Patient-reported outcomes in ovarian cancer clinical trials

M. L. Friedlander1* & M. T. King2

1Australia and New Zealand Gynaecology Oncology Group (ANZGOG) Level 4, Camperdown, New South Wales; 2Cancer Australia Chair in Cancer Quality of Life, PsychoOncology Co-Operative Research Group (PoCoG), School of Psychology, The University of Sydney, Sydney, New South Wales, Australia

There is general acceptance of the importance of incorporating patient-reported outcome (PRO) measures including health-related quality of life (HRQOL) into clinical trials, and there are now a number of guidance documents available on how to use PRO’s for regulatory authorities and in comparative effectiveness research. The methods used to collect, analyse and report PRO data in clinical trials have received considerable scrutiny, revealing many shortcomings in the standard of reporting of HRQOL in clinical trials as well as in how PRO’s have been selected and analysed in clinical trials. This has led to the recent Consolidated Standards of Reporting Clinical Trials—PRO extension statement which lays down a framework for selection and reporting analysis of PROs, either as primary or secondary trial end points, thus ensuring scientific rigour. Adherence to these guidelines can only improve the conduct of clinical trials and interpretation of their results, which may help avoid missing out on opportunities as in the past. We review pertinent literature on PRO measures and discuss how various recent PRO guidance documents should be applied to ovarian cancer clinical trials.

Key words: HRQOL, patient-reported outcomes, ovarian cancer

introduction

A ‘patient-reported outcome’ (PRO) is defined by the USA’s Federal Drug Agency as the ‘measurement of any aspect of a patient’s health status that comes directly from the patient without the interpretation of the patient responses by a physician or anyone else’ [1]. The term PRO is fast eclipsing ‘quality of life’ in the health lexicon as an all-encompassing term for a wide range of possible impacts of disease and treatment on a patient’s well-being. Just like the term ‘health-related quality of life’ (HRQOL), it encompasses symptoms of disease, side-effects of treatment, anxiety, depression and various aspects of functioning (e.g. physical, role, social, sexual). In the UK, the slight variant PRO measure has emerged, and is used for the same purpose. These terms, PRO and PRO measurement, solve the intractable problem of coming up with a standard definition of quality of life that fits all purposes, but still acknowledge and respect the centrality of patients’ perceptions to good health care, policy and research. Examining and measuring the patient’s subjective experience in prospective comparative effectiveness research (CER) is now considered essential to informed clinical decision-making and health policy [2]. Regulatory authorities such as the FDA and EMEA routinely consider evidence from PRO’s in the evaluation of treatment benefit and have provided guidance to the industry on how to use PRO’s to support labelling claims [1, 3]. There is a great deal of science involved in producing good-quality PRO measurements, and there are a number of key issues to be considered in their development [4]. There are also many challenges in the good conduct of PRO research, including how PRO measurements are selected, evaluated, analysed and interpreted [5, 6].

PRO measurements provide the means to validly and reliably quantify subjective information provided by patients in response to specific questions using carefully developed and rigorously validated instruments. There are many instruments to choose from. There is no ‘right’ PRO measure in any absolute sense; one needs to carefully select the best measures for any specific clinical trial from available candidate measures [7, 8]. Here, the ‘best’ questionnaire is the one that best matches the specific aims and objectives of the study. This in turn depends on the expected effects of interventions under study on the target patient population.

PROs have long been included in clinical trials, but their value and importance is now more widely appreciated and accepted. As a consequence, the methods used to collect, analyse and report PRO data are receiving greater scrutiny. Rather than simply adding an instrument to measure HRQOL as an afterthought in clinical trials, there needs to be a close dialogue with QOL experts and experienced statisticians during the design and planning phase of the clinical trial in order to select the most appropriate instruments for the study as well as getting them involved in the development of a statistical analysis plan in order to generate meaningful patient-centred outcome data. The results of PRO data analyses are commonly only briefly covered in the primary publication of clinical ovarian cancer trials or included in an on-line appendix or published in journals with a lower impact factor, published at a later date, or

*Correspondence to: Dr M. Friedlander, Australia and New Zealand Gynaecology Oncology Group (ANZGOG) Level 4, 92-94 Parramatta Road, Camperdown, New South Wales, Australia. E-mail: m.friedlander@unsw.edu.au

© The Author 2013. Published by Oxford University Press on behalf of the European Society for Medical Oncology. All rights reserved. For permissions, please email: journals.permissions@oup.com.
perhaps not at all. Given the importance of PRO’s, the detailed results should ideally be included in the primary publication or in a companion paper in the same journal.

Brundage et al. [9] recently raised important concerns about the standard of reporting HRQOL in clinical trials of which all clinical researchers should take note of. The authors carried out a systematic review of 794 randomised trials undertaken between 2002 and 2008 that reported HRQOL across a range of medical conditions and found that only 56% provided a rationale for the selected outcome measure, only 50% provided a HRQOL hypothesis, only 28% provided information about missing data and 36% did not discuss HRQOL findings in the context of other trial outcomes [9]. It is unlikely that the reporting of HRQOL in clinical trials for ovarian cancer is any better, and the ovarian cancer research community needs to step up to the challenge of improving design and reporting of PRO’s in clinical trials.

A PRO extension to the Consolidated Standards of Reporting Clinical Trials (CONSORT)—PRO statement has recently been published and its recommendations should be considered by all clinical trialists as well journal editors [10]. Essentially, it recommends 5 CONSORT-PRO checklist items be considered for randomised, clinical trials where PROs are the primary or secondary end points. These recommendations include: (i) that the pro(s) be identified as a primary or secondary outcome(s) in the abstract; (ii) a description of the hypothesis of the pros and relevant domains is provided; (iii) evidence for the validity and reliability of the pro instrument(s) be provided; (iv) statistical approaches for dealing with missing data be explicitly stated; and (v) that the pro-specific limitations of the study findings and generalizability of results to other populations and clinical practice be discussed [10]. Adhering to these recommendations should improve the reporting and interpretation of PROs in clinical trials.

**instruments to measure HRQOL**

HRQOL is a complex multidimensional construct with a range of conceptual definitions [11]. There is wide agreement that HRQOL assessment should include the core domains of physical, social and emotional functioning or well-being, as well as a number of disease-related or treatment-related symptoms such as pain, fatigue and nausea. Symptoms are often the main reason for administering treatment, and some PRO measurements cover only symptoms. Comprehensive coverage of symptom experience includes three aspects: prevalence, severity and distress [5]. A symptom assessment tool should also be easy to understand and complete, must be reliable and valid for the symptoms it is supposed to measure and ideally present minimal burden to patients and clinical research staff.

The recent effectiveness guidance document on PRO’s developed by the Center for Medical Technology Policy recommended 5 PRO measures: the European Organisation for Research and Treatment of Cancer (EORTC) core questionnaire, QLQ-C30; the Functional Assessment of Cancer Therapy-General (FACT-G); M. D. Anderson Cancer Centre’s Symptom Index (MDASI); PRO-CTCAE; and the Patient Reported Outcome Measurement Information System (PROMIS) [12]. These are all quite general and the authors made the point that the measure selection should be based on the needs of the study, the psychometric properties of the PRO measure and the characteristics of the population.

The QLQ-C30 and FACT-G are the two most widely used cancer-specific HRQOL questionnaires. Luckett et al. provide guidance in selecting between them [13]. They constitute the core questionnaires of the EORTC and the Functional Assessment of Chronic Illness Therapy (FACTIT) measurement systems. Both are modular systems, such that the general issues covered by QLQ-C30 and FACT-G, respectively, can be augmented by site- and/or treatment-specific modules. Both EORTC and FACTIT have developed ovarian cancer-specific modules, and because these are often used in clinical trials, they are briefly reviewed below, along with the core module of each suite.

The FACT-O, like all the FACTIT measures, starts with the 27 items of the FACT-G version 4 [14]. These cover the four core well-being domains: physical, functional, social/family and emotional well-being. The FACT-O also contains 12 additional concerns specific to ovarian cancer. The recall period is ‘the past 7 days’, and all items are rated on a 5-point Likert scale: 0 = ‘not at all’, 1 = ‘a little bit’, 2 = ‘somewhat’, 3 = ‘quite a bit’, 4 = ‘very much’. The scoring algorithm allows for eight summary scales: the four core well-being subscales, a subtotal of the 27 core items, a subtotal of the 12 ovarian-specific additional concerns, a grand total of the 39 items, and a trial outcome index (sum of the 17 physical and functional wellbeing items plus the 12 ovarian-specific items) [14]. The FACIT (www.facit.org/FACITOrg/Questionnaires) ovarian cancer symptom index (FOSI) is a shorter, more focused subset of the FACT-O items, which summarises as a single scale, offering both a shorter form for patients to complete and just one summary scale to analyse, report and interpret. The FOSI comes in two forms: the original eight items form [15, 16] and the more comprehensive 18-item form (NFOSI-18) [17]. The general FACIT scoring algorithm simply sums component items within a scale, after reversing negatively phrased items, such that a higher score represents better wellbeing and less symptom burden.

The item pool for the FACT-O was developed through semi-structured interviews with five gynaecological oncology nurses and 17 ovarian cancer patients with a range of disease severity, subsequently reviewed by an expert panel at the GOG meeting in 1995 [5]. The FACT-O was then found to be a reliable and valid assessment of quality of life of women with ovarian cancer in a study of 232 heterogeneous patients at MD Anderson Cancer Centre [18]. However, it should be noted that the study items are quite general and may not fully reflect all the disease specific symptoms that patients experience. The validation study included patients with both early and advanced ovarian cancer and almost half were receiving routine surveillance while most of those on chemotherapy were receiving first-line chemotherapy.

The EORTC QLQ-C30 version 3 is a widely used measure of HRQOL in oncology and is extensively validated [19]. It contains 30 items which assess five domains of functioning (physical, role, emotional, social, cognitive), global HRQOL, and nine symptoms which commonly occur in cancer, regardless of primary site (pain, fatigue, nausea, vomiting, constipation, diarrhoea, dyspnoea, problems with sleep,
It is noteworthy that neither the QLQ-OV28 nor the FACT-O were specifically developed and validated in patients with platinum resistant recurrent ovarian cancer, where the aim of treatment is symptom benefit and palliation. These instruments have been widely used in ovarian cancer clinical trials, including studies of patients with recurrent ovarian cancer where the symptom burden may be high. However, none of these trials has detected a difference in HRQOL between trial arms, and only rarely detected improvements in HRQOL over time, in patients with recurrent ovarian cancer [21].

The 3rd Ovarian Cancer Consensus meeting concluded that objective response rates and progression-free survival alone were inadequate end points of treatment in patients with platinum-resistant ovarian cancer and recommended the development of a specific instrument to measure symptom benefit that could be used in clinical trials [22]. In 2010, the 4th Ovarian Cancer Consensus meeting emphasised the need for the development and validation measures of clinical benefit end points in clinical trials, including HRQOL and, more generally, PROs [23].

The Symptom Benefit working group was established under the auspices of the Gynecologic Cancer Intergroup (GCIG) to address the recommendations of the 3rd and 4th Ovarian Cancer Consensus meetings, and has developed an instrument called Measure of Ovarian Cancer Symptoms and Treatment Concerns (MOST). It comprises 35 items. The first 15 items relate to disease symptoms, identified in a study of 126 patients with platinum-resistant recurrent ovarian cancer. A further 17 items deal with adverse effects of treatment. Each of these 32 symptoms is rated over a recall period of ‘the last 3–4 weeks’ (reflecting the typical time between cycles of chemotherapy) on an 11-point numeric rating scale, from 0 = ‘no trouble at all’ to 10 = ‘worst I can imagine’, with intermediate verbal anchors at 2 = ‘mild’, 5 = ‘moderate’ and 8 = ‘severe’. Pending psychometric analyses, we envisage two symptom index style summary scales: one based on the average of the 15 disease symptoms, the other on the 17 adverse effects of treatment. The MOST also contains three items that assess well-being: physical, emotional and overall, on a scale from 0 = worst possible to 10 = best possible, also ‘in the past 3–4 weeks’. These three items will be reported individually. This instrument is currently undergoing validation in stage 2 of the GCIG Symptom Benefit Study. The aim is that this instrument will offer both clinical utility and statistical efficiency in capturing changes in symptoms over time as well as the adverse effects of treatment. The two symptom index scales are intended to provide statistically efficient PRO measurements for clinical trials.

**Recommendations for incorporating PRO’s into clinical trials and clinical comparative research**

The Center for Medical Technology Policy (CMTP) has developed a detailed effectiveness guidance document (EGD) on incorporating PRO’s into clinical comparative research in adult oncology (http://www.cmtpnet.org) [12]. This was published online in May 2012 and has been well summarised by Basch et al. [24]. Significant effort and extensive consultation went into developing this guidance document with input from many individuals and organisations. The purpose of the document is to provide clear recommendations and guidance to clinical investigators on the appropriate inclusion of PRO measures and how they should be incorporated into prospective CER and is essential reading for all involved in clinical trial research. The ultimate aim is to provide good evidence and reliable data that reflects the patient experience. The FDA has also developed a separate PRO guidance document entitled ‘Patient Reported Outcomes Measures: Use in Medical Product Development to Support Labelling Claims’ which provides information regarding what is required if PRO’s are going to be used to support a US labelling claim [1].

Fifteen specific recommendations were made in the CMPT effectiveness guidance document which is divided into three categories including: (i) selection of measures; (ii) implementation methods and (iii) data analysis and reporting which are reviewed by Basch et al. as well and covered in detail in the full guidance document [12, 24]. The purpose of the effectiveness guidance document is to better align the design of clinical research with the information needs of patients, clinicians and funding agencies, and the guidelines are intended to set a minimum standard to ensure that studies directly measure patients’ reported experience. They make the point that patient subjective experience constitutes information that is essential to any study investigating a specific treatment or intervention and that PRO’s add value to clinical research and clinical care. PRO data are more reflective of underlying health status than clinician reporting and provides meaningful clinical outcomes including survival and treatment adverse effects. QOL experts and statisticians should be included early and be involved in the study design. They should be involved in the selection of the most appropriate instruments, based on the primary and secondary aims of the study, and should develop a sample size or power calculations for the key patient-reported end points. They should also play a key role in the development of a detailed statistical analysis plan for the PROs, which should address how missing patient reported data, which is a common problem, will be handled and reported [25]. There should be a concerted effort to engage patients and clinical trials research staff and all participating clinicians and impress upon them the
importance of collecting PRO data to limit the amount of missing data.

Traditionally, analyses of PRO data have focused on comparisons of means between study arms, but there may be better ways to analyse the results. For example reporting the proportion of patients experiencing a specific change from baseline at a predetermined time point may be a more meaningful way to look at the results. The CMPT effectiveness guidance document also suggests that the cumulative distribution of PRO changes from baseline be included in the statistical analysis plan. This is a powerful graphical display which allows the reader to readily assess the proportion of patients who experience change in a specific measure at a time point of interest compared with baseline, and to readily compare this between trial arms. This would be particularly useful in studies of symptomatic patients with recurrent ovarian cancer where the aim of treatment is palliation. This would enable the spectrum of response across the study population to be reported and both improvements and decrements in scores from baseline would be very helpful in evaluating the impact of treatment, both within and between treatments.

lessons to be learned regarding using PRO measurements in clinical trials in other cancers

A recent study comparing the quality of life and in patients with metastatic pancreatic randomised to either FOLFIRINOX or gemcitabine is particularly pertinent as there are many parallels with patients with platinum-resistant recurrent ovarian cancer with respect to symptom burden and survival, making the methodology very applicable to clinical trials in recurrent ovarian cancer [24]. The combination chemotherapy was associated with significantly more toxicity than gemcitabine but there was no difference between the treatment arms for EORTC QLQ-C30 domains, with the exception of diarrhoea which was worse in the combination arm. It is not unusual to find no or minimal differences between the treatment arms when overall quality of life is compared in clinical trials. However, the investigators also looked at the time until definitive deterioration (TUDD) in each QLQ-C30 scale. Importantly, they measured QOL every 2 weeks until progression—frequent measurement is important when trying to assess the impact of treatment on symptoms. They showed that TUDD was significantly greater for the combination arm for global health status (GHS), physical role, cognitive and social functioning as well as six symptom domains (fatigue, pain, nausea and vomiting, dyspnoea, anorexia and constipation). This was particularly well demonstrated in a Kaplan–Meier plot for TUDD > 20 points on the GHS scale [HR 4.7, confidence interval (CI) 2.3–9.5] [26]. This is a very informative way to graphically compare treatments and a good approach to analyse PRO data, particularly when the aim is to assess symptom benefit and the impact of treatment on various aspects of QOL. Similar approaches have been used in evaluating chemotherapy in lung cancer studies [27] and provide good examples for how to design and analyse the results of studies of palliative chemotherapy in populations of patients who are very symptomatic and have a relatively short survival.

conclusions

There is wide agreement and acceptance of the importance of HRQOL, and more generally, PROs, in clinical decision-making as well as for CER and informing health policy. Although HRQOL and PRO measurements are commonly included in ovarian cancer clinical trials, and may be the most important outcome in some studies, there are still problems and inconsistencies with how the data are analysed and reported which could be avoided or improved with relatively little additional effort at the time of protocol development. This has important clinical and ethical implications which has been well expressed by Calvert et al. who wrote ‘the ethics of enrolling a substantial number of patients in clinical trials in which HRQOL is recognised as an important outcome, yet is reported in a way that is ineffective in advising patient care or health policy is questionable’ [28]. There are now guidance documents available that detail what is required to improve the assessment of PROs in adult oncology as well as recently published recommendations by the CONSORT-PRO extension group that provides guidance on reporting of PROs in randomised, controlled trials where PROs are either a primary or secondary outcome.

funding

The GCIG Symptom benefit study has been supported by an NHMRC project grant ID 570893.

disclosure

The authors have declared no conflicts of interest.

references


