Development of the Measure of Ovarian Symptoms and Treatment concerns (MOST): aiming for optimal measurement of patient reported symptom benefit with chemotherapy for symptomatic ovarian cancer

M T King[1], M R Stockler[2], P Butow[1], R O'Connell[2], M Voysey[2], A M Oza[3], K Gillies[2], H S Donovan[4], Rebecca Mercieca-Bebber[1], J Martyn[2], K Sjoquist[2], and M L Friedlander[2],

[1] School of Psychology and Psycho-Oncology Co-operative Research Group (PoCoG), University of Sydney [2] ANZGOG and NHMRC Clinical Trials Centre, University of Sydney [3]Princess Margaret Hospital, Toronto [4] School of Nursing, University of Pittsburgh

ABSTRACT

Objective. To determine the optimal patient-reported outcome measure (PROM) for assessing symptom benefit in trials of palliative chemotherapy for women with symptomatic ovarian cancer.

Methods. Candidate PROMs were: EORTC QLQ-C30 plus ovarian-specific QLQ-OV28; Functional Assessment of Cancer Therapy-Ovarian (FACT-O); FACT-Ovarian Symptom Index (FOSI), gynecologic cancer-specific Symptom Representation Questionnaire (SRQ). Pre-defined optimality criteria were: inclusion of all symptoms necessary for the specified purpose; recall period covering typical length of palliative chemotherapy; numerical item rating scales; all necessary symptoms included in a single symptom index. Qualitative and quantitative methods were applied to data from Stage 1 of the Gynecologic Cancer Intergroup Symptom Benefit Study (GCIG-SBS) to determine the set of necessary symptoms, and to objectively assess candidate PROMs against the optimality criteria.

Results. Ten necessary symptoms were identified: pain, fatigue, abdominal bloating/discomfort, sleep disturbance, bowel disturbance, nausea and vomiting, shortness of breath, poor appetite, urinary symptoms and weight changes. While QLQ-C30/OV28 together cover all these symptoms, they split them into numerous scales, dissipating potential symptom benefit signal. Conversely, FACT-O does not cover all necessary symptoms and contains many other HRQOL-related items and treatment side-effects, diluting potential symptom benefit signal when summed into scales. Item response scales and composite scoring of all candidate PROMs were suboptimal to our specific purpose. We therefore developed a new PROM, the Measure of Ovarian Symptoms and Treatment concerns (MOST), to provide optimal measurement for the specified purpose.

Conclusions. This paper documents the development of the MOST, a new PROM designed to assess patient-reported benefits and burden as endpoints in clinical trials of palliative chemotherapy for women with symptomatic ovarian cancer. The validity, reliability and statistical efficiency of the MOST, relative to the best candidate scales of existing PROMs, will be assessed in Stage 2 of GCIG-SBS.

Key words. Health-related quality of life, patient reported outcomes, symptom benefit, symptom index, ovarian cancer

INTRODUCTION

The primary objective of chemotherapy for women with symptomatic ovarian cancer is symptom control. The 3rd Ovarian Cancer Consensus meeting concluded that response rates and progression free interval were inadequate measures of the palliative benefit of chemotherapy and that symptom control and health-related quality of life (HRQoL) should also be considered as measures of treatment benefit [1]. This was endorsed by the 4th Ovarian Cancer Consensus meeting [2]. The Symptom Benefit working group was convened under the auspices of the Gynecologic Cancer Intergroup (GCIG), and the GCIG Symptom Benefit Study (SBS) (Clinical trial register - ACTRN12607000603415) was initiated after extensive consultation.

The Symptom Benefit Study was designed in two stages. Stage 1 had two aims: 1) to describe and document the symptom burden, treatment and outcomes of women with platinum resistant/refractory recurrent ovarian cancer; 2) to determine the optimal instrument to measure subjective symptom benefit in clinical trials of palliative chemotherapy for ovarian cancer. The first aim is addressed in a companion paper [3]. We address the second aim, defining "optimal" as best able to provide efficient and focused measurement of symptom benefit for use in clinical trial endpoint analysis.

HRQoL Questionnaires (MOVED HERE FROM METHODS, AND SHORTENED)

Several patient-reported outcome measures (PROMs) have been designed to assess HRQoL in ovarian cancer [4]. The most widely used are the EORTC QLQ-OV28 [5], an ovarian-specific module used in conjunction with the core QLQ-C30 [6], and the FACT-O [7]. Since HRQoL is a multi-dimensional construct [4], these PROMs typically assess a range of issues, including various aspects of functioning and side effects of treatment, in addition to symptoms of ovarian cancer. Depending on the questionnaire's scoring algorithm, symptoms may be split among various scales or combined with other aspects of HRQoL. How this is done and the implications for clinical trial endpoint analysis are explored in this paper.

Symptom Indexes

Symptom indexes are PROMs which provide a more focused approach, typically including only those symptoms that are most likely to be ameliorated by palliative treatment, with all symptoms summed into a single index [8]. Two symptom indexes have been developed specifically for ovarian cancer, the Ovarian Symptom Index (FOSI) [8, 9] and the NFOSI-18 [10], subsets of 8 and 18 of the FACT-O items, respectively. The Symptom Representation questionnaire (SRQ), designed to assess the symptom experience of women with gynecologic cancers, also yields a composite index of symptom severity based on 22 items [11].

The objectives of GCIG-SBS Stage 1 addressed in this paper are:

1) to evaluate existing PROMs (QLQ-C30/OV28, FACT-O, FOSI-8, NFOSI-18 and SRQ) in terms of their optimality as measures of symptom benefit for use in endpoint analysis of clinical trials of palliative chemotherapy for ovarian cancer; and

2) if existing PROMs are found to be suboptimal in regard to 1), to develop a PROM optimally designed for that purpose.

METHODS

Participants and study procedures are described in detail in the companion paper [3]. The key points for this paper are: 1) patients with platinum resistant/refractory recurrent ovarian cancer self-completed a HRQoL booklet before starting palliative chemotherapy (baseline); and 2) a subset of patients completed a structured interview.

Patient-reported outcome measures

The HRQoL booklet included seven validated PROMs, in this order: Symptom Representation Questionnaire [11], FACT-O v4 [7], QLQ-C30 v3 [6], QLQ-OV28 [5], Patient Disease and Treatment Assessment Form (Pt DATA Form) [12]; Hospital Anxiety and Depression Scale (HADS) [13]; and Herth Hope Index (HHI) [14]. This paper utilizes baseline data from SRQ, QLQ-C30/OV28, and FACT-O/FOSI; data from the other timepoints and questionnaires are reported separately [3, 15].

Optimality criteria [NEW SECTION]

To determine the optimal PROM to measure subjective symptom benefit in clinical trials of palliative chemotherapy for ovarian cancer, we first defined "optimal" as best able to provide efficient and focused measurement of symptom benefit for use in clinical trial endpoint analysis. We operationalized this definition in terms of four optimality criteria derived from key elements of the first two steps of the process for developing a PRO instrument for use in clinical trials recommended by the Food and Drug Administration (FDA) [16]:

1) *content validity*: in this context, defined as: symptoms necessary and sufficient for the target population (women with symptomatic ovarian cancer) and purpose (to provide efficient and targeted measurement of symptom benefit for use in clinical trial endpoint analysis). The guiding principle was "everything that matters and no more", so items that were not necessary to this purpose were considered surplus to it;

2) *recall period*: should be long enough to capture all effects within a cycle of palliative chemotherapy, i.e. 3-4 weeks;

3) *numerical rating scale for items*: a systematic review has established that numerical rating scales have better responsiveness and ease of use and higher compliance rates relative to visual analogue or verbal rating scales [17];

4) *symptom index scoring*: whereby all necessary symptoms are summed into a single index, providing a statistically efficient trial endpoint [18].

As per the FDA's PRO guidance [16], qualitative and quantitative methods were used to objectively assess the optimality criteria.

Qualitative analysis

Structured interviews were conducted in a subsample (*n*=20) to assess patients' opinions about the completeness of coverage of issues in the HRQoL booklet, and importantly, whether any issues had been left out

(Criterion 1). Patients' preference for existing questionnaires and response formats were also elicited. Eligible patients: 1) were symptomatic at baseline; 2) completed ≥3 cycles of chemotherapy; 3) completed the HRQoL booklet prior to each cycle of chemotherapy. We sampled purposively to obtain a balanced mix of patients whose symptoms improved, persisted or worsened from baseline to their last HRQoL assessment, and until data saturation was achieved. Selected patients were invited to participate by their research nurse. An independent, professional interviewer then arranged a telephone interview with the participant, which was audiotaped. Participants received a summary of their HRQoL responses by post for reference during the interview. Transcribed interviews were content-analysed for reports of additional symptoms not covered in the HRQoL booklet, and patients' opinions about the comprehensiveness of coverage of relevant issues by the questionnaires, the ease of use of item response scales, and their preference among the questionnaires.

Quantitative Analysis

The purpose of the quantitative analysis was to determine which symptoms were necessary and sufficient for the target population and purpose (i.e. to specify the symptom set for Criterion 1). The SRQ was the ideal measure for this purpose, as it contains a comprehensive set of symptoms, allows patients to nominate a further three symptoms, and asks which three symptoms the patient has "noticed most in the last week" (referred to hereafter as the "Top 3"). We assessed two aspects of the SRQ data. First, symptoms that were noticed most across the sample were determined by calculating the frequency of each symptom nominated in the "Top 3" (including the three optional additional symptoms). We set a threshold of 5% as we judged this to be sufficiently common to warrant inclusion as an item in a symptom index, and that lower prevalence would add respondent burden for very little information gain. Second, the most prevalent or severe symptoms were determined by calculating the mean and standard deviation of ratings of all 22 symptoms listed on the SRQ. Analyses were conducted using SAS software, version 9.2 (SAS Institute).

Assessment of existing PROMs

The QLQ-C30/-OV28, FACT-O, FOSI and SRQ were assessed against the optimality criteria by considering item content (Criterion 1), recall period (Criterion 2), item response scale (Criterion 3) and scoring algorithm (Criterion 4).

[Development of the MOST – MOVED DOWN, CONSOLIDATED UNDER SAME HEADING IN RESULTS]

RESULTS

Participants

126 women participated in Stage 1 of GCIG-SBS, 124 completed baseline HRQoL booklets and 123 had at least 1 cycle of chemotherapy. Patient characteristics and details of their symptom experience at baseline are given in the companion paper [3]. Thirty women were approached for interviews, eight declined and 22 were interviewed, at which point data saturation had

been achieved. Two interviews were excluded due to poor quality audio recording. The 20 women included in the qualitative analysis ranged in age from 48 to 86 years (mean 65.5). At the time of interview: three participants had completed three cycles of chemotherapy, 13 had completed four cycles and four had completed five cycles; 10 reported at least two baseline symptom had improved by 2 or more points, three reported worsening by 2 or more points of at least two baseline symptoms, and seven remained stable.

Establishing the symptom set for content validity [AMENDED HEADING, PREVIOUSLY "Patients Top 3 Symptoms and coverage by PROMS"]

Interviewees generally agreed that all pertinent issues were covered in the HRQoL booklet, and none believed important issues were absent. Some participants did not complete the FACT-O sexuality items, stating that they were "intrusive" and not relevant. [THIS PARAGRAH MOVED UP, PREVIOUSLY UNDER HEADING "PATIENT INTERVIEWS", NOW REMOVED]

Table 1 summarises the 10 symptoms nominated by at least 5% of our sample as in the Top 3 symptoms are shown in Table 1, along with the mean (SD) rating. Items that were more frequently nominated in the Top 3 also tended to have higher severity ratings. On this basis, we considered these 10 symptoms necessary and sufficient to measure subjective symptom benefit in clinical trials of palliative chemotherapy for ovarian cancer, thereby establishing symptom set for Criterion 1.

Assessment of PROMs against the optimality criteria

Table 2 summarises the extent to which each of the candidate PROMs addresses the measurement optimality criteria. Table 3-5 provide further detail for Criteria 1 and 4.

Criteria1 and 4: content validity and item-into-scale aggregation

Table 3 summarises the number of items in QLQ-C30/OV28, FACT-O, FOSI-8 and NFOSI-18 that address the 10 necessary symptoms, and the number of items addressing other issues ("surplus items" for brevity).

Together, the QLQ-C30 and QLQ-OV28 cover all 10 necessary symptoms with a total of 20 items. The EORTC scoring algorithm integrates these 20 items into 10 separate summary scales, with a further 13 scales scored from the 38 surplus items (Table 4).

The FACT-O covers 8 of the 10 necessary symptoms with 10 items, asking a further 29 surplus items. The FACT-O scoring algorithm integrates the 10 necessary items into 8 non-independent summary scales (i.e. each item contributes to several scales) (Table 5). The scale with the highest ratio of necessary to surplus items is the trial outcomes index (TOI), containing all 10 necessary items plus 16 surplus items.

The FOSI-8 covers only 4 of the 10 necessary symptoms with 6 items, and the NFOSI-18 covers a further 2 of the 10 necessary symptoms (still missing 4), plus a further 8 surplus items (Table 5). The 9-item Disease-Related Symptoms – Physical (DSR-P) subscale of the NFOSI-18 contains the highest ratio of necessary to surplus items, containing 8 necessary items plus "I feel ill".

The SRQ contains all 10 necessary items, along with 12 others, and includes all 22 in a composite symptom index.

Criterion 2: Recall period

The recall period for all the existing PROMs is the past week.

Criterion 3: Item response scales

The EORTC and FACIT questionnaires use similar 4-point and 5-point verbal rating scales, respectively, while the SRQ uses an 11-point numeric rating scale. The verbal anchors for all these rating scales are described in the footnotes to Table 2.

Patients' preferences for existing PROMs

Overall, interviewees did not express consistent preferences between existing HRQoL questionnaires regarding they content or item response format; 17/20 did not have a preferred questionnaire, the SRQ and Patient DATA Form were preferred by one participant each, and one participant was not asked the question. Five participants preferred the EORTC 0-4 response format, three preferred the 0-10 format of the SRQ and Patient DATA Form, one participant preferred the EORTC Global QoL 0-7 format, one was not asked the question and the remaining 10 did not have a preference.

Participants explained the importance of presenting questions simply. Many participants found it difficult to complete questionnaires where the direction of coding was switched (i.e. FACT-O/FOSI). Some participants commented that a week was too short for the recall period.

The MOST (Measure of Ovarian Cancer Symptoms and Treatment Concerns)

Since none of the candidate PROMs met all the optimality criteria, we designed a new measure, the MOST, to meet all our optimality criteria, and taking into consideration patients' qualitative feedback. We adapted the Patient DATA form [12] for this purpose, adding the additional necessary symptoms. Following the FDA's PRO Guidance, we specified a conceptual framework that included two key concepts: symptom benefit and treatment burden. To address the latter, we included a comprehensive selection of chemotherapy side-effects, based on the items in the Treatment Concerns section of the Patient DATA Form [12].

The MOST contains 35 items (Figure 1): 15 assess disease symptoms, 17 assess adverse effects of treatment ("treatment concerns"), and 3 assess wellbeing (physical, emotional, overall). The wording and layout are simple, like that of the SRQ and Patient DATA Form.

The MOST has two forms, the main one being the recent status form, which aims to determine how troublesome various disease symptoms and treatment problems have been in the period between chemotherapy cycles; thus respondents are asked "how much that aspect troubled you on average during the last 3-4 weeks". This from is designed to be completed when patients attend for clinical assessment prior to their next cycle of chemotherapy. The 32 symptoms are rated 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”. The three wellbeing items range from 0 = “worst possible” to 10 = “best possible”. The MOST change form contains the same items, but asks patients to report their perceived change since “before you started this course of chemotherapy 6-8 weeks ago”. This form is designed to allow estimation of the minimally important difference [19, 20], which will be done in Stage 2 of GCIG-SBS.

DISCUSSION

Our analysis of data from Stage 1 of GCIG-SBS showed that the disease-related physical symptoms that were most noticed and most severe in this sample of women with platinum resistant/refractory recurrent ovarian cancer were pain, fatigue, abdominal bloating and discomfort, sleep disturbance, bowel problems, nausea and vomiting, shortness of breath, poor appetite, urinary symptoms and weight changes. Emotional problems were also commonly noticed and relatively severe, as were two treatment related symptoms (hot flushes, numbness and tingling). Our analysis of the content of existing ovarian-specific HRQOL questionnaires revealed that while the EORTC QLQ-C30 and QLQ-OV28 together cover all of these symptoms, their mandated scoring algorithms dissipate the impact of these symptoms by splitting them into numerous scales. The FACT-O covered all but two of the necessary symptoms, but also included many other HRQoL-related issues, leading to dilution in multi-item summary scales. The latter shortcoming has been recognized and addressed to some extent in the development of the two FACIT Ovarian Symptom Indexes [8-10]. However our analysis shows that even these are somewhat diluted by other issues. The scales that are most focused are the EORTC abdominal/gastrointestinal scale, FACT-O TOI, and FOSI-18 DSR-P.

This paper documents the preliminary development of the MOST, a new PROM designed to assess patient-reported benefits and burden as endpoints in clinical trials of palliative chemotherapy for women with symptomatic ovarian cancer. We used qualitative and quantitative methods to identify symptoms relevant to this patient population and clinical context, as recommended by the FDA for the development of PROMs for clinical trials [16] and for comparative effectiveness research [21]. We note that the MOST includes the 12 core symptoms recommended for consideration in studies in advanced or metastatic cancers [21]. Our choice of a numerical rating scale for the MOST items is supported by a systematic review showing that such scales have better responsiveness and ease of use and higher compliance rates relative to visual analogue or verbal rating scales [17]. Our work to this point, based on Stage 1 of the GCIG-SBS, covers Steps 1 and 2 of the iterative process recommended by the FDA for developing a PRO instrument for use in clinical trials (Figure 3, page 7) [16]. We will complete this iterative process, covering Steps 3-5 in Stage 2 of GCIG-SBS. This will include determining the scoring algorithm, psychometric properties, and interpretation guidelines (including the minimally important difference) for the MOST. Consistent with our conceptual framework, we envisage a scoring algorithm yielding one symptom index and one treatment problems index, such that the benefits and burden of chemotherapy can be quantified independently and

precisely in any given trial. Previous studies have confirmed that major gains in measurement precision can be achieved by including a relatively large number of items in a summary scale [18].

The MOST may also have broad clinical utility, with potential to enhance communication between patients and clinicians about symptom benefit and treatment burden and inform shared decision-making about whether or when to stop chemotherapy [22]. Using MOST in clinics would probably enable clinicians to better understand and treat the symptoms that matter most to these patients.

Arguably, the QLQ-C30/OV28, FACT-O and FOSI are currently the best available PROMs for ovarian cancer. Some of their scales are sensitive to differences in performance status, relapse versus no relapse, survivors versus general population, and chemotherapy naïve versus known neuropathy [4]. All six randomized controlled trials of chemotherapy in the platinum resistant/refractory recurrent ovarian cancer population that have assessed HRQoL have used QLQ-C30, FACT-O or FOSI [23, 24]. Yet only one has demonstrated changes over time in PROs [25], and none has demonstrated a difference in PROs between trial arms. We have developed the MOST as a complement, not a substitute, for these PROMs, with the aim of providing optimal assessment of subjective symptom benefit in future trials. A thorough validation of the MOST will be undertaken as part of Stage 2 of GCIG-SBS. Data are currently being collected in Australia, Canada, Denmark, England, France, Germany, Italy, Ireland, Japan, Sweden, and the United States of America. The QLQ-C30/OV28, FACT-O and FOSI are being used, alongside the MOST, in order to assess the MOST's measurement properties relative to those PROMs, in particular, responsiveness to clinically important change and relative statistical efficiency [18]. We will focus our comparisons on the scales that best target symptoms of ovarian cancer, i.e. QLQ-OV28 abdominal/gastrointestinal scale, FACT-O TOI, FOSI-18 DSR-P. If found to be valid, reliable and responsive, the MOST will be an excellent candidate PROM for clinical trials of ovarian cancer where the objective of treatment is symptom improvement.

References

1. du Bois, A., et al., *2004 consensus statements on the management of ovarian cancer: final document of the 3rd International Gynecologic Cancer Intergroup Ovarian Cancer Consensus Conference (GCIG OCCC 2004).* Annals of Oncology, 2005. **16**: p. 7-12.
2. Stuart, G.C., et al., *2010 Gynecologic Cancer InterGroup (GCIG) consensus statement on clinical trials in ovarian cancer: report from the Fourth Ovarian Cancer Consensus Conference.* International Journal of Gynecological Cancer, 2011. **21**(4): p. 750-5.
3. Friedlander, M.L., et al., *Symptom burden and outcomes of patients with platinum resistant/refractory recurrent ovarian cancer - a reality check! Results of Stage 1 of the GCIG Symptom Benefit Study.* International Journal of Gynecological Cancer, under review.
4. Luckett, T., et al., *Assessing health-related quality of life in gynecologic oncology: a systematic review of questionnaires and their ability to detect clinically important*

differences and change. International Journal of Gynecological Cancer, 2010. **20**(4): p. 664-84.
5. Greimel, E., et al., *An international field study of the reliability and validity of a disease-specific questionnaire module (the QLQ-OV28) in assessing the quality of life of patients with ovarian cancer.* Eur J Cancer, 2003. **39**(10): p. 1402-8.
6. Aaronson, N.K., et al., *The European Organization for Research and Treatment of Cancer QLQ-C30: a quality-of-life instrument for use in international clinical trials in oncology.* J Natl Cancer Inst, 1993. **85**(5): p. 365-76.
7. Basen-Engquist, K., et al., *Reliability and Validity of the Functional Assessment of Cancer Therapy–Ovarian.* Journal of Clinical Oncology, 2001. **19**(6): p. 1809-1817.
8. Cella, D., et al., *What are the most important symptom targets when treating advanced cancer? A survey of providers in the National Comprehensive Cancer Network (NCCN).* Cancer Investigation, 2003. **21**(4): p. 526-35.
9. Beaumont, J., et al., *Validation of the Functional Assessment of Cancer Therapy-Ovarian (FACT-O) Symptom Index (FOSI) in a phase II clinical trial of pertuzumab in patients with advanced ovarian cancer. .* 2007 ASCO Annual Meeting Proceedings Part I. Vol 25, No. 18S (June 20 Supplement), 2007. **25**: p. 18S.
10. Jensen, S.E., et al., *A new index of priority symptoms in advanced ovarian cancer.* Gynecologic Oncology, 2011. **120**(2): p. 214-9.
11. Donovan, H.S., et al., *Evaluation of the Symptom Representation Questionnaire (SRQ) for assessing cancer-related symptoms.* J Pain Symptom Management, 2008. **35**(3): p. 242-257.
12. Stockler, M.R., et al., *Effect of sertraline on symptoms and survival in patients with advanced cancer, but without major depression: a placebo-controlled double-blind randomised trial.* Lancet Oncol, 2007. **8**(7): p. 603-12.
13. Zigmond, A.S. and R.P. Snaith, *The hospital anxiety and depression scale.* Acta Psychiatr Scand, 1983. **67**(6): p. 361-70.
14. Herth, K., *Abbreviated instrument to measure hope: development and psychometric evaluation.* J Adv Nurs, 1992. **17**(10): p. 1251-9.
15. Sjoquist, K.M., et al., *Hope, Quality of Life and Benefit from Treatment in Women Having Chemotherapy for Platinum resistant/refractory Recurrent Ovarian Cancer - The GCIG Symptom Benefit Study.* The Oncologist, 2013 (published online ahead of print October 2013).
16. FDA, *Food and Drug Administration. Guidance for Industry on Patient-Reported Outcome Measures: Use in Medical Product Development to Support Labeling Claims.* Federal Register, 2009. **74**(235): p. 65132-65133.
17. Hjermstad, M.J., et al., *Studies comparing Numerical Rating Scales, Verbal Rating Scales, and Visual Analogue Scales for assessment of pain intensity in adults: a systematic literature review.* Journal of Pain & Symptom Management, 2011. **41**(6): p. 1073-93.
18. King, M.T., et al., *Responsiveness and relative statistical efficiency of QLQ-C30 versus FACT-G: implications for sample size requirements of health-related quality of life endpoints.* Journal of Clinical Epidemiology, in press (accepted 26/02/2013).
19. King, M.T., *A point of minimal important difference (MID): a critique of terminology and methods.* Expert Review of Pharmacoeconomics & Outcomes Research, 2011. **11**(2): p. 171-84.
20. Revicki, D., et al., *Recommended methods for determining responsiveness and minimally important differences for patient-reported outcomes.* Journal of Clinical Epidemiology, 2008. **61**(2): p. 102-9.
21. Basch, E., et al., *Recommendations for incorporating patient-reported outcomes into clinical comparative effectiveness research in adult oncology.* Journal of Clinical Oncology, 2012. **30**(34): p. 4249-55.

22. Valderas, J.M., et al., *The impact of measuring patient-reported outcomes in clinical practice: a systematic review of the literature.* Quality of Life Research, 2008. **17**(2): p. 179-93.
23. Mutch, D.G., et al., *Randomized phase III trial of gemcitabine compared with pegylated liposomal doxorubicin in patients with platinum-resistant ovarian cancer.* J Clin Oncol, 2007. **25**(19): p. 2811-8.
24. Colombo, N., et al., *Randomized, open-label, phase III study comparing patupilone (EPO906) with pegylated liposomal doxorubicin in platinum-refractory or -resistant patients with recurrent epithelial ovarian, primary fallopian tube, or primary peritoneal cancer.* J Clin Oncol, 2012. **30**(31): p. 3841-7.
25. ten Bokkel Huinink, W., et al., *Topotecan versus paclitaxel for the treatment of recurrent epithelial ovarian cancer.* J Clin Oncol, 1997. **15**(6): p. 2183-93.

Figure and Table Legends

Figure 1: The Measure of Ovarian Symptoms and Treatment concerns (MOST)

Table 1: Symptoms nominated most often by the patients in their "Top 3" symptomsa in the Symptom Representation Questionnaire (SRQ) at baseline, number (%) of patients who nominated each, and mean (SD) rating of each symptom at its worst (N = 124)

Table 2: Summary of the extent to which the existing candidate patient-reported outcome measures and the new questionnaire (MOST) met the optimality criteria for efficient and focused measurement of symptom benefit for clinical trial endpoint analysis.

Table 3: Number of questions in the QLQ-C30, QLQ-OV28, FACT-O and FOSI (8 and 18 item versions) that address the 10 symptoms that are necessary to measure subjective symptom benefit in clinical trials of palliative chemotherapy for ovarian cancer, and number that address other issues (surplus items).

Table 4: Number of summary scales in the QLQ-C30 and QLQ-OV28, and the number of questions in each that address the 10 symptoms necessary to measure subjective symptom benefit in clinical trials of palliative chemotherapy for ovarian cancer.

Table 5: Number of multi-item scales in the FACT-G, FACT-O and FOSI (8 and 18 item versions), and the number of items in each that address the 10 symptoms necessary and sufficient to measure subjective symptom benefit in clinical trials of palliative chemotherapy for ovarian cancer.

Table 1 Symptoms nominated most often by the patients in their "Top 3" symptoms[a] in the Symptom Representation Questionnaire (SRQ) at baseline, number (%) of patients who nominated each, and mean (SD) rating of each symptom at its worst (N = 124)

"Top 3" Symptoms[a]	**No. of patients**	**%**	**Mean**[b] **(SD)**
Pain	55	44	3.7 (3.1)
Fatigue	46	37	4.2 (2.8)
Abdominal bloating/discomfort	39	32	3.6 (3.2)
Sleep disturbance	30	24	3.6 (3.1)
Bowel problems	27	22	3.7 (3.4)
Nausea/vomiting	24	19	2.1 (2.6) Nausea
			1.1 (2.4) Vomiting
Emotional	20	16	2.5 (2.6) Depression
			1.9 (2.2) Mood swings
(Numbness/tingling)	13	11	1.6 (2.3)
Shortness of breath	13	11	2.0 (2.5)
Poor Appetite	12	10	2.7 (2.9)
(Hot flushes)	11	9	1.6 (2.4)
Urinary problems	9	7	1.4 (2.5)
Weight changes	7	6	1.2 (2.2) Weight loss
			0.9 (2.3) Weight gain
Miscellaneous:	14[c]	11	

[a] Patients first rated each of 22 specific symptoms (2 "other" unspecified symptoms could also be nominated and rated), then listed "the 3 symptoms that you noticed most in the last week", which we refer to as the "Top 3". Parentheses denote two symptoms which are likely to be long-term side-effects of previous course of chemotherapy.

[b] Each symptom was rated "at its worst in the past week" on an 11-point scale, where 0 = "did not have the symptom" to 10 = "as bad as I can imagine".

[c] 14 patients listed 15 other symptoms: leg cramping/cramping (unspecified), alopecia, cough, thirst, sexual problems, high pressure originating in my chest then neck, face sweeling, feet blisters, swollen leg, ringing in ear, fever

Table 2 Summary of the extent to which the existing candidate patient-reported outcome measures and the new questionnaire (MOST) met the optimality criteria for efficient and focused measurement of symptom benefit for clinical trial endpoint analysis.

Optimality Criteria	QLQ-C30/OV28	FACT-O	FOSI-8	FOSI-18	SRQ	MOST
1. content validity[1]	**Met:** all 10 necessary symptoms, plus 38 surplus items	**Partially met:** 8/10 necessary symptoms, plus 29 surplus items	**Partially met:** 6/10 necessary symptoms, and 2 surplus items	**Met:** all 10 necessary symptoms, and 8 surplus items	**Met:** all 10 necessary symptoms, plus 12 surplus items	**Met:** all 10 necessary symptoms, plus 20 surplus items
2. Recall period of 3-4 weeks[2]	**Not met:** recall of 1 week	**Not met:** recall of 1 week	**Not met:** recall of 1 week	**Not met:** recall of 1 week	**Not met:** recall of 1 week	**Met**
3. Numerical rating scales for item responses	**Not met:** 4-point verbal rating scale[3a]	**Not met:** 5-point verbal rating scale[3b]	**Not met:** 5-point verbal rating scale	**Not met:** 5-point verbal rating scale	**Met:** 0-10 numeric rating scale[3c]	**Met:** 0-10 numeric rating scale[3d]
4. Symptom index scoring: all necessary symptoms aggregated into a single summary score	**Not met:** 20 items covering 10 necessary symptoms separated into mutually independent 10 summary scales	**Partially met:** Total score includes10 items covering 8/10 necessary symptoms but also includes 29 surplus items	**Partially met:** total score includes 6/10 necessary symptoms plus 2 surplus items	**Partially met:** Total score includes all 10 necessary symptoms plus 8 surplus items, including some treatment side-effects and 3 items about function and wellbeing	**Partially met:** Total score includes all 10 necessary symptoms plus 12 surplus items	**Met:** psychometric properties to be assessed in Stage 2 of GCIG-SBS[4], along with the Treatment concerns index, containing 17 of the 20 surplus items

1. Content validity was defined as symptoms including all 10 symptoms necessary for the target population (women with symptomatic ovarian cancer) and purpose (to provide efficient and targeted measurement of symptom benefit for use in clinical trial endpoint analysis). The guiding principle was "everything that matters and no more", so items that were not necessary to this purpose were considered surplus to it.
2. Recall period of 3-4 weeks, matched to the typical length of a cycle of palliative chemotherapy for recurrent ovarian cancer.
3. 3a. All but two items are rated on a 4-point Likert scale: 1 = "not at all", 2 = "a little", 3="quite a bit", 4="very much". The two exceptions are the global QOL/health items, rated on a 7-point scale, anchored at 1 "very poor" and 7 = "excellent".

3b. All items are rated on a 5-point verbal rating scale: 0 = "not at all", 1 = "a little bit", 2 = "somewhat", 3="quite a bit", 4="very much"

3c. All items are rated on an 11-point numeric rating scale, from 0 = "did not have the symptom" to 10 = "as bad as I can imagine".

3d. All symptoms and treatment concerns rated on an11-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". The three wellbeing items range from 0 = "worst possible" to 10 = "best possible".

4. GCIG-SBS: Gynecologic Cancer Intergroup Symptom Benefit Study

Table 3 Number of questions in the QLQ-C30, QLQ-OV28, FACT-O and FOSI (8 and 18 item versions) that address the 10 symptoms that are necessary to measure subjective symptom benefit in clinical trials of palliative chemotherapy for ovarian cancer, and number that address other issues (surplus items).

	QLQ-C30	QLQ-OV28	EORTC C30+OV28	FACT-O	FOSI (8)	NCCN-FACT FOSI-18	SRQ
The 10 necessary symptoms:							
1. Pain - general	2	-	2	1	1	1	1
2. Abdominal bloating/discomfort	-	4	4	2	2	2	1
3. Fatigue	3	-	3	1	1	2	1
4. Sleep disturbance	1	-	1	1	-	1	1
5. Bowel problems	2	2	4	1	-	2	1
6. Nausea and vomiting	2	-	2	2	2	2	1
7. Shortness of breath	1	-	1	-	-	-	1
8. Appetite	1	1	2	1	-	-	1
9. Urinary problems	-	1	1	-	-	-	1
10. Weight changes	-	-	-	1	-	-	1
Number of items addressing necessary symptoms[a]	12	8	20	10	6	10	10
Number of surplus items[a]	18	20	38	29	2	8	12
Number of necessary symptoms missed	2	6	0	2	6	4	0

a. “Necessary items” are those that relate to the 10 symptoms that are necessary and sufficient to measure subjective symptom benefit in clinical trials of palliative chemotherapy for ovarian cancer. “Surplus items” are relate to other symptoms or issues.

Table 4 Number of summary scales in the QLQ-C30 and QLQ-OV28, and the number of questions in each that address the 10 symptoms necessary to measure subjective symptom benefit in clinical trials of palliative chemotherapy for ovarian cancer.

Type	Questionnaire	Summary scale, as specified by the EORTC standard scoring algorithms	Number of questions (items)	
			Necessary and sufficient symptoms[a]	Surplus symptoms / issues
Symptoms	QLQ-C30	Pain	2	
		Fatigue	3	
		Nausea &vomiting	2	
		Diarrhea	1	
		Constipation	1	
		Dyspnea	1	
		Appetite	1	
		Sleep	1	
		Financial concerns		1
	QLQ-OV28	Abdominal/GI	7	
		Peripheral neuropathy		3
		Hormonal/menopausal		2
		Other chemo side-effects	1	6
Functioning	QLQ-C30	Physical functioning		5
		Role functioning		2
		Social functioning		2
		Emotional functioning		4
		Cognitive functioning		2
	QLQ-OV28	Body image		2
		Sexuality		4
		Attiude to disease/treatment		3
Global	QLQ-C30	Global quality of life/health		2
Total items			20	38
Total scales			10	13

Table 5 Number of multi-item scales in the FACT-G, FACT-O and FOSI (8 and 18 item versions), and the number of items in each that address the 10 symptoms necessary and sufficient to measure subjective symptom benefit in clinical trials of palliative chemotherapy for ovarian cancer.

Questionnaire	Summary scales, as specified by the FACIT standard scoring algorithms	Number of items	
		Necessary and sufficient symptoms	Surplus symptoms / issues
FACT-G	Physical well-being (PWB)	3	4
	Functional well-being (FWB)	1	6
	Emotional well-being (EWB)	-	6
	Social/family well-being (SWB)	-	7
	FACT-G total score (TOT) = PWB+FWB+EWB+SWB	4	23
FACT-O	Additional (ovarian) concerns (AC)	6	6[a]
	FACT-O total score = TOT+AC	10	29
	Trial outcome index (TOI) = PWB+FWB+AC	10	16
FOSI-8	FOSI-8	6	2[b]
NFOSI-18	NFOSI-8	10	8[c]
	Disease-related symptoms – Physical (DRS-P)	8	1
	Disease-related symptoms – Emotional (DRS-E)	-	1
	Treatment side effects (TSE)	2	3
	Function and wellbeing (FWB)	-	3

a. The six other issues are: "I am bothered by hair loss", "I like the appearance of my body", "I am able to get around by myself", I am able to feel like a woman", "I am interested in sex", "I have concerns about my ability to have children".
b. The two other issues are: "I worry my condition will get worse" and "I am content with the quality of my life right now".
c. The eight other issues are: the two additional FOSI-8 questions plus "I am bothered by skin problems", "I am able to get around by myself", "I am able to enjoy life", "I feel ill", "I am bothered by hair loss" and "I am bothered by side effects of treatment".

Symptom Benefit (ANZGOG0701) Study No. 0 7 0 1 _ _ _ / _ _ Visit No.

examples: 0701005, 0701034

01 - Baseline
02 - Visit 2
03, 04, 05 etc. for all subsequent visits.

Today's Date _ _ / _ _ / _ _ _ _
DD MM YYYY

Measure of Ovarian Cancer Symptoms & Treatment Concerns - Recent

Please **circle** one number for each line to best show **how much that aspect troubled you** on average **during the last 3-4 weeks.**

		No trouble at all		Mild			Moderate			Severe		Worst I can imagine
1.	Pain (all and anywhere)	0	1	2	3	4	5	6	7	8	9	10
2.	Fatigue (tiredness)	0	1	2	3	4	5	6	7	8	9	10
3.	Poor appetite (or feeling full quickly)	0	1	2	3	4	5	6	7	8	9	10
4.	Abdominal pain, discomfort and/or cramps	0	1	2	3	4	5	6	7	8	9	10
5.	Abdominal swelling, bloating and/or fullness	0	1	2	3	4	5	6	7	8	9	10
6.	Trouble eating	0	1	2	3	4	5	6	7	8	9	10
7.	Indigestion	0	1	2	3	4	5	6	7	8	9	10
8.	Nausea	0	1	2	3	4	5	6	7	8	9	10
9.	Vomiting	0	1	2	3	4	5	6	7	8	9	10
10.	Diarrhoea	0	1	2	3	4	5	6	7	8	9	10
11.	Constipation	0	1	2	3	4	5	6	7	8	9	10
12.	Bladder problems	0	1	2	3	4	5	6	7	8	9	10
13.	Shortness of breath	0	1	2	3	4	5	6	7	8	9	10
14.	Leg swelling	0	1	2	3	4	5	6	7	8	9	10
15.	Trouble sleeping	0	1	2	3	4	5	6	7	8	9	10

Please **circle** one number for each line to show how you would have rated yourself on that aspect on average **during the last 3-4 weeks.**

		Best possible	Very Good		Good		Fair		Poor	Very poor	Worst possible
16.	Physical well-being	10	9	8	7	6	5	4	3	2	1	0
17.	Emotional well-being	10	9	8	7	6	5	4	3	2	1	0
18.	Overall well-being	10	9	8	7	6	5	4	3	2	1	0

20402

Symptom Benefit (ANZGOG0701) Study No. 0 7 0 1 _ _ _ / _ _ Visit No.

examples: 0701005, 0701034

01 - Baseline
02 - Visit 2
03, 04, 05 etc. for all subsequent visits.

Measure of Ovarian Cancer Symptoms & Treatment Concerns - Recent

Please **circle** one number for each line to best show **how much that aspect troubled you** on average **during the last month.**

		No trouble at all		Mild			Moderate			Severe		Worst I can imagine
19.	Altered sense of taste	0	1	2	3	4	5	6	7	8	9	10
20.	Sore mouth or throat	0	1	2	3	4	5	6	7	8	9	10
21.	Difficulty swallowing	0	1	2	3	4	5	6	7	8	9	10
22.	Loss of appetite	0	1	2	3	4	5	6	7	8	9	10
23.	Hair loss	0	1	2	3	4	5	6	7	8	9	10
24.	Skin rash	0	1	2	3	4	5	6	7	8	9	10
25.	Numbness or pins and needles	0	1	2	3	4	5	6	7	8	9	10
26.	Sore hands and feet	0	1	2	3	4	5	6	7	8	9	10
27.	Problems taking tablets	0	1	2	3	4	5	6	7	8	9	10
28.	Problems with needles or injections	0	1	2	3	4	5	6	7	8	9	10
29.	Inconvenience of treatment	0	1	2	3	4	5	6	7	8	9	10
30.	Thought of actually having treatment	0	1	2	3	4	5	6	7	8	9	10
31.	Trouble concentrating	0	1	2	3	4	5	6	7	8	9	10
32.	Anxiety (feeling worried)	0	1	2	3	4	5	6	7	8	9	10
33.	Depression (feeling sad)	0	1	2	3	4	5	6	7	8	9	10
34.	Problems doing what I wanted	0	1	2	3	4	5	6	7	8	9	10
35.	Problems for my family or friends	0	1	2	3	4	5	6	7	8	9	10