 20	0 (0%) of 20
Included baseline as a covariate‡	2 (11%) of 18	11 (61%) of 18	5 (28%) of 18	6 (27%) of 22	12 (55%) of 22	4 (18%) of 22	5 (25%) of 20	12 (60%) of 20	3 (15%) of 20
Assessing the prevalence of, and handling of missing data									
Intention-to-treat population§	4 (20%)	10 (50%)§	6 (30%)§	6 (25%)	10 (42%)§	8 (33%)§	4 (18%)	8 (36%)§	10 (45%)§
Baseline compliance rates for each treatment arm¶	7 (39%) of 18	11 (61%) of 18	NA	11 (50%) of 22	11 (50%) of 22	NA	10 (50%) of 20	10 (50%) of 20	NA
Follow-up compliance rates for each treatment arm	5 (25%)	15 (75%)	NA	6 (25%)	18 (75%)	NA	8 (36%)	14 (64%)	NA
Strategy to handle missing data	4 (20%)	16 (80%)	NA	9 (38%)	15 (63%)	NA	5 (23%)	17 (77%)	NA

Data are n (%). Note that some percentages do not add up to 100% because of rounding. RCT=randomised controlled trial. *"Unclear" means that the article reported some information for the variable but not enough to give a clear yes or no response for the specific variable. †"No" means that a broad hypothesis was reported and "not reported or unclear" means no hypothesis was reported. ‡RCTs that used a statistical method that does not allow for an inclusion of a covariate were coded as not applicable. §"No" means modified ITT was used and "not reported or unclear" means that the analysis population was not reported. ¶n is based on the number of studies that included a baseline assessment in their study design.

***Table 2*: Key parameters for the analysis of patient-reported outcome data in different time periods**

domain. Among the 38 RCTs, only six (16%) used a statistical correction to correct for multiple testing. Two (5%) of 38 RCTs reported PROs as an exploratory endpoint and assessed multiple outcomes—exploratory endpoints do not necessarily have to correct for multiple testing. Results remained largely the same after removing these two exploratory endpoints from the total score of PROs that assessed multiple outcomes (6 [17%] of 36). Combined, these numbers show that 27 (41%) of 66 trials addressed the issue of multiple testing either by statistically correcting for multiple scales and domains or assessing only one scale or domain (often identified a priori as the most relevant scale or domain). No clear pattern was seen in these findings over time (2001–06, 11 [55%] of 20; 2007–12, seven [29%] of 24; and 2013–17, nine [41%] of 22).

53 (80%) RCTs analysed data with repeated assessments at follow-up (defined as more than one follow-up assessment), and eight (12%) RCTs analysed data with a single follow-up assessment. Among the 53 RCTs that used multiple follow-up assessment points in their primary PRO analysis, 33 (62%) used a statistical technique that considered the repeated measurements of the data (eg, time-to-event or linear mixed models), or statistically corrected for them if these repeated measures were tested independently from one another. Overall, 41 (62%) of the 66 trials addressed the issue of multiple testing either by statistically correcting for multiple domains by use of a statistical technique that took into account the repeated measurements, or by analysing only one follow-up timepoint. These findings remain consistent over time (2001–06, 13 [65%] of 20; 2007–12, 14 [58%] of 24; and 2013–17, 14 [64%] of 22).

Most RCTs (55 [83%]) reported PRO scores descriptively, such as mean scores or mean change scores by trial arms, either on their own or as a support for a comparative analysis. This reporting has remained quite consistent over time (2001–06, 16 [80%] of 20; 2007–12, 19 [79%] of 24; and 2013–17, 20 [91%] of 22).

When analysing PRO data, we identified more than six primary statistical analysis techniques. The two most commonly used statistical techniques were (generalised) linear mixed models (18 [27%] of 66 RCTs) and Wilcoxon rank-sums test or *t* test (11 [17%] of 66). Many RCTs (15 [23%] of 66) did not report the statistical technique used; a p value was reported but it was not mentioned how this value was obtained. When comparing findings over time, the most commonly used statistical techniques between 2001–06 were (generalised) linear mixed models (eight [40%] of 20) and Wilcoxon rank-sums test or *t* test (five [25%] of 20); between 2007–12 were ANOVA or linear regression (seven [29%] of 24), (generalised) linear mixed models (three [13%] of 24) and Wilcoxon rank-sums test or *t* test (three [13%] of 24); and between 2013–17 were (generalised) linear mixed models (seven [32%] of 22) and time-to-event (five [23%] of 22). No single technique was used in most trials. Moreover, across all periods, a substantial proportion of RCTs did not report the statistical technique used (2001–06, five [25%] of 20; 2007–12, six [25%] of 24; and 2013–17, four [18%] of 22).

Less than half of the RCTs (28 [42%] of 66) addressed the clinical relevance of their findings. Among the trials that reported whether a finding was clinically relevant or not, the methods used varied; results were reported either as a change of X points from baseline (18 [64%] of 28), an X points difference between treatment arms (nine [32%] of 28), or both (one [4%] of 28). The percentage of RCTs reporting the clinical relevance of their findings increased somewhat over time (2001–06, five [25%] of 20; 2007–12, 11 [46%] of 24; and 2013–17, 12 [55%] of 22).

Baseline assessment

Most RCTs (60 [91%] of 66) included a baseline PRO assessment. From these 60 studies, 36 (60%) compared PRO baseline scores between treatment arms and 13 (22%) included the baseline score as a covariate. The inclusion of a baseline PRO assessment in most RCTs has been consistent over time (2001–06, 18 [90%] of 20; 2007–12, 22 [92%] of 24; and 2013–17, 20 [91%] of 22); however, the number of studies reporting whether PRO baseline scores are comparable between treatment arms has declined over time (2001–06, 13 [72%] of 18; 2007–12, 14 [64%] of 22; and 2013–17, nine [45%] of 20). Additionally, including baseline scores as a covariate has not necessarily improved over time (2001–06, two [11%] of 18; 2007–12, six [27%] of 22; and 2013–17, five [25%] of 20).

Amount and handling of missing data

Many RCTs (24 [36%] of 66) did not report, or did not clearly specify, the analysis population for the primary PRO analysis. This absence of reporting did not improve over time—2001–06, six (30%) of 20; 2007–12, eight (33%) of 24; and 2013–17, ten (45%) of 22. 14 (21%) RCTs used the intention-to-treat (ITT) population in their analysis, with a greater number of RCTs (28 [42%]) using a modified ITT (mITT) population. These numbers were relatively comparable over time (table 2). Five different definitions of mITT were used across the 28 trials indicating that there was no consistent definition used—18 (64%) of the 28 RCTs used baseline PRO and one or more post-assessment measure, four (14%) used baseline PRO, two (7%) used more than one PRO data-point, and two (7%) used baseline PRO and trial-specific follow-up point of interest (appendix).

Among the RCTs that assessed baseline PRO (60 [91%] of 66), only 8 (47%) reported baseline PRO compliance for each treatment arm. 19 (29%) of the 66 RCTs reported whether compliance between treatment groups differed throughout the follow-up assessments. Most studies (48 [73%]) did not report how they dealt with missing data. These findings were relatively comparable over time (table 2).

MN, USA (J Sloan PhD)**; and Leeds Institute of Cancer and Pathology, University of Leeds, St James's Hospital, Leeds, UK** (Prof G Velikova PhD)

Correspondence to:
Dr Madeline Pe, Quality of Life Department, European Organisation for Research and Treatment of Cancer, 1200 Brussels, Belgium
madeline.pe@eortc.org

See **Online** for appendix

Discussion

The aim of this systematic review was to assess the PRO analysis component in RCTs of advanced breast cancer. Our findings showed that clear heterogeneity on how PRO data were analysed among the 66 eligible RCTs.

Most trials did not report a specific research hypothesis (88%), even those from the past 6 years (2012–17, 82%). This finding is consistent with previous reviews[94–97] and could reflect a lack of knowledge about the probable HRQOL trajectory for novel treatments, or a lack of consideration of PRO-specific hypotheses at the design stage and specification in the trial protocol. This latter point is consistent with reviews[98,99] regarding the content of trial protocols. Our findings highlight an area of poor practice that does not meet International Society for Quality of Life Research and CONSORT-PRO reporting standards.[100,101] Failure to state a clear PRO hypothesis a priori opens up the possibility that inappropriate statistical techniques could be used. For instance, if a study aimed to measure HRQOL changes over a 6-week period, a cross-sectional HRQOL analysis at 6 weeks is not equivalent to an area under the curve analysis within the same timeframe; in fact, these two analytical techniques could yield different results. If the PRO objective is not clearly stated, different statistical approaches could be reported as equivalent ways of addressing the same PRO objective when in fact they focus on different aspects of the data and so respond to different research objectives. Divergent findings, however, might not necessarily invalidate the PRO data analysis but rather illustrate the importance of a well-defined a-priori hypothesis and responding to them with appropriate statistical techniques. Therefore, it is crucial that researchers clearly define their hypotheses and statistical analyses in the protocol or statistical analysis plan,[102] and that results are described to accurately represent the key patterns in the data and can be understood by non-statisticians.

The most commonly used statistical technique (linear mixed models) was only used in 18 (27%) of the 66 RCTs. Wilcoxon rank-sums test or *t* tests—statistical techniques appropriate for single time points or change scores—were also commonly used (11 [17%]), although this strategy might not always be appropriate since most trials involved analysing data with more than two repeated assessments (53 [80%]). There was an increase in the use of time-to-event analysis over time; between 2001–07, one (5%) of 20 RCTs included a time-to-event analysis compared with five (23%) of 22 between 2013–17 (table 2). However, a major concern remains that some RCTs (15 [23%] of 66) did not clearly report the statistical technique that was used to analyse PRO data, which is still the case in recent years (2013–17, four [18%] of 22).

Analysis of a PRO endpoint, such as HRQOL, often involves multiple outcomes. When drawing conclusions about treatment efficacy, type 1 errors (false positive findings) should be avoided by adjusting crucial p values for multiple comparisons when multiple outcomes are used to test a multidimensional endpoint (eg, HRQOL). Many RCTs (30 [79%] of 38) did not do this adjustment including those studies that were published in the past 6 years (2013–17; 10 [91%] of 11), which could have led to erroneous conclusions about the PRO endpoint due to excess type 1 errors.[103] Since the results of RCTs can lead to new standards of care being set, this practice should be avoided. Ongoing work from SPIRIT-PRO to standardise what needs to be included in the design stage of a trial (protocol) and statistical analysis plans could promote improved reporting on these issues.[102]

The sample size estimation required for a trial is typically calculated for the primary clinical endpoint only. Since PRO endpoints, such as HRQOL, are often secondary endpoints the sample size can be much larger (or smaller) than that needed for the primary endpoint. Since statistical significance is highly dependent on sample size, a large sample size can produce statistically significant results but the clinical relevance of the change in the PRO endpoint could be negligible.[104] Therefore, clinical relevance has been recommended to be reported alongside statistical significance. Similar to other reviews,[94–97,105] our findings showed that reporting the clinical relevance of PRO data is not common practice; less than half of the RCTs (28 [42%] of 66) reported whether their findings were clinically relevant although this practice has increased in the past 6 years (in 2001–06, five [25%] of 20 RCTs reported clinical relevance compared with 12 [55%] of 22 in 2013–17).

Most RCTs (90%) in this Review reported having a baseline assessment and this observation has been consistent over the years (2001–06, 2007–12, and 2013–17). These findings suggest that this practice has been widely accepted. Assessing baseline (or pretreatment) scores is essential in PRO analysis. Since individuals can differ in their baseline scores, it is important to take this into account when assessing differences between treatment arms and individual changes over time. This adjustment increases the efficiency of the statistical analysis by reducing the influence of baseline differences in the analysis.[106] Most RCTs (60 [91%] of 66) collected baseline PRO information but 24 (40%) of these did not check whether there were baseline differences between treatment arms. Additionally, only a small number of trials (13 [22%] of 60) reported making use of the baseline PRO scores as a covariate. These findings remain comparable over the years. Overall, these results highlight the lack of consistency between investigators on how to use baseline information in their analyses.

To assess the amount of missing data, trials should report the set or subset of participants that will be included in the analysis (the analysis population),[107] and PRO completion (or compliance) over time.[108] Only a small number of RCTs (14 [21%] of 66) used an ITT analysis, including RCTs done between 2013–17 (four [18%] of 22). Additionally, some RCTs that

purported to use an ITT analysis apparently did not adhere to the ITT principle (ie, all randomised participants should be analysed according to the allocated treatment[109]). For example, some RCTs reported that an ITT analysis would be used but removed a patient for the statistical tests if an assessment was missing (eg, when a statistical test involves calculating a change score[35,60]). Probably because of the difficulty of making use of the ITT population for PRO analysis, some RCTs used an mITT analysis instead; however, no consensus exists on which mITT approach should be used as shown by the definitions of mITT seen in these RCTs (eg, patients with baseline PRO and patients with baseline PRO plus one follow-up assessment).

Measuring compliance is another way of understanding the amount of missing data in a trial.[108] Our findings showed that although more than half of the RCTs reported baseline compliance data, a smaller number of RCTs reported follow-up compliance within the study timeframe, and not all trials compared compliance between treatment groups. This lack of information makes it difficult to evaluate whether a statistical technique is appropriate for the analysis population (eg, some statistical techniques assume that the dataset has no missing data or that data are missing at random) and whether the conclusions are generalisable to the population of interest.

Strategies to deal with missing data in statistical analyses were reported in only 18 (27%) of 66 RCTs and this practice has not changed over time (between 2001–06, four [20%] of 20 RCTs reported strategies for dealing with missing data compared with five [33%] of 22 RCTs between 2013–17). However, missing data is known to be a challenge in the analysis of PRO data in cancer trials.[8,106,110] Since patients with cancer often have disease-related and treatment-related illness and mortality, so missing assessments are often inevitable.[111] Because missing data can bias results, sensitivity analyses should be done to explore the robustness of the primary findings.[112] That is, investigators are encouraged to reanalyse the data with a statistical model that makes different missing data assumptions compared with that of the primary analysis. If results are reasonably consistent across the different analyses, there is increased confidence that the presence of missing data did not compromise the original findings.[113] The lack of information on how missing data were handled suggests that this problem is often ignored or regarded as unimportant when reporting PRO findings. This situation should not be acceptable.

Although our Review was robust and used a systematic approach, our work also has several limitations. Findings were based on published articles, and the articles selected could reflect publication bias (ie, statistically significant positive results tend to have a better chance of being published).[114] Protocols or a-priori statistical analysis plans were not checked alongside these published reports; however, information classified as not reported could have been recorded in the protocol but was not included in the article due to space limitations. However, our findings are consistent with those from other systematic reviews[98,99] of protocols and other reviews[94–97,105] evaluating the quality of PRO reporting in RCTs showing that these issues are indeed present in the PRO field. We excluded non-English publications in our search so some relevant trials could have been excluded. As the focus was on advanced breast cancer, it might not be generalisable to all cancer types, although we have no reason to think that the analysis problems that have been reported would be different in other disease sites. Indeed, the converging results from other systematic reviews[16,17,95] in different cancer sites suggest a general problem that is not specific to one cancer site. Because no standards on PRO analyses in RCTs have been defined, the evaluation criteria of these trials were based on authors' selection of statistical issues that were deemed as crucial for the analysis of PRO data, but remain broadly in line with ongoing work on guidelines for statistical analysis plans.[102] Although this Review focuses on standards in statistical analysis, we emphasise the importance of a high-quality study design and choosing appropriate PRO measures and assessment points that capture the effect of both the disease and treatment on the patient experience. Even if the most robust statistical approach is used, findings from an RCT would be of little relevance if the study design is of poor quality and inappropriate outcomes and follow-up assessment points are used.[102]

Conclusion

Our Review highlights the many statistical issues that need to be addressed to improve the analysis and interpretation of PRO data, including HRQOL. The lack of consensus on how to analyse PRO data makes it difficult to draw robust conclusions regarding PRO endpoints and compare findings across trials. Although the increased inclusion of PRO endpoints in RCTs is a substantial step towards a more patient-centred approach, standards and guidelines are needed for the analysis of PRO data in cancer RCTs. The SISAQOL consortium was set up to address this need and develop recommendations on how to analyse PRO data in RCTs,[11] of which they will produce such guidelines in the future.

Contributors

All authors conceptualised the idea during the SISAQOL consortium meeting in Brussels, Belgium in January, 2016. MPe and CCo conceptualised and developed the relevant statistical issues needed to be assessed for the analysis of PRO data. MPe did the systematic review with LDo as the second reviewer. MPe, LDo, CCo, and AB contributed to the initial interpretation of the results. MPe took the lead in drafting the manuscript. MPe and LDo drafted the initial summary of findings. LDo took the lead in the presentation of the raw results found in the appendix. AB supervised the findings and writing of this work. All authors discussed the results, provided critical feedback, and reviewed the manuscript. All authors approved the final version of the manuscript.

Declaration of interests

MC reports personal fees from Astellas and grants from NIHR, outside the submitted work, and is Chair for the International Society for Quality of Life Research, Best Practices for patient-reported outcomes (PROs) in Trials Taskforce. AC is an employee of Genentech. KC reports consultancy fees for PRO work and statistical work from Amgen, BMS, and Celgene; consultancy fees for clinical trial statistics from Endomag; and is employed part-time by Adelphi Values, outside the submitted work. ND reports grants from the EuroQol Group and the Association of the British Pharmaceutical Industry outside the submitted work. IG is an employee of Boehringer Ingelheim that provided an unrestricted education grant to the European Organisation for Research and Treatment of Cancer (EORTC). MKo reports grants from EORTC, Biofrontera, and Komitee Forschung Naturmedizin eV; and personal fees from Janssen-Cilag outside the submitted work. KO reports grants for the International Brain Tumour Alliance from AbbVie, Accuray, Antisense Pharma, Apogenix, Archimedes, Ark Therapeutics, AstraZeneca, Boehringer Ingelheim, Brain Tumour Network (USA), Brain Tumor Resource and Information Network (USA), Bristol-Myers Squibb (BMS), Celldex Therapeutics, Crusade, Dijon Designs (UK), Elekta, Eli Lilly, Gerry & Nancy Pencer Brain Trust (Canada), Gosling Foundation (UK), GlaxoSmithKline, Ivy Foundation (USA), Lully, Link Pharmaceuticals, MagForce, Medac, Merck Serono, Merck, MGI Pharma, MSD Oncology, NeoPharm, Neuroendoscopy (Australia), Northwest Biotherapeutics, Novartis, Novocure, Pediatric Brain Tumor Foundation (USA), Pfizer, Photonamic, Roche, Schering-Plough (Global), Sontag Foundation (USA), Spink (UK), to-BBB, Vane Percy (UK), VBL Therapeutics, and the Wallerstein Foundation (USA), all of which are outside the submitted work. GV reports personal fees and non-financial support from Roche and Eisai; personal fees from Novartis; grants from the National Institute for Health Research (NIHR; England), Yorkshire Cancer Research, Breast Cancer Now, and EORTC Quality of Life Group outside the submitted work. AB reports grants from Boehringer Ingelheim and the EORTC research fund during the conduct of the study, grants from Merck outside the submitted work, and is a member of the EORTC Quality of Life Group executive committee. All other authors declare no competing interests. This study received no National Institutes of Health (NIH) funding. None of the authors were fully or partly NIH funded, employed by NIH, or are in receipt of an NIH grant.

Acknowledgments

We would like to acknowledge Irina Ghislain (European Organisation for Research and Treatment of Cancer, Brussels, Belgium) for her help in the initial extraction of the relevant papers. We would also like to acknowledge Mariana Brandao and Noam Pondé (Institut Jules Bordet, Brussels, Belgium) for their help in extracting the treatment arms and pointing out which trials are practice changing. We would like to dedicate this manuscript to Setting International Standards in the Analyzing Patient-Reported Outcomes and Quality of Life Data consortium member, Franck Bonnetain, who passed away in May, 2017. Financial support for this research was provided by the European Organisation for Research and Treatment of Cancer research fund and by an unrestricted education grant from Boehringer Ingelheim.

References

1 Institute of Medicine. Crossing the quality chasm: a new health system for the 21st century. Washington (DC): National Academy Press, 2001.

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

3 European Medicines Agency. Appendix 2 to the guideline on the evaluation of anticancer medicinal products in man: the use of patient-reported outcome (PRO) measures in oncology studies. April 1, 2016. http://www.ema.europa.eu/docs/en_GB/document_library/Other/2016/04/WC500205159.pdf (accessed Nov 21, 2017).

4 FDA-NIH Biomarker Working Group. BEST (Biomarkers, EndpointS, and other Tools) Resource. 2016. https://www.ncbi.nlm.nih.gov/books/NBK338448/ (accessed March 15, 2018).

5 Bottomley A. The cancer patient and quality of life. *Oncologist* 2002; **7:** 120–25.

6 LeBlanc TW, Abernethy AP. Patient-reported outcomes in cancer care—hearing the patient voice at greater volume. *Nat Rev Clin Oncol* 2017; **14:** 763–72.

7 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.

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

9 Fayers PM, Machin D. Quality of life: assessment, analysis and interpretation of patient-reported outcomes, 2nd edn. Somerset, NJ: John Wiley & Sons, 2013.

10 US Food and Drug Administration. Patient-reported outcome measures: use in medical product development to support labeling claims. December, 2009. https://www.fda.gov/downloads/Drugs/Guidances/UCM193282.pdf (accessed Nov 21, 2017).

11 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.

12 Cardoso F, Costa A, Norton L, et al. ESO-ESMO 2nd international consensus guidelines for advanced breast cancer (ABC2). *Ann Oncol* 2014; **25:** 1871–88.

13 Higgins JPT, Green S. Cochrane handbook for systematic reviews of interventions version 5.1.0. March, 2011. http://handbook-5-1.cochrane.org/ (accessed March 1, 2016).

14 Moher D, Liberati A, Tetzlaff J, Altman DG, PRISMA Group TP. Preferred reporting items for systematic reviews and meta-analyses: the PRISMA statement. *PLoS Med* 2009; **6:** e1000097.

15 Ghislain I, Zikos E, Coens C, et al. Health-related quality of life in locally advanced and metastatic breast cancer: methodological and clinical issues in randomised controlled trials. *Lancet Oncol* 2016; **17:** e294–e304.

16 Fiteni F, Anota A, Westeel V, Bonnetain F. 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:** 122.

17 Hamel J-F, Saulnier P, Pe M, et al. A systematic review of the quality of statistical methods employed for analysing quality of life data in cancer randomised controlled trials. *Eur J Cancer* 2017; **83:** 166–76.

18 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:** 1908–15.

19 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. *Curr Med Res Opin* 2013; **29:** 1463–73.

20 Welslau M, Diéras V, Sohn JH, et al. Patient-reported outcomes from EMILIA, a randomized phase 3 study of trastuzumab emtansine (T-DM1) versus capecitabine and lapatinib in human epidermal growth factor receptor 2-positive locally advanced or metastatic breast cancer. *Cancer* 2014; **120:** 642–51.

21 Hurvitz SA, Dirix L, Kocsis J, et al. Phase II randomized study of trastuzumab emtansine versus trastuzumab plus docetaxel in patients with human epidermal growth factor receptor 2-positive metastatic breast cancer. *J Clin Oncol* 2013; **31:** 1157–63.

22 Svensson H, Einbeigi Z, Johansson H, Hatschek T, Brandberg Y. Quality of life in women with metastatic breast cancer during 9 months after randomization in the TEX trial (epirubicin and paclitaxel w/o capecitabine). *Breast Cancer Res Treat* 2010; **123:** 785–93.

23 Brufsky A, Hoelzer K, Beck T, et al. A randomized phase II study of paclitaxel and bevacizumab with and without gemcitabine as first-line treatment for metastatic breast cancer. *Clin Breast Cancer* 2011; **11:** 211–20.

24 Thomas ES, Gomez HL, Li RK, et al. Ixabepilone plus capecitabine for metastatic breast cancer progressing after anthracycline and taxane treatment. *J Clin Oncol* 2007; **25:** 5210–17.

25 Corey-Lisle PK, Peck R, Mukhopadhyay P, et al. Q-TWiST analysis of ixabepilone in combination with capecitabine on quality of life in patients with metastatic breast cancer. *Cancer* 2012; **118:** 461–68.

26 Nuzzo F, Morabito A, Gravina A, et al. Effects on quality of life of weekly docetaxel-based chemotherapy in patients with locally advanced or metastatic breast cancer: results of a single-centre randomized phase 3 trial. *BMC Cancer* 2011; **11:** 75.

27 Di Leo A, Jerusalem G, Petruzelka L, et al. Results of the CONFIRM phase III trial comparing fulvestrant 250 mg with fulvestrant 500 mg in postmenopausal women with estrogen receptor-positive advanced breast cancer. *J Clin Oncol* 2010; **28:** 4594–600.

28 Ellis MJ, Gao F, Dehdashti F, et al. Lower-dose (6 mg daily) versus high-dose (30 mg daily) oral estradiol therapy of hormone-receptor-positive, aromatase-inhibitor-resistant advanced breast cancer: a randomized phase 2 study. *JAMA* 2009; **302:** 774–80.

29 Zhou X, Cella D, Cameron D, et al. Lapatinib plus capecitabine versus capecitabine alone for HER2+ (ErbB2+) metastatic breast cancer: Quality-of-life assessment. *Breast Cancer Res Treat* 2009; **117:** 577–89.

30 Hopwood P, Watkins J, Ellis P, Smith I. Clinical interpretation of quality-of-life outcomes: an investigation of data from the randomized trial of gemcitabine plus paclitaxel compared with paclitaxel alone for advanced breast cancer. *Breast J* 2008; **14:** 228–35.

31 Moinpour CM, Donaldson GW, Liepa AM, Melemed AS, O'Shaughnessy J, Albain KS. Evaluating health-related quality-of-life therapeutic effectiveness in a clinical trial with extensive nonignorable missing data and heterogeneous response: results from a phase III randomized trial of gemcitabine plus paclitaxel versus paclitaxel monothe. *Qual Life Res* 2012; **21:** 765–75.

32 Chia S, Gradishar W, Mauriac L, et al. Double-blind, randomized placebo controlled trial of fulvestrant compared with exemestane after prior nonsteroidal aromatase inhibitor therapy in postmenopausal women with hormone receptor-positive, advanced breast cancer: results from EFECT. *J Clin Oncol* 2008; **26:** 1664–70.

33 Reyno L. Phase III study of *N,N*-diethyl-2-[4-(phenylmethyl) phenoxy]ethanamine (BMS-217380-01) combined with doxorubicin versus doxorubicin alone in metastatic/recurrent breast cancer: National Cancer Institute of Canada Clinical Trials Group study. *J Clin Oncol* 2004; **22:** 269–76.

34 Liu J, Tu D, Dancey J, et al. Quality of life analyses in a clinical trial of DPPE (tesmilifene) plus doxorubicin versus doxorubicin in patients with advanced or metastatic breast cancer: NCIC CTG Trial MA.19. *Breast Cancer Res Treat* 2006; **100:** 263–71.

35 Jones SE, Erban J, Overmoyer B, et al. Randomized phase III study of docetaxel compared with paclitaxel in metastatic breast cancer. *J Clin Oncol* 2005; **23:** 5542–51.

36 Therasse P, Mauriac L, Welnicka-Jaskiewicz M, et al. Final results of a randomized phase III trial comparing cyclophosphamide, epirubicin, and fluorouracil with a dose-intensified epirubicin and cyclophosphamide + filgrastim as neoadjuvant treatment in locally advanced breast cancer: an EORTC-NCIC-SAKK Mult. *J Clin Oncol* 2003; **21:** 843–50.

37 Bottomley A, Therasse P, Piccart M, et al. Health-related quality of life in survivors of locally advanced breast cancer: an international randomised controlled phase III trial. *Lancet Oncol* 2005; **6:** 287–94.

38 Fountzilas G, Kalofonos HP, Dafni U, et al. Paclitaxel and epirubicin versus paclitaxel and carboplatin as first-line chemotherapy in patients with advanced breast cancer: a phase III study conducted by the Hellenic Cooperative Oncology Group. *Ann Oncol* 2004; **15:** 1517–26.

39 Howell A, Robertson JFR, Albano JQ, et al. Fulvestrant, formerly ICI 182,780, is as effective as anastrozole in postmenopausal women with advanced breast cancer progressing after prior endocrine treatment. *J Clin Oncol* 2002; **20:** 3396–403.

40 Osborne CK, Pippen J, Jones SE, et al. Double-blind, randomized trial comparing the efficacy and tolerability of fulvestrant versus anastrozole in postmenopausal women with advanced breast cancer progressing on prior endocrine therapy: results of a North American trial. *J Clin Oncol* 2002; **20:** 3386–95.

41 Buzdar A, Douma J, Davidson N, et al. Phase III, multicenter, double-blind, randomized study of letrozole, an aromatase inhibitor, for advanced breast cancer versus megestrol acetate. *J Clin Oncol* 2001; **19:** 3357–66.

42 Conte PF, Guarneri V, Bruzzi P, et al. Concomitant versus sequential administration of epirubicin and paclitaxel as first-line therapy in metastatic breast carcinoma: results from the gruppo oncologico nord ovest randomized trial. *Cancer* 2004; **101:** 704–12.

43 Keller AM, Mennel RG, Georgoulias VA, et al. Randomized phase III trial of pegylated liposomal doxorubicin versus vinorelbine or mitomycin C plus vinblastine in women with taxane-refractory advanced breast cancer. *J Clin Oncol* 2004; **22:** 3893–901.

44 O'Shaughnessy J, Miles D, Vukelja S, et al. Superior survival with capecitabine plus docetaxel combination therapy in anthracycline-pretreated patients with advanced breast cancer: phase III trial results. *J Clin Oncol* 2002; **20:** 2812–23.

45 Cortés J, Baselga J, Im Y-H, et al. Health-related quality-of-life assessment in CLEOPATRA, a phase III study combining pertuzumab with trastuzumab and docetaxel in metastatic breast cancer. *Ann Oncol* 2013; **24:** 2630–35.

46 Lück H-J, Du Bois A, Loibl S, et al. Capecitabine plus paclitaxel versus epirubicin plus paclitaxel as first-line treatment for metastatic breast cancer: efficacy and safety results of a randomized, phase III trial by the AGO Breast Cancer Study Group. *Breast Cancer Res Treat* 2013; **139:** 779–87.

47 Gianni L, Romieu GH, Lichinitser M, et al. AVEREL: a randomized phase III trial evaluating bevacizumab in combination with docetaxel and trastuzumab as first-line therapy for her2-positive locally recurrent/metastatic breast cancer. *J Clin Oncol* 2013; **31:** 1719–25.

48 Bachelot T, Bajard A, Ray-Coquard I, et al. Final results of ERASME-4: a randomized trial of first-line docetaxel plus either capecitabine or epirubicin for metastatic breast cancer. *Oncology* 2011; **80:** 262–68.

49 Blackwell KL, Burstein HJ, Storniolo AM, et al. Randomized study of lapatinib alone or in combination with trastuzumab in women with ErbB2-positive, trastuzumab-refractory metastatic breast cancer. *J Clin Oncol* 2010; **28:** 1124–30.

50 Wu Y, Amonkar MM, Sherrill BH, et al. Impact of lapatinib plus trastuzumab versus single-agent lapatinib on quality of life of patients with trastuzumab-refractory HER2+ metastatic breast cancer. *Ann Oncol* 2011; **22:** 2582–90.

51 Schröder CP, De Munck L, Westermann AM, et al. Weekly docetaxel in metastatic breast cancer patients: no superior benefits compared to three-weekly docetaxel. *Eur J Cancer* 2011; **47:** 1355–62.

52 Sherrill B, Amonkar MM, Sherif B, Maltzman J, O'Rourke L, Johnston S. Quality of life in hormone receptor-positive HER-2+ metastatic breast cancer patients during treatment with letrozole alone or in combination with lapatinib. *Oncologist* 2010; **15:** 944–53.

53 Sherrill B, Di Leo A, Amonkar MM, et al. Quality-of-life and quality-adjusted survival (Q-TWiST) in patients receiving lapatinib in combination with paclitaxel as first-line treatment for metastatic breast cancer. *Curr Med Res Opin* 2010; **26:** 767–75.

54 Meier CR, Illiger HJ, Steder M, et al. Weekly vinorelbine versus docetaxel for metastatic breast cancer after failing anthracycline treatment. *Onkologie* 2008; **31:** 447–53.

55 Fountzilas G, Dafni U, Dimopoulos MA, et al. A randomized phase III study comparing three anthracycline-free taxane-based regimens, as first line chemotherapy, in metastatic breast cancer: a Hellenic Cooperative Oncology Group study. *Breast Cancer Res Treat* 2009; **115:** 87–99.

56 Miller K, Wang M, Gralow J, et al. Paclitaxel plus bevacizumab versus paclitaxel alone for metastatic breast cancer. *N Engl J Med* 2007; **357:** 2666–76.

57 Cella D, Wang M, Wagner L, Miller K. Survival-adjusted health-related quality of life (HRQL) among patients with metastatic breast cancer receiving paclitaxel plus bevacizumab versus paclitaxel alone: results from Eastern Cooperative Oncology Group Study 2100 (E2100). *Breast Cancer Res Treat* 2011; **130:** 855–61.

58 Crump M, Gluck S, Tu D, et al. Randomized trial of high-dose chemotherapy with autologous peripheral-blood stem-cell support compared with standard-dose chemotherapy in women with metastatic breast cancer: NCIC MA.16. *J Clin Oncol* 2008; **26:** 37–43.

59 Karamouzis MV, Ioannidis G, Rigatos G. Quality of life in metastatic breast cancer patients under chemotherapy or supportive care: a single-institution comparative study. *Eur J Cancer Care (Engl)* 2007; **16:** 433–38.

60 Cassier PA, Chabaud S, Trillet-Lenoir V, et al. A phase-III trial of doxorubicin and docetaxel versus doxorubicin and paclitaxel in metastatic breast cancer: results of the ERASME 3 study. *Breast Cancer Res Treat* 2007; **109:** 343–50.

61 von Minckwitz G, Chernozemsky I, Sirakova L, et al. Bendamustine prolongs progression-free survival in metastatic breast cancer (MBC): a phase III prospective, randomized, multicenter trial of bendamustine hydrochloride, methotrexate and 5-fluorouracil (BMF) versus cyclophosphamide, methotrexate and 5-fluo. *Anticancer Drugs* 2005; **16:** 871–77.

62 Miller KD, Chap LI, Holmes FA, et al. Randomized phase III trial of capecitabine compared with bevacizumab plus capecitabine in patients with previously treated metastatic breast cancer. *J Clin Oncol* 2005; **23:** 792–99.

63 Bottomley A, Biganzoli L, Cufer T, et al. Randomized, controlled trial investigating short-term health-related quality of life with doxorubicin and paclitaxel versus doxorubicin and cyclophosphamide as first-line chemotherapy in patients with metastatic breast cancer: European Organization for Research and Treatment of Cancer Breast Cancer Group, Investigational Drug Branch for Breast Cancer and the New Drug Development Group Study. *J Clin Oncol* 2004; **22:** 2576–86.

64 Winer EP, Berry DA, Woolf S, et al. Failure of higher-dose paclitaxel to improve outcome in patients with metastatic breast cancer: cancer and leukemia group B trial 9342. *J Clin Oncol* 2004; **22:** 2061–68.

65 Nabholtz JM, Falkson C, Campos D, et al. Docetaxel and doxorubicin compared with doxorubicin and cyclophosphamide as first-line chemotherapy for metastatic breast cancer: results of a randomized, multicenter, phase III trial. *J Clin Oncol* 2003; **21:** 968–75.

66 Sledge GW, Neuberg D, Bernardo P, et al. Phase III trial of doxorubicin, paclitaxel, and the combination of doxorubicin and paclitaxel as front-line chemotherapy for metastatic breast cancer: an intergroup trial (E1193). *J Clin Oncol* 2003; **21:** 588–92.

67 Osoba D, Slamon DJ, Burchmore M, Murphy M. Effects on quality of life of combined trastuzumab and chemotherapy in women with metastatic breast cancer. *J Clin Oncol* 2002; **20:** 3106–13.

68 Vogel CL, Cobleigh MA, Tripathy D, et al. Efficacy and safety of trastuzumab as a single agent in first-line treatment of *HER2*-overexpressing metastatic breast cancer. *J Clin Oncol* 2002; **20:** 719–26.

69 Rugo H, Brammer M, Zhang F, Lalla D. Effect of trastuzumab on health-related quality of life in patients with HER2-positive metastatic breast cancer: data from three clinical trials. *Clin Breast Cancer* 2010; **10:** 288–93.

70 Eiermann W. Trastuzumab combined with chemotherapy for the treatment of HER2-positive metastatic breast cancer: pivotal trial data. *Ann Oncol* 2001; **12:** 57–62.

71 Chan S, Romieu G, Huober J, et al. Phase III study of gemcitabine plus docetaxel compared with capecitabine plus docetaxel for anthracycline-pretreated patients with metastatic breast cancer. *J Clin Oncol* 2009; **27:** 1753–60.

72 Del Mastro L, Fabi A, Mansutti M, et al. Randomised phase 3 open-label trial of first-line treatment with gemcitabine in association with docetaxel or paclitaxel in women with metastatic breast cancer: a comparison of different schedules and treatments. *BMC Cancer* 2013; **13:** 164.

73 Park YH, Jung KH, Im SA, et al. Phase III, multicenter, randomized trial of maintenance chemotherapy versus observation in patients with metastatic breast cancer after achieving disease control with six cycles of gemcitabine plus paclitaxel as first-line chemotherapy: KCSG-BR07-02. *J Clin Oncol* 2013; **31:** 1732–40.

74 Park YH, Jung KH, Im SA, et al. Quality of life (QoL) in metastatic breast cancer patients with maintenance paclitaxel plus gemcitabine (PG) chemotherapy: results from phase III, multicenter, randomized trial of maintenance chemotherapy versus observation (KCSG-BR07-02). *Breast Cancer Res Treat* 2015; **152:** 77–85.

75 Gelmon KA, Boyle FM, Kaufman B, et al. Lapatinib or trastuzumab plus taxane therapy for human epidermal growth factor receptor 2-positive advanced breast cancer: final results of NCIC CTG MA.31. *J Clin Oncol* 2015; **33:** 1574–83.

76 Kaufman PA, Awada A, Twelves C, et al. Phase III open-label randomized study of eribulin mesylate versus capecitabine in patients with locally advanced or metastatic breast cancer previously treated with an anthracycline and a taxane. *J Clin Oncol* 2015; **33:** 594–601.

77 Cortes J, Hudgens S, Twelves C, et al. Health-related quality of life in patients with locally advanced or metastatic breast cancer treated with eribulin mesylate or capecitabine in an open-label randomized phase 3 trial. *Breast Cancer Res Treat* 2015; **154:** 509–20.

78 Park IH, Ro J, Lee KS, Kim SN, Yun YH, Nam BH. Phase II study of gemcitabine in combination with vinorelbine versus gemcitabine followed by vinorelbine for metastatic breast cancer. *Invest New Drugs* 2010; **28:** 659–69.

79 Pivot X, Spano JP, Espie M, et al. Patients' preference of trastuzumab administration (subcutaneous versus intravenous) in HER2-positive metastatic breast cancer: results of the randomised MetaspHer study. *Eur J Cancer* 2017; **82:** 230–36.

80 Perez EA, Barrios C, Eiermann W, et al. Trastuzumab emtansine with or without pertuzumab versus trastuzumab plus taxane for human epidermal growth factor receptor 2-positive, advanced breast cancer: primary results from the phase III MARIANNE study. *J Clin Oncol* 2017; **35:** 141–48.

81 Shiroiwa T, Fukuda T, Shimozuma K, et al. Long-term health status as measured by EQ-5D among patients with metastatic breast cancer: comparison of first-line oral S-1 and taxane therapies in the randomized phase III SELECT BC trial. *Qual Life Res* 2017; **26:** 445–53.

82 Cinieri S, Chan A, Altundag K, et al. Final results of the randomized phase II NorCap-CA223 trial comparing first-line all-oral versus taxane-based chemotherapy for HER2-negative metastatic breast cancer. *Clin Breast Cancer* 2017; **17:** 91–99.

83 Rochlitz C, Bigler M, von Moos R, et al. SAKK 24/09: safety and tolerability of bevacizumab plus paclitaxel vs. bevacizumab plus metronomic cyclophosphamide and capecitabine as first-line therapy in patients with HER2-negative advanced stage breast cancer—a multicenter, randomized phase III trial. *BMC Cancer* 2016; **16:** 780.

84 Gradishar WJ, Tjulandin S, Davidson N, et al. Phase III trial of nanoparticle albumin-bound paclitaxel compared with polyethylated castor oil-based paclitaxel in women with breast cancer. *J Clin Oncol* 2005; **23:** 7794–803.

85 Howell A, Robertson JFR, Abram P, et al. Comparison of fulvestrant versus tamoxifen for the treatment of advanced breast cancer in postmenopausal women previously untreated with endocrine therapy: a multinational, double-blind, randomized trial. *J Clin Oncol* 2004; **22:** 1605–13.

86 Takashima T, Mukai H, Hara F, et al. Taxanes versus S-1 as the first-line chemotherapy for metastatic breast cancer (SELECT BC): an open-label, non-inferiority, randomised phase 3 trial. *Lancet Oncol* 2016; **17:** 90–98.

87 Yamamoto D, Sato N, Rai Y, et al. Efficacy and safety of low-dose capecitabine plus docetaxel versus single-agent docetaxel in patients with anthracycline-pretreated HER2-negative metastatic breast cancer: results from the randomized phase III JO21095 trial. *Breast Cancer Res Treat* 2017; **161:** 473–82.

88 Pagani O, Klingbiel D, Ruhstaller T, et al. Do all patients with advanced HER2 positive breast cancer need upfront-chemo when receiving trastuzumab? Randomized phase III trial SAKK 22/99. *Ann Oncol* 2017; **28:** 305–12.

89 Harbeck N, Saupe S, Jäger E, et al. A randomized phase III study evaluating pegylated liposomal doxorubicin versus capecitabine as first-line therapy for metastatic breast cancer: results of the PELICAN study. *Breast Cancer Res Treat* 2017; **161:** 63–72.

90 Vrdoljak E, Marschner N, Zielinski C, et al. Final results of the TANIA randomised phase III trial of bevacizumab after progression on first-line bevacizumab therapy for HER2-negative locally recurrent/metastatic breast cancer. *Ann Oncol* 2016; **27:** 2046–52.

91 Turner NC, Ro J, André F, et al. Palbociclib in hormone-receptor–positive advanced breast cancer. *N Engl J Med* 2015; **373:** 209–19.

92 Harbeck N, Iyer S, Turner N, et al. Quality of life with palbociclib plus fulvestrant in previously treated hormone receptor-positive, HER2-negative metastatic breast cancer: patient-reported outcomes from the PALOMA-3 trial. *Ann Oncol* 2016; **27:** 1047–54.

93 Bell T, Crown JP, Lang I, et al. Impact of palbociclib plus letrozole on pain severity and pain interference with daily activities in patients with estrogen receptor-positive/human epidermal growth factor receptor 2-negative advanced breast cancer as first-line treatment. *Curr Med Res Opin* 2016; **32:** 959–65.

94 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:** 653–64.

95 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.

96 Schandelmaier S, Conen K, von Elm E, et al. Planning and reporting of quality-of-life outcomes in cancer trials. *Ann Oncol* 2015; **26:** 1966–73.

97 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:** 3335–42.

98 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.

99 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.

100 Calvert M, Blazeby J, Altman DG, et al. Reporting of patient-reported outcomes in randomized trials. *JAMA* 2013; **309:** 814.

101 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:** 1161–75.

102 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.

103 The European Agency for the Evaluation of Medicinal Products (EMEA). Points to consider on multiplicity issues in clinical trials. Sept 19, 2002. http://www.ema.europa.eu/docs/en_GB/document_library/Scientific_guideline/2009/09/WC500003640.pdf (accessed Nov 21, 2017).

104 Ranganathan P, Pramesh CS, Buyse M. Common pitfalls in statistical analysis: clinical versus statistical significance. *Perspect Clin Res* 2015; **6:** 169–70.

105 Cocks K, King MT, Velikova G, Fayers PM, Brown JM. Quality, interpretation and presentation of European Organisation for Research and Treatment of Cancer quality of life questionnaire core 30 data in randomised controlled trials. *Eur J Cancer* 2008; **44:** 1793–98.

106 Fairclough DL. Design and analysis of quality of life studies in clinical trials. Florida: Chapman & Hall, 2002.

107 Chan A-W, Tetzlaff JM, Gotzsche PC, et al. SPIRIT 2013 explanation and elaboration: guidance for protocols of clinical trials. *BMJ* 2013; **346:** e7586.

108 Mercieca-Bebber R, Palmer MJ, Brundage M, Calvert M, Stockler MR, King MT. Design, implementation and reporting strategies to reduce the instance and impact of missing patient-reported outcome (PRO) data: a systematic review. *BMJ Open* 2016; **6:** e010938.

109 Montedori A, Bonacini MI, Casazza G, et al. Modified versus standard intention-to-treat reporting: are there differences in methodological quality, sponsorship, and findings in randomized trials? A cross-sectional study. *Trials* 2011; **12:** 58.

110 Rombach I, Rivero-Arias O, Gray AM, Jenkinson C, Burke Ó. The current practice of handling and reporting missing outcome data in eight widely used PROMs in RCT publications: a review of the current literature. *Qual Life Res* 2016; **25:** 1613–23.

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

112 Fielding S, Ogbuagu A, Sivasubramaniam S, MacLennan G, Ramsay CR. 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.

113 White IR, Horton NJ, Carpenter J, Pocock SJ, Pocock SJ. Strategy for intention to treat analysis in randomised trials with missing outcome data. *BMJ* 2011; **342:** d40.

114 Dubben H-H, Beck-Bornholdt H-P. Systematic review of publication bias in studies on publication bias. *BMJ* 2005; **331:** 433–34.