Breast Cancer Research and Treatment **87:** 233–243, 2004.

Report

Quality of life in randomized trials of cytotoxic or hormonal treatment of advanced breast cancer. Is there added value?

Roldano Fossati, Carlo Confalonieri, Paola Mosconi, Vanna Pistotti, and Giovanni Apolone
Istituto di Ricerche Farmacologiche "Mario Negri", Milan, Italy

Key words: advanced breast cancer, cytotoxic-hormonal treatment, quality of life

Summary

Background. Since most advanced cancers are still incurable, oncologic clinical research pays considerable attention to palliation, increasingly valuing subjective measures of outcome such as quality of life (QoL). We reviewed randomised clinical trials (RCT) of cytotoxic or hormonal treatments in advanced breast cancer (ABC), published before December 2003, to evaluate the methodological quality of QoL assessment and assess its added value (over classical clinical endpoints (CCE), i.e. survival, response, time to progression, toxicity) in the choice of the best treatment option.

Methods. RCTs were classified according to treatment characteristics and the CCEs. A descriptive analysis was based on the methodological aspects of QoL assessment and the clinical value of QoL findings was judged by counting the frequency of reporting in the study abstracts and the assessment of QoL combined with CCEs.

Results. We retrieved 33 eligible RCTs (10,791 patients); only 20 reported the number of patients considered in QoL principal analysis and only 69% of randomized patients were included in such analyses. A total of 17 different QoL questionnaires were used, 11 only once. QoL assessment lasted from less than 12 weeks to progression, and timing of questionnaires from 2 to 12 weeks. Compliance rates were 85.7% for baseline forms and 67% for overall assessment, but this information was available for only 18 and 20 trials, respectively. Wide variability emerged in analysis strategies and statistical approaches. QoL findings were reported in 12 study abstracts (37% of patients). Eight studies reported a significant difference in QoL scores but since QoL data often failed to parallel the clinical findings (e.g. better QoL scores were reported in two of 17 trials with better CCEs and in six of 20 with significant differences in toxicity profiles), the QoL added value was difficult to ascertain and, on the whole, only moderate.

Conclusion. In ABC trials, QoL assessment added relatively little value to CCEs in helping select the best treatment option, apparently largely because of sub-optimal methodological standards.

Introduction

Most advanced common cancers are still incurable and new therapies are often selected by comparing their chances of postponing death. If improved survival, not cure, is the goal, clinicians have still to prove that the toxic effects of therapies do not offset any gain in survival among patients with a limited life expectancy. Furthermore, when the unfortunate reality is that expected survival gains are small if any, better palliation becomes the actual goal. Well codified clinical endpoints such as tumor response, time to progression or toxicity assessments have been used as surrogate endpoints for palliation but are ultimately deemed inadequate to understand the patient's perception of the impact of therapy [1]. This has encouraged clinical researchers to value other subjective endpoints of clinical benefit such as the so-called health related quality of life [2].

Since the 80s there has been a steep rise in the use of quality of life (QoL) evaluation in clinical research in cancer. Advanced breast cancer (ABC) was soon recognised as an ideal clinical context for QoL assessment on account of its high prevalence, the relatively long survival after metastasis, and the vast array of cytotoxic or hormonal treatments with widely varying side effects which could have an impact on patients' subjective health status. Two systematic reviews have evaluated QoL assessment in randomized trials for breast cancer [3, 4]. These papers gave very good analytical descriptions of each study but the question remains as to whether and how often QoL assessment provides clinicians with reliable information not already conveyed by classical clinical endpoints (CCE).

We evaluated randomised clinical trials (RCT) in ABC that used formal QoL outcome measures (i.e. standardised self-reported measures of functioning and well-being) in order to describe how QoL was assessed and check whether a fuller description of the effects of treatment would enable physicians to select a treatment regimen better.

Methods

Identification and retrieval of randomized trials

The procedure for Pubmed search (1966 and, to December 2003) listed in Table 1 was used to identify RCTs. For further verification, a supplementary EMBASE search was conducted using the same adapted search strategy.

Abstracts were discarded if a full text of the article was not retrieved and similarly we excluded studies that enrolled patients with tumors other than breast, and studies not evaluating cytotoxic or hormonal treatments.

Since two people (RF and CC) independently reviewed all potentially relevant material to establish whether it met the inclusion criteria, any disagreement in the rejection process was first handled between them. When this could not be done, a third independent reviewer's opinion was sought (GA).

Type of data and evaluation of QoL reporting

The following information was gathered from each report: randomisation starting year, line of palliative treatment, description of treatments and type of comparison, QoL as primary or secondary endpoint, number of patients randomised, mean/median age, number of patients offered participation in the QoL assessment, number of patients included in the QoL principal analysis, type and number of questionnaires used, duration of QoL assessment, timing of QoL investigations, summaries of compliance with QoL forms (at baseline and overall), the authors' use of QoL findings, independently from their statistical significance, in the choice of treatment as reported in the study

Table 1. Pubmed search used to identify RCTs

Database	Query No	Search terms	Results
	1	"Breast Neoplasms" [MeSH:NoExp]	110016
	2	Breast AND (neoplasm* OR tumor* OR tumour* OR cancer*)	179821
	3	1 OR 2	179821
	4	3 AND (random OR controlled trial)	9471
	5	3 AND "controlled clinical trial" [Publication Type]	946
	6	3 AND "randomized controlled trial" [Publication Type]	4472
Pubmed	7	4 OR 5 OR 6	11172
	8	3 AND 7	11172
	9	"Quality of Life" [MeSH]	38658
	10	"Quality of Life"	58210
	11	8 AND (9 OR10)	575
	12	11 AND (advanc* OR metasta*)	256
	13	12 Limits: Publication Date from 1966 to 2003	252

Note: English language was selected.

abstract, statistical significance of traditional clinical outcomes (response rate, progression-free/treatment failure survival, overall survival, toxicity) and QoL scores.

We classified QoL results as 'not significant' when all measured QoL dimensions or when the primary QoL endpoints, according to the authors' definition, showed no statistically significant difference (again, according to the authors' definition, but with p value at least <0.05). Lacking a validated tool to assess the quality of measuring and reporting QoL (methodological quality) we used the checklist proposed by Chassany et al [5] to improve standards of QoL studies and we assigned a numerical score to a selection of those items. We selected nine items as most important and presented them as questions to elicit "yes" or "no" answers, namely: (1) Was the QoL-related research hypothesis clearly explained? (2) Was there evidence of psychometric validation of the questionnaire? (3) Was the timing and frequency of QoL assessment reported? (4) Was the mode and the site of administration of questionnaires reported? (5) Were procedures for handling of missing data reported? (6) Was it specified whether QoL analysis followed an "intention to treat" approach? (7) Were inferential statistics specified for evaluating group differences? (8) Were statistical procedures adopted to correct for multiple tests and preserve the overall type I error rate? (9) Was patients' participation rate at study entry and/or afterwards reported? Each "yes" was given a score of one point. A further point was added if the overall methodological quality of the study was satisfactory (i.e. more than three points according to Jadad's validated quality scale [6]). Overall, this quality rating could produce a score from 0 (lowest quality) to 10 (highest quality).

Analysis

The analysis of these trials began by describing the clinical characteristics of treatment comparisons, looking for any significant changes in the traditional clinical endpoints mentioned. Changes were considered significant when the p value was less than 0.05. When assessing QoL aspects of the trials we first reported methodological details of the evaluation (type and number of questionnaires used, duration of assessment, etc.), giving averages weighted by the study sample size as simple summary measures of baseline and overall compliance. The summary score for methodological quality of QoL assessment was used to evaluate any trends in quality over time (specifically, the year randomization began) and to test differences between predefined categories such as studies reporting or not reporting QoL in their abstracts, studies considering QoL as primary or secondary endpoint, and studies yielding statistically significant differences in QoL results or not. Associations between classificative variables were explored by Spearman's correlation coefficient and with the Mann–Whitney U test.

Then we evaluated the added value of QoL data that could help decide about treatment options. First, we looked at abstracts to check whether QoL findings were reported. Second, we made a cross-evaluation of the significant differences in QoL and CCE of efficacy and toxicity. Operationally, we created five mutually exclusive groups (listed in Table 5) from the trials based on whether the treatments compared were more, less or equally (i.e. nonsignificant) efficacious or toxic. Then we analysed the distribution of studies with significant differences in QoL scores inside these groups. From their relationships we could summarize the interplay between CCE and QoL data.

Results

Data search

The 252 records identified through the search strategy shown in Table 1 shrank to 35 useful papers (which reported 30 eligible trials) when 217 articles were excluded because: 85 were reviews or non-randomised trials, 35 reported treatments other than cytotoxic or hormonal, 21 were published in non-English language journals, 19 reported on mixed forms of tumors (including breast cancer) or other tumors, 16 were methodological papers reporting the evaluation of QoL measures, 16 reported adjuvant therapies; 15 were preliminary analyses or duplicate publications, eight had data inadequate for analysis, and two did not actually report on QoL. Three further trials were retrieved by a hand search, totalling 33 randomised trials whose results were reported in 38 different articles that included QoL as an explicit endpoint [7–44].

Descriptive analyses of trials and QoL assessment method

These 33 trials enrolled 10,791 patients, mean age 58.6 years (62.7 and 53.9 respectively for patients receiving hormonal treatments or chemotherapy). Four trials (1020 patients) started randomisation between 1981 and 1986 [7, 14, 15, 37, 39], 13 (3752 patients) between 1987 and 1992 [8, 9, 11, 15, 17, 18, 20–22, 24, 25, 29, 33, 36, 40], 16 (6019 patients) between 1993 and 1998 [10, 12, 13, 16, 18–20, 22–24, 26–28, 30–35, 38, 41–44]. Table 2 lists some details of these trials and the clinical results.

QoL assessment was considered a primary endpoint in seven trials (1731 patients) [9, 14, 15, 17, 25, 35, 36, 41] while the other 26 specified that QoL was a secondary endpoint according to the authors or judged as such in the case of sample size calculations on different outcome measures. In 29 trials (9629 patients) [8, 10, 11, 13–22, 24, 26–44], the authors reported the number of patients asked to participate in QoL assessment (9232 patients or 96% of those randomised) but only 20 trials (5772 patients) [8, 13–15, 17–22, 24, 28–36, 38–42] clearly reported how many patients were considered in the QoL principal analysis (4005 patients or 69% of those randomised).

Methods adopted to measure QoL are shown in Table 3 (in brackets the number of studies and numbers of patients for whom the information was available). A total of 17 different instruments were used in the 33 articles reviewed. Of these, 11 were used only once. Compliance rates in the collection of self-reported QoL data are reported in Table 4.

Several strategies were followed to study differences in QoL between treatments. Trialists compared QoL data at each individual time point, or compared differences between the post baseline mean scores and baseline, between last measurement and baseline, between a single time point and baseline, between the best/worst/last response and baseline; they compared the proportion of patients whose baseline score improved by two points, or the time to improvement of the baseline score by 10 points. Sometimes the strategies were not clearly reported. Statistical analysis of these summary measures showed a similar wide array of parametric, non-parametric, univariate and multivariate approaches that relied on methods such as the Wilcoxon rank sum test, Mann–Whitney U test, Wei–Johnson test, Student's *t* test, ANOVA, ANCOVA, means with confidence intervals, least-squares means, logistic regression, repeated measures ANOVA and MANOVA, random coefficients model, and the Kaplan–Meier method. All these were used to analyse either a single global index of QoL or comprehensive sets of separate scores of specific aspects of QoL.

QoL results

In 12 trials (4041 patients, 37%) [12, 14, 15, 18, 22, 24, 27, 29, 35–37, 39–41] the authors used QoL findings in the abstract as a part of their decision-

Table 2. Details of RCTs and clinical findings

		No. trials	No. pts
Type of treatment	Hormonal therapy [8, 12, 16, 18, 19, 21, 22, 26–28, 29, 38, 40, 43, 44]	13	5213
	Chemotherapy [7, 9–11, 13–15, 17, 20, 23–25, 30–37, 39, 41, 42]	20	5578
Treatment line	First line [7, 10, 11, 14, 15, 17, 20–24, 29–31, 34–37, 39–42]	19	5654
	Second line [8, 9, 12, 13, 16, 18, 19, 25–27, 28, 32, 33, 38, 43, 44]	14	5137
Type of comparison	Different drugs [7–10, 12, 13, 15–27, 30–34, 37, 38, 40–44]	26	9306
	Same drugs but different doses/schedules [11, 14, 15, 28, 29, 35, 36, 39]	7	1485
Statistically significant clinical outcomes	Response [11, 13–17, 19, 23, 30–32, 34, 38, 41, 42]	12	4454
	Time to progression or treatment failure [9, 14, 15, 17, 19, 23, 25, 27, 28, 30–32, 34, 35, 37, 38, 41, 42]	13	4091
	Survival [9, 10, 23, 25, 27, 32, 37]	6	2065
	At least one of three [9, 10, 11, 13–17, 19, 23, 25, 27, 28, 30–32, 34, 35, 37, 38, 41, 42]	18	6024

Table 3. Methods adopted to measure QoL

		No. studies	No. pts randomized
Questionnaires	LASA[a] based questionnaires (Priestman and Baum) [7–10, 14, 15, 17, 21, 25, 28, 37]	10	2621
	European Organization for Research and Treatment of Cancer (EORTC) Quality of Life questionnaire C30 [13, 16, 19, 20, 23, 27, 30–35, 38, 42]	10	3725
	Rotterdam Symptoms Check List [12, 20, 24, 26, 30, 31, 34, 36]	6	1573
	Spitzer index[b] [10, 14, 15, 35]	4	979
	Functional Living Index-Cancer (FLIC) [11, 18 ,29]	3	1276
	Functional Assessment of Cancer Therapy (FACT) Breast [41, 43, 44]	3	1534
	Other questionnaires[c]	10	2863
Number of questionnaires used for each trial	1 [7, 11–13, 16, 18, 19, 22–24, 26, 28, 32, 33, 36–38, 40–44]	20	7051
	2 [8–10, 14, 15, 20, 21, 25, 27, 30, 31, 34, 35, 39]	11	3332
	3 or more [17, 29]	2	408
Duration of QoL assessment (29 studies, 9382 pts) mean 35.8 weeks	Up to 12 weeks [28, 29, 36]	3	517
	13–24 weeks [10, 11, 17, 19, 20, 23, 30, 31, 33, 34, 38, 39, 41]	10	2821
	25 weeks or more [12, 14, 15, 22, 24, 26, 35, 40]	7	2097
	Until progression [8, 13, 16, 18, 27, 32, 42–44]	9	3947
Timing of questionnaires (32 studies, 10458 pts) mean 7.3 weeks	Every 2–4 weeks [9–13, 17, 19, 22, 23, 25, 29, 37–40, 43, 44]	14	4007
	Every 6–16 weeks [8, 14–16, 18, 20, 21, 24, 26–28, 30–36, 41, 42]	18	6451

[a] Linear Analogue Self-Assessment (LASA) scale, also known as Visual Analogue Scale (VAS) consists of a 10 cm line anchored at both ends with words describing the extremes of the dimension being measured.
[b] HRQoL as evaluated by clinician; never used as sole instrument of assessment.
[c] Other questionnaires were (no. studies, no. pts randomized): Brunner [22, 40] (1, 260), ECOG Analgesic [21] (1, 648), HADS [20] (1, 116), Nottingham [17] (1, 40), Pacis [8] (1, 177), POMS [39] (1, 133), Qualitator Diary [17] (1, 40), Rand [29] (1, 368), Southwest [9,25] (1, 179), Tannock [39] (1, 133), TRSS [27] (1, 769).

making process for choosing the best treatment option. Three of these 12 studies considered QoL as a primary endpoint [14, 15, 35, 36]. Eight trials (2828 patients, 26%) [12, 14, 15, 22, 24, 26, 27, 29, 36, 40] reported significantly ($p < 0.05$) more favourable outcomes in terms of QoL (five compared different drugs, three compared different schedules) and in seven it was also possible to identify the specific QoL dimension involved (one only physical dimension [24], two only psychological dimension [26, 36], four both [12, 14, 15, 27, 29]). Two of the eight studies considered QoL as a primary outcome measure [14, 15, 36].

The relationship between clinical outcomes (response, time to progression/treatment failure, survival, toxicity) and QoL scores is shown in Table 5. Clinical efficacy was considered better when a significant improvement was reported in at least one of the traditional clinical endpoints.

Five trials (1156 patients, 14%) [8, 14, 15, 17, 30, 31, 33, 34] evaluated pre-treatment QoL status as a predictor of survival and four of them (979

Table 4. Compliance rates for completion of QoL questionnaires

Compliance	% (range)
Baseline form (18 trials, 5032 patients) [8, 14, 15, 17, 19, 20, 23, 24, 26, 29–36, 38, 39, 41, 42]	85.7 (64–100)
Overall (20 trials, 5601 patients) [8, 11, 13, 14, 15, 17, 19, 20, 23, 24, 26, 28, 29–36, 38, 39, 41, 42]	67.0 (30–90)

Table 5. Relation between classical clinical endpoints and QoL results

Clinical efficacy and toxicity profiles[a]	Statistically significant difference in Qol scores		Total	Mean methodological quality score
	Yes	No		
Better clinical efficacy and better toxicity	1 (769) [27]	2 (760) [10, 16]	3 (1529)	3.3
Better clinical efficacy but worse toxicity[b] [5, 17, 19, 23, 28, 30, 31, 34, 37, 38, 41, 42]	0	9 (2763)	9 (2763)	5.2
Better clinical efficacy and same toxicity	1 (305) [14, 15]	4 (971) [9, 13, 25, 32, 35]	5 (1276)	6.6
Same clinical efficacy but different toxicity	4 (1309) [22, 24, 26, 29, 40]	4 (759) [7, 8, 20, 39, 42]	8 (2068)	4.6
Same clinical efficacy and same toxicity	2 (445) [12, 36]	5 (2254) [18, 21, 33, 43, 44]	7 (2699)	4.3
Total	8 (2828)	24 (7507)	32 (10335)	4.8

[a] "Better", "worse" or "different" when p was at least <0.05; please note that "same" is not to be interpreted as "equivalence" but simply indicates "a non-significant result".
[b] One study excluded because QoL analysis was not done [11].

pts) [14, 15, 17, 30, 31, 33, 34] indicated that QoL had significant prognostic value for survival.

Quality of measuring and reporting QoL

Global scores for methodological quality over time are illustrated in Figure 1, where the area of each circle has been made proportional to the sample size. Although there was a weak direct association in quality score over time it did not reach statistical significance (Spearman's correlation coefficient = 0.29, $p = 0.10$). The methodological score was marginally better in trials that considered QoL as primary endpoint (mean scores 5.8 versus 4.7) and in those that found significant differences in QoL results (mean scores 5.1 versus 4.8), and slightly worse in studies reporting QoL findings in their abstracts (mean scores 4.5 versus 5.0). None of these comparisons reached statistical significance.

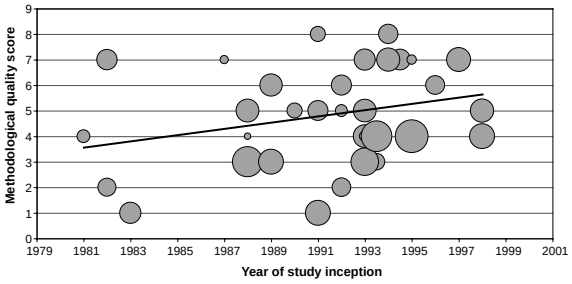

Figure 1. Methodological quality score by year of study inception.

Discussion

This review indicates that after two decades of randomised trials enrolling thousands of patients the contribution of QoL evaluation to raising clinical research standards and to our understanding of patients with metastatic breast cancer is only partially satisfactory. As expected, the clinical setting was favourable to QoL assessment since marked differences in classical endpoints were rare (e.g. only six therapeutic regimens imparted a survival benefit over the comparison and,

overall, a little more than half the studies showed a significant difference in at least one clinical endpoint, see Table 2). Nevertheless, the efficiency of QoL research was still sub-optimal as only 12 trials reported QoL findings in the abstract where one would expect to find data that warrant a prominent place in the results; only eight of the 33 trials reported significant differences in QoL scores. Even out of the seven that considered QoL as a primary outcome measure, only three reported QoL data in the abstract, and significant differences were found in two.

The results in Table 5 do not lend support to the notion that there is any correlation between better clinical efficacy or toxicity profiles and better QoL [45]. Significant differences in QoL scores were reported in only two of 17 studies with better clinical endpoints and only six of 20 with significant differences in toxicity profiles. It is remarkable that no difference in QoL emerged in the nine studies in which one treatment option was more effective even though more toxic than the comparator. Although the relationship between response and well-being has been observed in non-randomised [46] and randomised [17, 47, 48] clinical trials in ABC and other advanced forms of cancer [49, 50] the lack of relationship in this subset of seven studies is still surprising. It would appear that the better efficacy of these treatments was offset by the overall burden of toxic effects, thus cancelling any subjective clinical benefit. However, in 13 trials (rows 2 and 4 of Table 5) the significant difference in toxicity – in four of them not even counterbalanced by greater efficacy – did not translate into significant differences in QoL score.

It must be borne in mind that RCTs in ABC are rarely double or even single blinded [51] and therefore patients randomised to receive a promising new agent may experience feelings of optimism and well-being due to expectations about the treatment itself, thus compensating any negative impact of treatment toxicity (''hope bias'') [14, 46, 52–55]. Moreover, some patients even expect some relation between the severity of side effects and treatment efficacy. Finally, two of seven studies in the bottom row of Table 5 yielded significant differences in QoL data, thus providing useful elements to help clinicians choose between treatments with similar traditional clinical profiles (non-significant differences in efficacy or toxicity).

Of course, another plausible explanation is that a significant proportion might actually be false-negative studies. This might be due to the small sample sizes of some studies but, more likely, to the concurrence of several procedural and methodological problems that still plague QoL research, such as those listed here.

(a) Differences in tools and timing

Seventeen different questionnaires for the 33 studies reviewed seem too many, although the increase in popularity of the EORTC and the RSCL questionnaires is certainly encouraging. Thirteen studies used more than one instrument thus further enhancing the lack of uniformity in the QoL domains measured.

Length of observation covered just the duration of treatments for some studies but lasted until progression for others. In the first case it is hard to believe that the assessment could usefully encompass the whole time frame during which treatments elicited their effects on QoL. The timing of questionnaires spanned from every 2 weeks to every twelve and often changed during the study, being more frequent while on active treatment and further apart during follow-up. The reports also often overlooked other important logistic aspects such as the exact timing of completing the QoL forms (assessment completed immediately before or after treatment) in relation to the comparison groups. When therapies cause cyclical changes in patients' well-being, like chemotherapy, the inappropriate intervals between the study arms cannot be ignored [56].

Overall, this methodological heterogeneity hampers cross-study comparisons and may obscure the practical clinical implications of QoL results.

(b) Missing data

The overall compliance rate for completion of questionnaires is vital for an unbiased, longitudinal QoL assessment, as differences in response can introduce serious selection bias. Although there is strong evidence that missing QoL data are rarely ''missing by chance'' [57], since patients failing to complete QoL questionnaires often have the lowest scores [54], in many trials substantial amounts

of missing data could be "administratively missing", i.e missing data due to not informative causes [58]. Sorting out these alternatives is important for unbiased treatment comparisons, but even so 13 out of 33 studies reported no information at all about compliance, and the other 20 studies usually provided only rough overall estimates (the ideal is to report the proportion of patients completing assessments as scheduled separately for each treatment arm). As an assessment restricted to patients returning the questionnaires may overestimate QoL, paradoxically a poorly performing treatment might be judged more positively due to this selection/drop-out bias. Although the wide variability in compliance rates (from 30 to 100%) suggests that the most important determinant of compliance in these studies presumably involves the logistic aspects of QoL data collection and not only a deterioration in patients' condition, it is impossible to distinguish these reasons for missing data.

Strategies to minimize the biases due to incomplete data sets, such as the use of auxiliary QoL outcomes (proxy ratings of health status) or specific models and methods for handling non-random missing data, are still debated [57, 59, 60] and were used very little in the studies considered in this review.

(c) Differences in analyses

QoL measurement and analysis require trained experts since QoL data are much more difficult to interpret than, for example, survival time or toxicity ratings. Although most of the summary measures and the statistical techniques adopted in these studies can be considered appropriate, each method having specific advantages and disadvantages, such disparate methodological approaches certainly do not help clinicians assess the meaning of QoL data better. In the absence of absolute units of QoL, which are probably impossible to define, and of a strictly standardized way to report differences, constructive comparison of QoL findings across studies and the calibration of QoL scales against objective reality will remain an elusive goal. Even when we tried to translate these QoL results into a measure of the magnitude of the effect that is metric-free, the effect size, we were frustrated by the fact that only five out of 16 studies using the well known and more recent EORTC, RSCL or FACT questionnaires gave sufficient data to estimate this parameter (group mean difference and its standard deviation).

Concluding remarks

Less than ten years ago, QoL was still considered a new and exciting field in oncology [61], a new tool to enhance our understanding of patients' perceptions of the treatment burden. QoL has been vigorously promoted by the pharmaceutical industry, which is eager to expand classical efficacy parameters with marketing aims [62–63]. While patients' evaluations still remain the point of reference in a theoretical decision-making process, QoL, as it has been assessed so far, does not seem to constantly translate this expectation into a clinical reality, at least in ABC patients. The inconsistent correlation between traditional clinical endpoints and QoL results adds to the methodological shortcomings of QoL assessments that jeopardize internal validity, and casts doubts on the uncritical use of patients' evaluations as a "standard" against which to measure clinical benefit. Moreover, patients' psychosocial responses in a palliative setting largely depend on the illness trajectory, that too often moves fast toward its unfavourable outcome. In such cases possible differences in QoL related to treatments might actually be obscured by the "background noise" of the disease progression [29]. Accordingly, regulatory agencies such as the US Food and Drug Administration and the European EMEA do not require this kind of data for market authorization of anti-cancer drugs and only recommend QoL assessment as complementary evidence to support traditional clinical outcome measures [64].

Finally, a minority of studies took into account the relationship between baseline QoL scores and prognosis. The existence of such a correlation is an interesting by-product of QoL research [48] but currently its clinical relevance in the choice of treatments seems marginal.

This paper has several limitations that may hamper the validity of results. Most are related to the different types of selection bias that are intrinsic to any data-pooling approach. Since the principal selection bias is the publication bias that makes positive results more likely to be published,

one might assume that this situation should favour and not devalue the clinical relevance of QoL assessment. Secondly, the score we derived from the literature to classify the papers according to their methodological quality is a un-validated tool used for the first time on this set of papers. Finally, we cannot deny that we had a negative inclination toward QoL while commenting on the results and thus we recognize that the "QoL glass" in ABC might be half-full rather than half-empty.

Recently, two other reviews addressed QoL in breast cancer patients [3, 4]. These provide a through analytical description of most of the studies we analysed and, overall, make the same criticisms about QoL research so far. The aim of our work was to summarize the published findings of QoL in ABC patients and give a summary picture of the overall yield of two decades of QoL assessment while highlighting the relations between QoL and classical outcome measures and outlining the methodological quality of these assessments over time.

If we still aim at rigorous evaluation of how much QoL can contribute to balance treatment activity with considerations of well-being, and since the methodology of QoL assessment has been rather unsatisfactory so far, we recommend that investigators conform to the methodological standards promoted by experts and regulatory agencies [3–5, 64, 65]. Although we should beware of the seductive numeracy of the score we assembled to classify QoL methodology, Figure 1 shows that QoL science has already moved ahead but there is still room for improvement.

The QoL measures used in this series were developed at a time when researchers in the field of health outcome assessment believed that most of the reliable variability in self-measured outcomes was due to the clinical effect of treatments and QoL questionnaires were believed to capture a monotonic balance between efficacy and toxicity, with few alternatives. Nowadays, the QoL questionnaire that best matches our particular purpose should bear in mind that the control of common side effects of cancer therapies such as nausea and vomiting has improved dramatically over the last decade and that the specific impact of the disease and its treatments on patients' life is constantly evolving [66]. This implies that we have to keep on searching for reliable and finely tuned QoL instruments. A final warning: we should not fail to recognize that in ABC patients QoL can be deeply affected not only by the type of treatment but also by other factors usually not pertaining to medical care and medical interventions such as different degrees of social support.

References

1. Tassinari D: Surrogate end points of quality of life assessment: have we really found what we are looking for? Health Qual Life Outcomes 1(1): 71, 2003
2. Guyatt GH, Feeny DH, Patrick DL: Measuring health related quality of life. Ann Intern Med 118: 622–629, 1993
3. Bottomley A, Therasse P: Quality of life in patients undergoing systemic therapy for advanced breast cancer. Lancet Oncol 3: 620–628, 2002
4. Goodwin PJ, Black JT, Bordeleau LJ, Ganz PA: Health-related quality-of-life measurement in randomized clinical trials in breast cancer-taking stock. J Natl Cancer Inst 95: 263–281, 2003
5. Chassany O, Sagnier P, Marquis P et al.: Patient reported outcomes: the example of health related quality of life-a European guidance for the improved integration of HRQoL assessment in the drug regulatory process. Drug Inf J 36: 209–238, 2002
6. Jadad AR: Randomized controlled trias. A user's guide. BMJ Books, BMA House, Tavistock Square, London WC1H 9JR, 1998
7. Bennet JM, Muss HB, Doroshow JH et al.: A randomized multicenter trial comparing mitoxantrone, cyclophosphamide, and fluorouracil with doxorubicin, cyclophosphamide, and fluorouracil in the therapy of metastatic breast carcinoma. J Clin Oncol 6: 1611–1620, 1988
8. Bernhard J, Castiglione-Gertsch M, Hsu Schmitz SF et al.: Quality of life in postmenopausal patients with breast cancer after failure of tamoxifen: formestane versus megestrol acetate as second-line hormonal treatment. Eur J Cancer 35: 913–920, 1999
9. Bertsch LA and Donaldson G: Quality of life analyses from vinorelbine (Navelbine) clinical trials of women with metastatic breast cancer. Semin Oncol 22 (Suppl 5): 45–54, 1995
10. Bishop JF, Dewar J, Toner GC et al.: Initial paclitaxel improves outcome compared with CMFP combination chemotherapy as front-line therapy in untreated metastatic breast cancer. J Clin Oncol 17: 2355–2364, 1999
11. Brufman G, Colatori E, Ghilezan N et al.: Doubling epirubicin dose intensity (100 mg/m^2 versus 50 mg/m^2) in the FEC regimen significantly increases response rates. An international randomised phase III study in metastatic breast cancer. Ann Oncol 8: 155–162, 1997
12. Buzdar AU, Jones SE, Vogel CL et al.: A phase III trial comparing anastrozole (1 and 10 milligrams), a potent and selective aromatase inhibitor, with megestrol acetate in postmenopausal women with advanced breast carcinoma. Cancer 79: 730–739, 1997
13. Chan S, Friedrichs K, Noel D et al.: Prospective randomised trial of docetaxel versus doxorubicin in patients with

metastatic breast cancer. J Clin Oncol 17: 2341–2354, 1999
14. Coates A, Gebski V, Stat M et al.: Improving the quality of life during chemotherapy for advanced breast cancer. A comparison of intermittent and continuous treatment strategies. N Engl J Med 317: 1490–1495, 1987
15. Coates A, Gebski V: Quality of life studies of the Australian New Zealand Breast Cancer Trial Group: approaches to missing data. Statist Med 17: 533–540, 1998
16. Dombernowsky P, Falkson SG, Leonard R et al.: Letrozole, a new oral aromatase inhibitor for advanced breast cancer: double-blind randomised trial showing a dose effect and improved efficacy and tolerability compared with megestrol acetate. J Clin Oncol 16: 453–461, 1998
17. Fraser SCA, Dobbs HJ, Ebbs SR et al.: Combination or mild single agent chemotherapy for advanced breast cancer? CMF vs epirubicin measuring quality of life. Br J Cancer 67: 402–406, 1993
18. Goss PE, Winer EP, Tannock IF et al.: Randomised phase III trial comparing the new potent and selective third-generation aromatase inhibitor vorozole with Megestrol acetate in postmenopausal advanced breast cancer patients. J Clin Oncol 17: 52–63, 1999
19. Hakamies-Blomqvist L, Luoma ML, Sjostrom J et al.: Quality of life in patients with metastatic breast cancer receiving either docetaxel or sequential methotrexate and 5-fluorouracil. A multicentre randomised phase III trial by the Scandinavian Breast Group. Eur J Cancer 36: 1411–1417, 2000
20. Harper-Wynne C, English J, Meyer L et al.: Randomized trial to compare the efficacy and toxicity of cyclophosphamide, methotrexate and 5-fluorouracil (CMF) with methotrexate mitoxantrone (MM) in advanced carcinoma of the breast. Br J Cancer 81: 316–322, 1999
21. Hayes DF, Van Zyl JA, Hacking A et al.: Randomised comparison of tamoxifen and two separate doses of toremifene in postmenopausal patients with metastatic breast cancer. J Clin Oncol 13: 2556–2566, 1995
22. Heidemann E, Souchon R, Stoger H et al.: First-line monochemotherapy with mitoxantrone versus combination with fluorouracil, epirubicin and cyclophosphamide in high-risk metastatic breast cancer: a prospective randomized multicenter clinical trial. Onkologie 23: 54–59, 2000
23. Jassem J, Pienkowski T, Pluzanska A et al.: Doxorubicin and paclitaxel versus fluorouracil, doxorubicin, and cyclophosphamide as first-line therapy for women with metastatic breast cancer: final results of a randomized phase III multicenter trial. J Clin Oncol 19: 1707–1715, 2001
24. Joensuu H, Holli K, Heikkinen M et al.: Combination chemotherapy versus single-agent therapy as first- and second-line treatment in metastatic breast cancer: a prospective randomized trial. J Clin Oncol 16: 3720–3730, 1998
25. Jones S, Winer E, Vogel C et al.: Randomized comparison of vinorelbine and melphalan in anthracycline-refractory advanced breast cancer. J Clin Oncol 13: 2567–2574, 1995
26. Jonat W, Howell A, Blomqvist C et al.: A randomized trial comparing two doses of the new selective aromatase inhibitor anastrozole (Arimidex) with megestrol acetate in postmenopausal patients with advanced breast cancer. Eur J Cancer 32A: 404–412, 1996
27. Kaufmann M, Bajetta E, Dirix LY et al.: Exemestane is superior to megestrol acetate after tamoxifen failure in postmenopausal women with advanced breast cancer: results of a phase III randomized double-blind trial. J Clin Oncol 18: 1399–1411, 2000
28. Kloke O, Klaassen U, Oberhoff C et al.: Maintenance treatment with medroxyprogesterone acetate in patients with advanced breast cancer responding to chemotherapy: results of a randomized trial. Breast Cancer Res Treat 55: 51–59, 1999
29. Kornblith AB, Hollis DR, Zuckerman E et al.: Effect of megestrol acetate on quality of life in a dose-response trial in women with advanced breast cancer. J Clin Oncol 11: 2081–2089, 1993
30. Kramer JA, Curran D, Piccart M et al.: Randomised trial of paclitaxel versus doxorubicin as first-line chemotherapy for advanced breast cancer: quality of life evaluation using the EORTC QLQ-C30 and the Rotterdam Symptom Checklist. Eur J Cancer 36: 1488–1497, 2000
31. Kramer JA, Curran D, Piccart M et al.: Identification and interpretation of clinical and quality of life prognostic factors for survival and response to treatment in first-line chemotherapy in advanced breast cancer. Eur J Cancer 36: 1498–1506, 2000
32. Nabholtz J-M, Senn HJ, Bezwoda WR et al.: Prospective randomized trial of docetaxel versus mitomycin plus vinblastine in patients with metastatic breast cancer progressing despite previous anthracycline-containing chemotherapy. J Clin Oncol 17: 1413–1424, 1999
33. Norris B, Pritchard KI, Myles J et al.: Phase III comparative study of vinorelbine combined with doxorubicin versus doxorubicin alone in disseminated metastatic/recurrent breast cancer: National Cancer Institute of Canada Clinical Trials Group Study MA8. J Clin Oncol 18: 2385–2394, 2000
34. Paridaens R, Biganzoli L, Bruning P et al.: Paclitaxel versus doxorubicin as first-line single-agent chemotherapy for metastatic breast cancer: a European Organization for Research and Treatment of Cancer randomized study with cross-over. J Clin Oncol 18: 724–733, 2000
35. Riccardi A, Tinelli C, Brugnatelli S et al.: Doubling of the epirubicin dosage within the 5-fluorouracil epirubicin and cyclophosphamide regimen: a prospective, randomized, multicentric study on antitumor effect and quality of life in advanced breast cancer. Inter J Oncol 16: 769–776, 2000
36. Richards MA, Hopwood P, Ramirez AJ et al.: Doxorubicin in advanced breast cancer: influence of schedule on response, survival and quality of life. Eur J Cancer 28A: 1023–1028, 1992
37. Stewart DJ, Evans WK, Shepherd FA et al.: Cyclophosphamide and fluorouracil combined with mitoxantrone versus doxorubicin for breast cancer: superiority of doxorubicin. J Clin Oncol 15: 1897–190, 1997
38. Sjostrom J, Blomqvist C, Mouridsen H et al.: Docetaxel compared with sequential methotrexate and 5-fluorouracil in patients with advanced breast cancer after anthracycline failure: a randomised phase III study with crossover on progression by the Scandinavian Breast Group. Eur J Cancer 35: 1194–1201, 1999
39. Tannock IF, Boyd NF, DeBoer G et al.: A randomised trial of two dose levels of cyclophosphamide, methotrexate, and

fluorouracil chemotherapy for patients with metastatic breast cancer. J Clin Oncol 6: 1377–1387, 1988
40. Heidemann E, Stoeger H, Souchon R et al.: Is first-line single-agent mitoxantrone in the treatment of high-risk metastatic breast cancer patients as effective as combination chemotherapy? No difference in survival but higher quality of life were found in a multicenter randomised trial. Ann Oncol 13: 1717–1729, 2002
41. 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 21: 588–592, 2003
42. Nabholtz JM, Falkons 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 21: 968–975, 2003
43. 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 20: 3386–3395, 2002
44. Howell A, Roberstron JFR, Quaresma Albano J 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 20: 3396–3403, 2002
45. Cardoso S, Di Leo A, Lohrisch C et al.: Second and subsequent lines of chemotherapy for metastatic breast cancer: what did we learn in the last two decades? Ann Oncol 13: 197–207, 2002
46. Ramirez AJ, Towlson KE, Leaning MS et al.: Do patients with advanced breast cancer benefit from chemotherapy? Br J Cancer 78: 1488–1494, 1998
47. Coates A, Gebski V, Signorini D et al.: Prognostic value of Quality-of-Life scores during chemotherapy for advanced breast cancer. J Clin Oncol 10: 1833–1838, 1992
48. Bernhard J, Thurlimann B, Hsu Schmitz S-F et al.: Defining clinical benefit in postmenopausal patients with breast cancer under second-line endocrine treatment: does qualità of life matter? J Clin Oncol 17: 1672–1679, 1999
49. Kaasa S, Mastekaasa A, Naess S: Quality of life of lung cancer patients in a randomized clinical trial evaluated by psychosocial well-being questionnaire. Ann Oncol 27: 335–342, 1988
50. Glimelius B, Hoffmar K, Olafsdottir M et al.: Quality of life during cytostatic therapy for advanced symptomatic colorectal carcinoma: a randomised comparison of two regimens. Eur J Cancer Clin Oncol 25: 829–835, 1989
51. Fossati R, Confalonieri C, Apolone G et al.: Does a drug do better when is new? Ann Oncol 13: 470–473, 2002
52. Coates AS, Hurny C, Peterson HF et al.: Quality-of-life scores predict outcome in metastatic but not early breast cancer. J Clin Oncol 18: 3768–3774, 2000
53. McEvoy MD, McCorkle R: Quality of life issues in patients with disseminated breast cancer. Cancer 66: 1416–1421, 1990
54. Seidman AD, Portenoy R, Yao T-J et al.: Quality of life in phase II trials: a study of methodology and predictive value in patients with advanced breast cancer treated with paclitaxel plus granulocyte colony-stimulating factor. J Natl Cancer Inst 87: 1316–1322, 1995
55. Cohen L, de Moor C, Amato RJ: The association between treatment-specific optimism and depressive symtomatology in patients enrolled in a phase I cancer clinical trial. Cancer 91: 1949–1955, 2001
56. Hakamies-Blomqvist L, Luoma M-L, Sjostrom J et al.: Timing of quality of life (QoL) assessments as a source of error in oncological trails. J Adv Nursing 35: 709–716, 2001
57. Bernhard J, Cella DF, Coates A et al.: Missing quality of data in cancer clinical trials: serious problems and challenges. Statist Med 17: 517–532, 1998
58. Fairclough DL, Gelber RD: Quality of life: statistical issues and analysis. In: Spilker B, (ed.) Quality of life and Pharmacoeconomics in Clinical Trials. Lippincott-Raven Publishers, Philadelphia, pp. 427–436, 1996
59. Simes J, Greatorex V, Gebski VJ: Practical approaches to minimize problems with missing quality of life data. Statist Med 17: 725–737, 1998
60. Fairclough DL, Peterson HF, Cella D et al.: Comparison of several model-based methods for analysing incomplete quality of life data in cancer clinical trials. Statist Med 17: 781–796, 1998
61. Winer EP: Quality-of-life research in patients with breast cancer. Cancer 74: 410–415, 1994
62. Cote I, Gregoire JP, Moisan J: Health-related quality of life measurement in hypertension. Pharmacoeconomics 18: 435–450, 2000
63. Apolone G, De Carli G, Brunetti M et al.: Health-related quality of life (HR-QOL) and regulatory issues. An assessment of the European Agency for the Evaluation for Medicinal Products (EMEA) recommendations on the use of HR-QOL measures in drug approval. Pharmaeconomics 19: 187–195, 2001
64. European Medicines Evaluation Agency. Note for guidance on evaluation of anticancer medicinal products in man. CPMP/EWP/205/95 rev.1 http://www.emea.eu.int/pdfs/human/ewp/020595en.pdf
65. Beitz J, Gnecco C, Justice R: Quality-of-Life end points in cancer clinical trial: the U.S. Food and Drug administration perspective. Monogr Natl Cancer Inst 20: 7–9, 1996
66. Carelle N, Piotto E, Bellanger A et al.: Changing patient perceptions of the side effects of cancer chemotherapy. Cancer 95: 155–163, 2002

Address for offprints and correspondence: Roldano Fossati, Department of Oncology, Laboratory of Clinical Research in Oncology, Istituto di Ricerche Farmacologiche Mario Negri, Via Eritrea 62, 20157 Milano, Italy; *Tel.:* +39-02-39014467; *Fax:* +39-02-33200231; *E-mail:* fossati@marionegri.it