Food and Drug Administration
CENTER FOR DRUG EVALUATION AND RESEARCH
Office of Oncology Drug Products

Briefing Document

Oncologic Drugs Advisory Committee Meeting
March 13, 2006
Gaithersburg Hilton, Gaithersburg, MD

NDA 20-509/S-039
Gemzar (gemcitabine HCL)
Eli Lilly & Company

Table of Contents

FDA ODAC Briefing Document for Gemzar
March 13, 2006

NDA 20509/S039

DRUG Gemzar (gemcitabine HCL)

APPLICANT Eli Lilly

DATE RECEIVED June 17, 2005

PROPOSED INDICATION Gemzar in combination with carboplatin is indicated for the treatment of patients with advanced ovarian cancer that has relapsed at least 6 months after completion of platinum-based therapy.

EXECUTIVE SUMMARY

Gemzar was studied in a randomized Phase 3 study of 356 patients with advanced ovarian cancer that had relapsed at least 6 months after first-line platinum-based therapy. Patients were randomized 1:1 to receive either Gemzar in combination with carboplatin or carboplatin alone. The Gemzar/carboplatin combination adds 2.8 months to median progression-free survival (PFS) with no apparent effect on survival at a cost of increased toxicity, mainly anemia, neutropenia and thrombocytopenia, requiring increased RBC and platelet transfusions and increased use of granulocyte stimulating factors and erythropoietic agents. Independently assessed tumor response rates were Gemzar/carboplatin 46.3% and carboplatin alone 35.6%. This trial was conducted entirely outside of the United States and the FDA had no input into its design or conduct.

The main issue is whether adding 2.8 months to median PFS at a cost of additional toxicity with no apparent effect on survival is a sufficient basis for Gemzar approval for this use. Important considerations are that the combination of paclitaxel and carboplatin has been shown to prolong survival in this setting. In addition, a large international gynecologic group (including the NCI, GOG, RTOG and NCIC) at a Consensus Conference on Ovarian Cancer in 2004 indicated that in the setting of second-line chemotherapy for advanced ovarian cancer "progression-free survival does not seem to be a good surrogate for survival". "Progression-free survival data remain of interest but are unlikely to be sufficiently persuasive to shift practice patterns". See DISCUSSION below for details.

CLINICAL EVALUATION

This SNDA is supported by a single RCT conducted by the Arbeitsgemeinschaft Gynaekologische Onkologie, Studiengruppe (AGO) and by some data from Phase 2 studies.

Study Design and Treatment Regimens

Gemzar was studied in a randomized Phase 3 study of 356 patients with advanced ovarian cancer that had relapsed at least 6 months after first-line platinum-based therapy. Patients were randomized 1:1 to receive either

Test regimen: Gemzar 1000 mg/m^2 on Days 1 and 8 of a 21-day cycle and carboplatin AUC 4 administered after Gemzar on Day 1 of each 21 day cycle

Control: single-agent carboplatin AUC 5 administered on Day 1 of each 21-day cycle

Patients were stratified prior to randomization by progression-free time, primary platinum therapy and bidimensionally measurable disease. See Table 1 for details.

Study Endpoints

Primary The primary endpoint was progression-free survival (PFS).

Secondary Response rate.
Response duration.
Survival time.
Toxicity.
Changes in quality of life (QoL), measured using the European Organization for the Research and Treatment of Cancer (EORTC) QLQC30 and OV28 QoL instruments.

Study Results

Table 1

Distribution by stratification factors

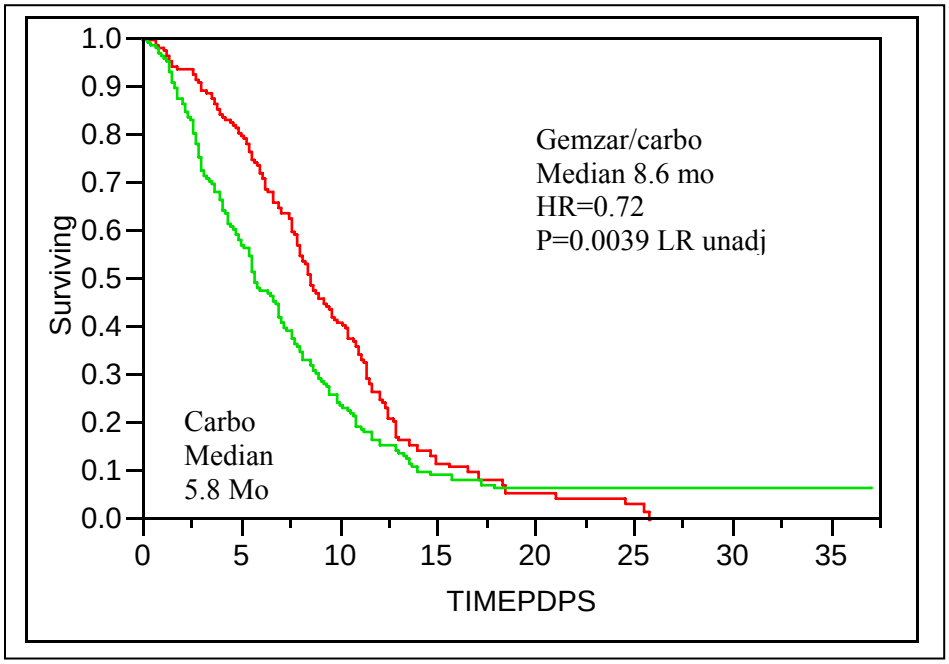
Stratification Factor		Number (%) of Patients	
		GCb Arm (N=178)	Cb Arm (N=178)
Progression-free time	<6 months	1 (0.5)	0
	6 to 12 months	71 (39.9)	71 (39.9)
	>12 months	106 (59.6)	107 (60.1)
Primary platinum therapy	Platinum plus paclitaxel	122 (68.5)	120 (67.4)
	Other platinum-containing therapies	56 (31.5)	58 (32.6)
Bidimensionally measurable disease	Yes	163 (91.6)	170 (95.5)
	No (evaluable disease only)	14 (7.9)	5 (2.8)
	No measurable or evaluable disease	1 (0.6)	3 (1.7)

Progression-Free Survival

Progression-Free Survival (PFS) was the primary efficacy endpoint. The protocol specified primary statistical analysis was unadjusted. Time to progression was defined as the time from the date of randomization to the date of progression or death from any cause. CA-125 was not a basis for progression in this study. The protocol criteria for progression were "50% increase or an increase of 10 cm2 (whichever is smaller) in the sum of products of all measurable lesions over smallest sum observed (over baseline if no decrease) using the same techniques as baseline, OR clear worsening of any evaluable disease, OR reappearance of any lesion which had disappeared, OR appearance of any new lesion/site, OR failure to return for evaluation due to death or deteriorating condition (unless clearly unrelated to this cancer). For 'scan-only' bone disease, increased uptake does not constitute clear worsening. Worsening of existing nonevaluable disease does not constitute progression.

The results are shown in Figure 1. Median PFS time was 8.6 months for the Gemzar/carboplatin group and 5.8 months for the carboplatin alone group, HR=0.72, p=0.0039, unadjusted LR, 2-sided. Eighty-seven per cent of patients had an event and only 13% were censored.

Figure 1 Progression-Free Survival (unadjusted)



Survival

The survival analysis results are shown in Figure 2. The modest 2.8 month effect on median PFS is not reflected in any survival effect. PFS does not appear to be a surrogate for survival in this study. Median survival time was 17.97 months in the Gemzar/carboplatin group and 17.31 months in the carboplatin alone group, HR= 0.985, p= 0.898, unadjusted LR, 2-sided. When adjusted for the 3 prerandomization stratification factors using stratified Cox regression, the HR=0.9, p=0.41.

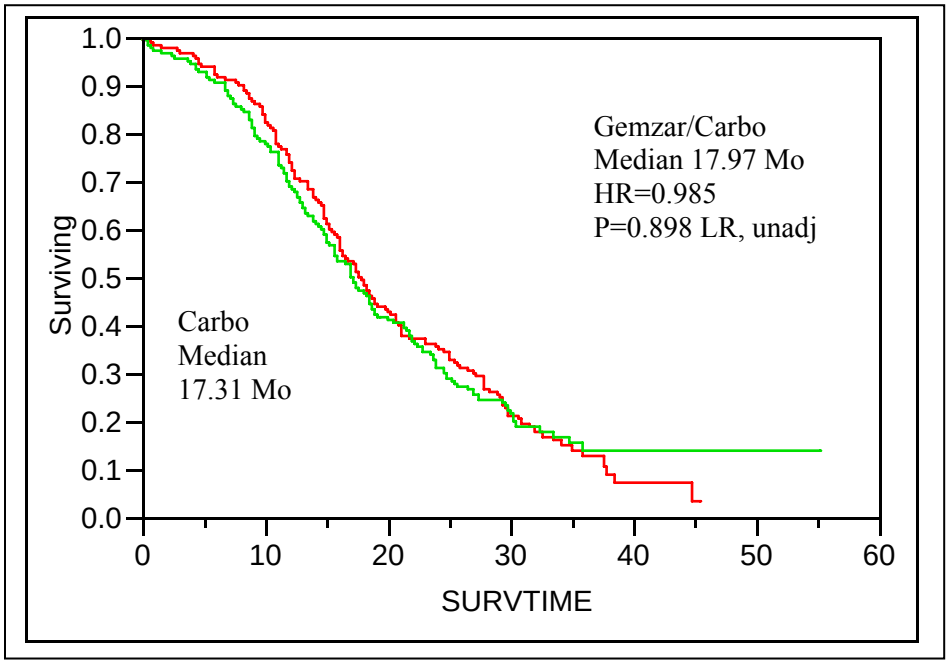
In the Final Protocol Addendum approved on January 28, 2002, the Applicant pointed out that survival time would be compared between regimens using the log-rank test, without further details. Later, the Applicant developed two Statistical Analysis Plans. The first was approved by the Applicant on February 5, 2003 for the cut-off date of February 28, 2003 and the second was approved by the Applicant on February 7, 2005 for the cut-off date of February 15, 2005. The second

SAP was submitted to the FDA in the briefing document for the March 23, 2005 Pre-NDA meeting (February 19, 2005: IND 29,653, SN 1122). In these SAPs the following additional analyses were added for the overall survival time.

The SAPs indicated that the Cox proportional hazards model will be used to study the adjusted treatment effect on overall survival. Covariates used in the Cox model will include age, Eastern Cooperative Oncology Group (ECOG) performance status, prior platinum therapy, total size of tumor, disease status, duration of platinum-free interval, post-study chemotherapy, and study therapy. Each factor will be assessed individually for prognostic value ($p<0.05$). Factors that are deemed to have prognostic value will be included in a multiple regression analysis to assess their significance in the presence of the other factors. Backward elimination will be used to identify the final set of prognostic factors. Treatment will then be added to this final model to assess the effect of treatment when adjusted for other significant factors in a multiple regression model.

The FDA considers such an analysis only exploratory and no conclusions can be drawn from it. To be considered otherwise by the FDA, the analysis would need to be pre-specified in the protocol as the primary analysis and the covariates would also need to be pre-specified. Neither is the case. The statistical analysis plan did not designate this as the primary survival analysis and the plan specified several covariates from which the covariates used in the Cox regression analysis were selected.

Figure 2 Survival By Treatment ITT (Unadjusted)



Post Study Chemotherapy

Post study chemotherapy was administered to 73 of 178 (41%) patients in each treatment group. In the carboplatin alone group 13 (7%) of patients received Gemzar after progression. It does not appear that differences in post study chemotherapy account for the failure to demonstrate a Gemzar survival effect.

Objective Tumor Response

Investigator assessed objective tumor response using SWOG criteria is shown in Table 2.

Table 2
Summary of Best Tumor Response
Investigator-Assessed

	Number (%) of Patients		
Best Tumor Response	**GCb Arm (N=178)**	**Cb Arm (N=178)**	**P-Value**
Complete response	26 (14.6%)	11 (6.2%)	
Partial response	54 (30.3%)	43 (24.2%)	
PRNM	4 (2.2%)	1 (0.6%)	
Total responders	84 (47.2%)	55 (30.9%)	0.0016 Chi Sq
Response Duration	HR=0.81		0.2511 LR

Abbreviations: Cb = carboplatin monotherapy; CI = confidence interval; CR = complete response; GCb = gemcitabine plus carboplatin combination therapy; N = total population size; PR = partial response; PRNM = partial response, nonmeasurable disease.

Objective tumor response was also evaluated by blinded independent reviewers. Results are shown in Applicant Table 3. All patients who had available radiologic scans from baseline and post-baseline were included in the independent radiologic review process. Patients for whom only baseline scans were available and those who were assessed only by ultrasound scans or physical examination were omitted from the independent review process.

Table 3
Summary of Best Tumor Response
Independently Assessed

Best Tumor Response	Number (%) of Patients		
	GCb Arm (N=121)	Cb Arm (N=101)	P Value Chi Square
Complete response	11 (9.1%)	4 (4.0%)	
Partial response	45 (37.2%)	32 (31.7%)	
Total responders [CR + PR + PRNM]	56 (46.3%)	36 (35.6%)	0.1091

Abbreviations: Cb = carboplatin monotherapy; CI = confidence interval; CR = complete response; GCb = gemcitabine plus carboplatin combination therapy; N = total population size; PR = partial response; PRNM = partial response, nonmeasurable disease.

Quality of Life

Health-related quality of life (HRQL) was assessed using the EORTC QLQ-C30 and the EORTC QLQ-OV28 patient reported outcome (PRO) questionnaires. These HRQL assessments can not be used as the basis of Gemzar approval because the study was not blinded, the effect of concurrent medications was not assessed, on some items the carboplatin alone group did better and the effect on "global quality of life," although statistically significant, is not clinically meaningful.

Extent of Drug Exposure

The median number of completed cycles was 6 for both treatment arms. Among the 175 treated patients on the GCb Arm a total of 961 cycles of therapy were completed. Among the 174 treated patients on the Cb Arm a total of 888 cycles were completed.

The per cent of planned dose actually administered is shown in the following Applicant Table 4.

Table 4

Treatment Dose Intensity

Treatment Arm Study Drug	Number of Patients Treated	Dose Intensity		
		Planned Mean Dose per Patient per Week	Actual Mean Dose per Patient per Week	Percent of Planned Mean Dose (Actual/Planned)
GCb Arm				
Gemcitabine Carboplatin	175 175	666.7 mg/m2 AUC 1.33	504.2 mg/m2 AUC 1.28	75.6% 96.2%
Cb Arm				
Carboplatin	174	AUC 1.67	AUC 1.64	98.2%

Safety

The following Applicant Table 5 shows the adverse events occurring in at least 10 % of patients on at least one arm. The Gemzar/Carboplatin group had a higher incidence of anemia, neutropenia and thrombocytopenia, required more RBC and platelet transfusions and required more granulocyte growth factors and erythropoietic agents.

Table 5

Adverse Events ≥10% of Patients[a]

	Gemzar plus Carboplatin N=175			Carboplatin N=174		
	% All Grades	% Grade 3	% Grade 4	% All Grades	% Grade 3	% Grade 4
Laboratory[b]						
Hematologic						
Anemia	86	22	6	75	9	2
RBC Transfusion[c]	38			15		
Neutropenia	90	42	29	58	11	1
Leukopenia	86	48	5	70	6	<1
Thrombocytopenia	78	30	5	57	10	1
Platelet Transfusion[c]	9			3		
Non-laboratory[d]						
Alopecia	49	0	0	18	0	0
Neuropathy-sensory	30	1	0	27	2	0
Nausea	69	4	0	61	2	0
Fatigue	39	2	<1	29	2	0
Vomiting	42	3	0	33	1	<1
Diarrhea	15	2	0	8	0	0
Anorexia	16	<1	0	8	0	0
Stomatitis/pharyngitis	21	<1	0	12	0	0
Constipation	30	5	0	24	2	0

[a] Grade based on Common Toxicity Criteria (CTC) Version 2.0 (all grades ≥10%).
[b] Regardless of causality.
[c] Percent of patients receiving transfusions. Transfusions are not CTC-graded events. Blood transfusions included both packed red blood cells and whole blood.
[d] Non-laboratory events were graded only if assessed to be possibly drug-related.

Colony stimulating factors were administered more frequently with combination therapy than with monotherapy (granulocyte growth factors: 23.6% and 10.1%, respectively; erythropoietic agents: 7.3% and 3.9%, respectively).

Table 6

Hospitalizations

	GCb Arm (N=178)	Cb Arm (N=178)
Hospitalizations		
Febrile neutropenia	2 (1.1%)	0
Other drug-related adverse events	27 (15.2%)	16 (9.0%)
Number of days of hospitalization		
Febrile neutropenia	11	0
Other drug-related adverse events	150	108

Dose Reductions and Discontinuations

In the Gemzar plus carboplatin versus carboplatin alone study, dose reductions occurred with 10.4% of Gemzar injections and 1.8% of carboplatin injections on the combination arm, versus 3.8% on the carboplatin alone arm.

On the combination arm, 13.7% of Gemzar doses were omitted and 0.2% of carboplatin doses were omitted, compared to 0% of carboplatin doses on the carboplatin alone arm.

DISCUSSION

The main issue for this SNDA is whether a 2.8 month improvement in median PFS at a cost of additional toxicity that is not reflected in an improvement in overall survival is an adequate basis for drug approval for treatment of patients with advanced ovarian cancer who have relapsed at least 6 months after completion of platinum-based therapy.

First, there are chemotherapy regimens that have been shown in RCTs to prolong survival in this setting. The ICON4/AGO-OVAR-2.2 trial randomized 802 women with platinum sensitive relapsed ovarian cancer to conventional platinum based chemotherapy (mostly single agent carboplatin) and paclitaxel plus platinum based chemotherapy. This is a combined analysis of data from three parallel protocols. Median survival was 29 months in the paclitaxel plus platinum group compared to 24 months in the conventional platinum group with HR 0.82, p=0.02, stratified LR. (1)

It is argued that the results of this trial might not be applicable in the USA because only 43% of trial patients had prior taxane treatment while this percentage is higher in the USA. However, the survival effect in this trial was similar in patients with and without prior taxane treatment.

It is also argued that residual taxane neurotoxicity may prevent some patients form receiving the paclitaxel/carboplatin combination and these patients might benefit from the Gemzar/carboplatin combination. However, a study of the Gemzar/carboplatin combination has not been conducted in this patient group.

In addition, a randomized Phase 3 trial in 474 patients with relapsed ovarian cancer compared Pegylated Liposomal Doxorubicin (PLD) to Topotecan. Median survival times were 60 weeks and 56.7 weeks, respectively for PLD and Topotecan, HR not reported, p=0.34, stratified LR. Patients were stratified prior to randomization by platinum sensitivity (platinum-free interval ≤ 6 months or > 6 months). In the subgroup of platinum sensitive patients median survival times were 108 weeks and 71 weeks respectively for PLD and Topotecan, HR not reported, p=0.008, stratified LR. (2)

Second, The Gemzar RCT did not show a survival effect. The lack of a survival effect is not explained by crossover of carboplatin patients to Gemzar after progression because only 13 patients crossed over. The Applicant states that the trial was not adequately powered to detect a survival effect. There were 283 deaths in the 356 patient RCT. To have had 80% power to detect a 30% survival effect about 460 events would be required.

If further accrual were allowed so that a final analysis would occur after an additional 177 events (total of 460 events), an observed hazard ratio estimate for the next 177 events of 0.64 (0.74) from a non-stratified (stratified) Cox regression model would be needed to get a statistically significant result for the analysis based on 460 events. If this additional period were powered at a treatment versus control hazard ratio of 0.77 (0.77=1/1.3, meaning we are assuming a 30% Gemzar survival effect), then the probability for detecting an overall survival difference between the two therapies after 460 events would be 0.10 (0.38) using the non-stratified (stratified) Cox regression analysis. This may represent a more favorable scenario than the actual case, since the hypothesized non-stratified HR of 0.77 is outside the current 95% confidence interval for the HR.

Using another approach, if we use the overall survival results based on the non-stratified (stratified) Cox regression analysis of the first 283 events to establish a prior distribution for the true hazard ratio for the next 177 events, then the probability of detecting a difference between the two therapies is 0.012 (0.15). Therefore if this trial were extended to allow a survival analysis at 460 events, it is very improbable that Gemzar would demonstrate a favorable effect on overall survival

Third, we note that the Applicant has performed a Cox multiple regression analysis of survival. Such an analysis is considered only exploratory by the FDA and no conclusions can be drawn from it. To be considered otherwise by the FDA, the analysis would need to be pre-specified in the protocol as the primary analysis and the covariates would also need to be prespecified. Neither is the case. The statistical analysis plan specified several covariates from which the covariates used in the Cox regression analysis were selected (See Survival section above for details).

Fourth, the Gynecologic Intergroup (GCIG) and its member organizations GOG (USA), RTOG (USA), NCI-US (USA), NCIC-CTG (Canada), [AGO-OVAR (Germany), ANZGOG (Australia – New Zealand), EORTC (Europe), GEICO (Spain), GINECO (France), , JGOG (Japan), MRC/NCRI (UK), NSGO (Scandinavia), SGCTG (Scotland)] make recommendations regarding endpoints for second-line chemotherapy Phase 3 studies in advanced ovarian cancer. (3, 4) The vote on this recommendation was **unanimous**.

"For phase III trials in the second-line setting, progression-free survival does not seem to be a good surrogate for survival: there are several examples where progression-free survival was significantly improved, with no survival impact. It can be argued that some of these studies were underpowered to detect survival improvements; however, the weight of evidence to consider progression-free survival a surrogate for survival, and thus a primary end point in the second-line setting, is not strong as yet. In the recurrent disease setting, overall survival remains an important primary end point (particularly if more costly or toxic therapy is being offered). Progression-free survival data remain of interest but are unlikely to be sufficiently persuasive to shift practice patterns. Furthermore, since the rationale for treating patients with relapsed disease is a desire to

improve symptoms and thus quality of life, an adequate measure of these factors would also be an appropriate primary end point for randomized trials. However, no universally acknowledged and standardized system of symptom measurement and analysis is readily available. GCIG will continue, through its working groups, to build a consensus on how meaningful improvements in disease-related symptoms can be quantified." (3)

Fifth, the Gemzar RCT was conducted entirely outside of the United States and the FDA had no input into its design or conduct.

CONCLUSION

The Gemzar/carboplatin combination adds 2.8 months to median PFS with no apparent effect on survival at a cost of increased toxicity, mainly anemia, neutropenia and thrombocytopenia, requiring increased RBC and platelet transfusions and increased use of granulocyte stimulating factors and erythrpoietic agents.

RECOMMENDATION

Deferred pending advice of the ODAC.

REFERENCES

1. ICON and AGO collaborators:
Writing committee—M K B Parmar, J A Ledermann, N Colombo, A du Bois, J-F Delaloye, G B Kristensen, S Wheeler, A M Swart, W Qian, V Torri, I Floriani, G Jayson, A Lamont, C Tropé.
Paclitaxel plus platinum-based chemotherapy versus conventional platinum-based chemotherapy in women with relapsed ovarian cancer: the ICON4/AGO-OVAR-2.2 trial. Lancet 2003; 361: 2099–2106.

2. Gordon, Alan N.; Fleagle, John T.; Guthrie, David; Parkin, David E.; Gore, Martin E.; Lacave, Angel J. Recurrent Epithelial Ovarian Carcinoma: A Randomized Phase III Study of Pegylated Liposomal Doxorubicin Versus Topotecan. Journal of Clinical Oncology Volume 19(14), 15 July 2001, pp 3312-3322.

3. J. B. Vermorken[1], M. K. B. Parmar[2], M. F. Brady[3], E. A. Eisenhauer[4], T. Hogberg[5], R. F. Ozols[3], J. Rochon[6], G. J. S. Rustin[2], S. Sagae[7] & R. H. M. Verheijen. Gynecologic Cancer Intergroup (GCIG) and its member organizations: 1EORTC (Europe); 2MRC/NCRI (UK); 3GOG (USA); 4NCIC-CTG (Canada); 5NSGO (Scandinavia); 6AGO-OVAR (Germany); 7JGOG (Japan)]. Clinical trials in ovarian carcinoma: study methodology. Annals of Oncology 16 (Supplement 8) viii20–viii29, 2005

4. A. du Bois[1], M. Quinn[2], T. Thigpen[3], J. Vermorken[4], E. Avall-Lundqvist[5], M. Bookman[3], D. Bowtell[2], M. Brady[3], A. Casado[4], A. Cervantes[6], E. Eisenhauer[7], M. Friedlaender[2], K. Fujiwara[8], S. Grenman[5], J. P. Guastalla[9], P. Harper[10], T. Hogberg[5], S. Kaye[11], H. Kitchener[10], G. Kristensen[5], R. Mannel[3], W. Meier[1], B. Miller[12], J. P. Neijt[13], A. Oza[7], R. Ozols[3], M. Parmar[10], S. Pecorelli[14], J. Pfisterer[1], A. Poveda[6], D.Provencher[7], E. Pujade-Lauraine[9], M. Randall[3], J. Rochon[1], G. Rustin[10], S. Sagae[8], F. Stehman[3], G. Stuart[7], E. Trimble[15], P. Vasey[11], I. Vergote[4], R. Verheijen[4] & U. Wagner[1].
Gynecological Cancer Intergroup (GCIG) and its member organizations: 1AGO-OVAR (Germany), 2ANZGOG (Australia – New Zealand), 4EORTC (Europe), 6GEICO (Spain), 9GINECO (France), 3GOG (USA), 8JGOG (Japan), 10MRC/NCRI (UK), 7NCIC-CTG (Canada), 15NCI-US (USA), 5NSGO (Scandinavia), 12RTOG (USA), 11SGCTG (Scotland), and with representation of 14 IGCS and 13th organizational team of the two prior International OCCC. 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 16 (Supplement 8) viii7–viii12, 2005

Annals of Oncology 16 (Supplement 8) viii7–viii12, 2005
doi:10.1093/annonc/mdi961

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)

A. du Bois[1]*, M. Quinn[2], T. Thigpen[3], J. Vermorken[4], E. Avall-Lundqvist[5], M. Bookman[3], D. Bowtell[2], M. Brady[3], A. Casado[4], A. Cervantes[6], E. Eisenhauer[7], M. Friedlaender[2], K. Fujiwara[8], S. Grenman[5], J. P. Guastalla[9], P. Harper[10], T. Hogberg[5], S. Kaye[11], H. Kitchener[10], G. Kristensen[5], R. Mannel[3], W. Meier[1], B. Miller[12], J. P. Neijt[13], A. Oza[7], R. Ozols[3], M. Parmar[10], S. Pecorelli[14], J. Pfisterer[1], A. Poveda[6], D. Provencher[7], E. Pujade-Lauraine[9], M. Randall[3], J. Rochon[1], G. Rustin[10], S. Sagae[8], F. Stehman[3], G. Stuart[7], E. Trimble[15], P. Vasey[11], I. Vergote[4], R. Verheijen[4] & U. Wagner[1]

Gynecological Cancer Intergroup (GCIG) and its member organizations: [1]AGO-OVAR (Germany), [2]ANZGOG (Australia – New Zealand), [4]EORTC (Europe), [6]GEICO (Spain), [9]GINECO (France), [3]GOG (USA), [8]JGOG (Japan), [10]MRC/NCRI (UK), [7]NCIC-CTG (Canada), [15]NCI-US (USA), [5]NSGO (Scandinavia), [12]RTOG (USA), [11]SGCTG (Scotland), and with representation of [14]IGCS and [13]the organizational team of the two prior International OCCC

International evidence-based consensus statements are important in order to define and update standards of care and to serve not only as guidelines to communities worldwide, but also to provide a rational basis for future research. Two previous successful international ovarian cancer consensus conferences (OCCC) were held in Elsinore, Denmark in 1993 [1] and 5 years later in Bergen aan Zee, The Netherlands [2], where consensus statements were developed on a number of issues including biological and prognostic factors, best current therapy, both surgical and medical, and directions for future research in advanced disease. Since then, international cooperation has become more extensive and intergroup studies are now common as a mechanism to conduct large randomized clinical trials. The Gynecological Cancer Intergroup (GCIG) now constitutes 13 national and international cooperative member groups and governmental organizations [3]. In 2002, the GCIG's general assembly voted to plan the 3rd International OCCC with a more formal process to achieve consensus among the study groups on methodology and standard requirements for clinical trials so as to guide other national and international study groups working in gynecologic oncology. It was anticipated that the OCCC statements would also guide the general medical community and support the pharmaceutical industry in developing appropriate strategies to improve the outcome of women with this disease.

Methodology

Planning committee and agenda

The GCIG is a cooperative organization (http://ctep.cancer.gov/resources/gcig) that includes representatives from four continents and most worldwide study groups performing trials in gynecologic oncology. The general assembly asked AGO-OVAR to be the host organization and formed a planning committee (PC) for the 3rd International OCCC including representatives of seven study groups from three continents: A.d.B. (chair PC, AGO-OVAR), J.P. (chair-elect GCIG, AGO-OVAR), M.Q. (ANZGOG), J.V. (past-chair GCIG, EORTC), T.T. (GOG), M.B. (GOG), M.P. (MRC/NCRI), G.S. (NCIC-CTG), E.A.-L. (chair GCIG, NSGO) and GCIG-secretary Monica Bacon (NCIC-CTG). The PC developed a proposal for the agenda which was approved by the GCIG assembly (first level of consensus) as including three core areas: (A) standard therapy and standard requirements for clinical trials in ovarian cancer; (B) study methodology; and (C) new treatment options and novel approaches. These three basic areas were covered by 12 questions considered as most relevant to direct future clinical and laboratory research via the GCIG's member study groups. Furthermore, GCIG agreed on the agenda, timetable and implementation of a semi-structured consensus process (see below).

The PC chair was delegated to develop funding plans, hire organizational help and select the venue of the meeting.

Selection of participants

The purpose of this meeting was not only to utilize the extensive expertise available through the GCIG, but also to develop a structured consensus process that would allow intellectual participation by all study groups worldwide, thereby ensuring that the eventual recommendations would have broad international acceptance. The GCIG covers four continents with membership of 13 national or international cooperative study groups and governmental/semi-governmental organizations. Each GCIG member organization was asked to provide a list of expert delegates who were regarded as being the most experienced and competent representatives. The number of delegates per organization varied from one to six and reflected the groups attributes (e.g. member institutions, population represented, history of completed

Correspondence to: Professor Dr A. du Bois, Department of Gynecology & Gynecologic Oncology, Dr Horst Schmidt Klinik (HSK), Ludwig-Erhard-Str. 100, D-65199 Wiesbaden, Germany.
E-mail: dubois.hsk-wiesbaden@uumail.de

© 2005 European Society for Medical Oncology

clinical trials). European groups sent proportionately fewer participants per group to ensure appropriate balance. Asia, however, remained underrepresented and no African group could be identified. In addition, one representative of the International Gynecological Cancer Society (IGCS) and the organizer of the two prior OCCC were invited to participate. Each participant was asked to agree on her/his tasks (see below). If one delegate refused to accept, the respective study group appointed a replacement. Overall, the invitation model resulted in an assembly of 52 experts.

The groups appointed their delegates to one of the three core areas, which were covered by working groups (groups A–C). Each group's delegates were divided into these working groups to guarantee diversity. Each working group was chaired by one responsible chairperson (underlined and *italic*) and two co-chairs (underlined): group A: *T.T. (GOG)*, A.d.B. (AGO-OVAR), G.S. (NCIC-CTG), M.F. (ANZGOG), S.K. (SGCTG), H.K. (MRC/NCRI), G.K. (NSGO), R.M. (GOG), A.P. (GEICO), F.S. (GOG), I.V. (EORTC), K.F. (JGOG), J.P.G. (GINECO), W.M. (AGO-OVAR), B.M. (RTOG), D.P. (NCIC-CTG); group B: *J.V. (EORTC)*, M.B. (GOG), M.P. (MRC/NCRI), E.E. (NCIC-CTG), R.O. (GOG), J.R. (AGO-OVAR), G.R. (MRC/NCRI), S.S. (JGOG), T.H. (NSGO); group C: *M.Q. (ANZGOG)*, E.A.-L. (NSGO), J.P. (AGO-OVAR), M.B. (GOG), D.B. (ANZGOG), S.G. (NSGO), P.H. (MRC/NCRI), E.P.-L. (GINECO), E.T. (NCI-US), P.V. (SGCTG), U.W. (AGO-OVAR), A.C. (EORTC), A.C. (GEICO), S.P. (IGCS), M.R. (GOG).

Gathering of evidence and structured consensus process

One presenter (p) and one discussant (d) were allocated to each of the 12 questions, making 24 p/d and nine chairpersons involved in the preparation of outlines for each question (four in group A, three in group B and five in group C). Again, the p and d for each question came from different groups. The p/d had several months to prepare two comprehensive outlines including all evidence they considered relevant to the appointed question. These outlines were discussed with the chairpersons of each working group and evidence not alluded to was included if appropriate. The modified outlines were then circulated among all members of the respective working group and discussed via e-mail by all participants and further modified before the conference (second level of consensus). The p/d prepared presentations for the conference and all materials were distributed prior to the meeting.

The first day of the 3-day conference was reserved for presentations of the outlines of presenters and discussants followed by a plenary discussion of each question. Through this discussion, working groups were able to gather additional views and evidence. At the end of each discussion a survey of the participants' opinions about the key points was collected to guide the working group activity. On the morning of the second day, working groups separated and discussed each of their questions resulting in agreed first drafts of answers (statements) (third level of consensus). That afternoon, each working group presented their drafts to the auditorium. Each statement was followed by an extensive discussion including comments from each group and suggestions for modifications were gathered for refinement. Again, working groups met separately and refined the statements including the suggestions from the general discussions. Additionally, working groups A and B met and discussed one overlapping issue. Finally, the second drafts of statements were provided that evening (fourth level of consensus).

The third day started with GCIG member study groups meeting separately to discuss their vote on each statement and to elect one voter per group. The consensus process included that each attendee had the opportunity to participate during discussion and in working group sessions, and that at least one member of each member study group commented on each question, but the final vote was limited to one vote per group. Each statement was read in the morning session and each study group commented on refinements required for approval (fifth level of consensus). All refinements per question were then voted on individually until a final statement was reached (majority vote; first level of acceptance). The final statements were considered at the final session and study groups were asked alphabetically whether they agreed or not (final level of acceptance). All 12 statements went through this structured consensus process and each study group commented and voted on each statement. A minority report would be included if one or more study group could not agree on a statement. The level of acceptance was reflected in the voting and is included in the section below.

Finally, the suggestions for a list of unmet needs and topics not included in this conference but important enough to be included in the next OCCC was completed [4]. Each working group had collected proposals for this list during the conference.

Results

12 questions and 12 statements: the 2004 consensus on ovarian cancer

The 12 questions were formally selected from a proposal by the PC and represent those questions regarded by the GCIG assembly as the most important with respect to current standards of care and future clinical trials. The questions and statements are printed in bold and outlined sequentially. The level of acceptance representing the final vote of the 13 member organizations is added to each question in italics. Further explanations were added after the conference and are not printed in bold. All these additions/explanations have been reviewed by all attendees of the GCIG OCCC 2004. Further details including the evidence on which the statements were based are outlined in the three working group documents published together with the statements [5–7].

1. A-1. Is there a need to strictly define the extent and type of surgery for patients in first-line trials?

- **Tissue should be obtained for histopathologic diagnosis to confirm the presence of primary ovarian or peritoneal carcinoma.**
- **Staging should be performed according to FIGO guidelines. For example, this includes at least lymph node sampling and peritoneal staging in early stage invasive disease (FIGO I–IIA).**
- **Up-front maximal surgical effort at cytoreduction with the goal of no residual disease should be undertaken.**
- **When cytoreductive surgery is not possible initially, it should be considered in patients who do not have progressive disease after three to five cycles of chemotherapy.**
- **Patients with ovarian cancer should have their surgery performed by an appropriately trained surgeon with experience in the management of ovarian cancer.**

Level of acceptance: 13/13 (i.e. 13 of 13 GCIG member organizations)

The first bullet point emphasizes that only patients for whom a histological diagnosis is available are included. The second bullet point focuses on the necessity of comprehensive staging, which is of utmost importance, especially in early ovarian cancer. This should include not only the above mentioned examples (lymph node and peritoneal staging) but all surgical procedures necessary to perform comprehensive FIGO staging (e.g. cytology, omentectomy, complete removal of the tumor,

hysterectomy and bilateral salpingo-oophorectomy in patients not suitable for fertility-sparing surgery).

The fourth bullet point focuses on patients who did not receive an appropriate and comprehensive effort at upfront cytoreduction by a trained surgeon (as outlined in third bullet) and who did not progress during chemotherapy. It is not meant that all patients who end up with bulky disease despite an appropriate surgical effort should receive interval debulking. If subsequent referral to a gynecologic oncology unit has taken place, a second surgical procedure prior to initiation of chemotherapy may be considered. The timing of interval debulking surgery should be flexible and the statement only reflects the current most common interval.

2. A-2. What is the impact of post-recurrence/progression treatment on the end points of first-line therapy? Do we need to standardize post-recurrence/progression therapy, or if not, how can we assess its impact on survival?

- **There is an impact of post-recurrence/progression therapy on overall survival (OS).**
- **It is not possible to standardize post-recurrence/progression therapy at the present time.**
- **Although OS is an important end point, progression-free survival (PFS) may be the preferred primary end point for trials assessing the impact of first-line therapy because of the confounding effect of the post-recurrence/progression therapy on OS. When PFS is the primary end point, measures should be taken to protect the validity of analysis of OS.**
- **There should be a clear definition of how to determine PFS.**

Level of acceptance: 13/13

The first bullet point refers to the recent observation that at least one large and two smaller trials have demonstrated a significant difference with respect to OS [5].

The third bullet point stresses the importance that the same schedule for follow-up has to be used in both arms and that the compliance with this schedule has to be assured.

3. A-3. Do we need a common 'GCIG recommended/accepted' standard arm for comparison with any new regimen/approach in first-line trials?

- **There should be a common 'GCIG recommended/accepted' standard arm for comparison with any new regimen/approach. Variations are allowed for clearly defined reasons.**

Level of acceptance: 13/13

4. A-4. Which regimen/kind of regimens can be regarded as standard comparator for future first-line trials?

- **Within a given trial the chemotherapy regimen should be standardized and consistent with respect to drugs, dose and schedule.**
- **The recommended standard comparator for trials of medical treatment in advanced ovarian cancer (FIGO IIB–IV) is carboplatin–paclitaxel.**
- **The recommended regimen is carboplatin with a dose of AUC 5–7.5 and paclitaxel 175 mg/m^2/3 h given every 3 weeks for six courses.**
- **The recommended standard in early stage (FIGO I–IIA) ovarian cancer patients in whom adjuvant chemotherapy is indicated should contain at least carboplatin AUC 5–7.5.**

Level of acceptance: 13/13

The first bullet point stresses that in future trials the regimens should be specified and consistent (e.g. not only 'platinum-based', but specifically which platinum and at what dose).

The last bullet point stresses that a carboplatin dose range is allowed, but that if doses above AUC 6 are considered, it should only be used within combination regimens containing paclitaxel and not as a single agent.

5. B-1. Which patient/disease characteristics should be considered as entry criteria or at least as strata for subgroup analysis in trials?

- **The following patient/disease characteristics should be formally considered for patient entry or as stratification factors: primary site, stage, prior treatment history, histological type, grade, residual disease, measurable or non-measurable disease, serum CA 125, performance status, age and co-morbidity, and other validated prognostic factors. For post-recurrence/progression trials: disease-related symptoms and treatment-free interval.**
- **Before exclusion of any particular patient group the following questions should be considered:**
  - **(a) Is the prognosis of these patients sufficiently different to the group as a whole to conclude without further information that it is inappropriate to include this group of patients?**
  - **(b) Is there good biological, medical or statistical evidence that the treatment is predicted to be considerably more or less effective (or even ineffective) in this group of patients?**
  - **(c) Is the result from the trial likely to be applied to this group of patients?**

Level of acceptance: 13/13

This statement should assist in decision making regarding study design. Taking into account that most study results are generalized for the whole patient population, study groups should ensure exclusion criteria are not too rigorous. Examples for each of these questions are: (a) patients with early ovarian cancer FIGO stage IA grade I most likely will be excluded from chemotherapy trials for ovarian cancer patients at higher risk for relapse; (b) patients whose tumors do not overexpress a specific biologic marker might be excluded from studies evaluating the role of an agent that is expected to work only in patients with tumors that overexpress that specific marker (e.g. HER2neu-negative patients in trastuzumab trials); (c) elderly patients should not be excluded from trials evaluating standard

chemotherapy regimens, because it is very likely that results of this trial will be generalized to this patient population also.

6. B-2. Which kind of phase III randomized study designs can be recommended to the study groups to make future trials quicker, cheaper and more reliable?

- **There is a continuing need to conduct large-scale randomized trials requiring international collaborations through the GCIG.**
- **The primary determinants for whether to use multi-arm or two-arm designs are study objectives, prioritization of the clinical questions and the availability of resources.**
- **When questions to be answered are of similar priority, multi-arm trials may be preferable.**

Level of acceptance: 13/13

7. B-3. Which are the recommended primary end points for future phase II and randomized phase III clinical trials in ovarian cancer? The recommended primary end points for future clinical trials in ovarian cancer are:

- **Phase II Screening for activity: Response* (objective RECIST or GCIG defined CA 125: to be specified in each protocol) (*for non-cytotoxic or biologic agents, other end points such as non-progression, immune response, etc., are being investigated, but are not yet validated).**
- **Phase III**

  Early ovarian cancer: Recurrence-free survival (note: recurrence = recurrent disease + deaths from any cause).

  Advanced first-line: Both PFS and OS are important end points to understand the full impact of any new treatment. Thus, either may be designated as the primary end point. Regardless of which is selected, the study should be powered so both PFS and OS can be appropriately evaluated.

  Maintenance following first-line: OS[1] minority statement

  Post-recurrence/progression trials: The choice of the primary end point needs to be fully justified with appropriate power calculations. Symptom control/quality of life (for early relapse) and OS (for late relapse) may be the preferred primary end points, although PFS should still be used in the assessment of new treatments. Whatever the primary end point, the ability of the study design to detect important differences in survival should be formally addressed.
- **Interim analysis: end points**

Time points for all efficacy analyses should be pre-specified in the protocol.

- **Early stopping/reporting for benefit:**

Primary end point.

If OS is not the primary end point then it is highly recommended that any stopping guidelines include specific criteria for stopping separately for both the primary end point and OS.

- **Early stopping for lack of benefit (in phase III or phase II–III)**

Primary or intermediate end points.

Level of acceptance: 10/13 (for whole statement)

[1]Minority vote by ANZGOG, RTOG, SGCTG: in certain situations PFS can also be considered a primary end point in maintenance trials following first-line therapy.

With the exception of 7. B-3, the level of acceptance was 13/13.

This statement deals with recommendations regarding primary end points for clinical trials. Other end points can be incorporated as secondary end points, and therefore recommendations should not be accepted exclusively. It should be mentioned that the selection of PFS as primary end point mandates rigorous definition of follow-up schedules (see statement 2). OS was regarded the preferable end point in studies evaluating maintenance therapies. However, there was a minority opinion that PFS may also be the primary end point in maintenance therapy trials.

8. C-1. Should maintenance/consolidation treatment be recommended for standard arms in future trials?

- **Current data do not support a recommendation of maintenance/consolidation treatment as a standard arm in future trials.**

Level of acceptance: 13/13

All participants believed that further investigations on the role of maintenance therapy are warranted and, ideally, such maintenance therapy should be compared with an observation arm.

9. C-2. Should dose-intense therapy or intraperitoneal therapy be a standard arm of clinical trials in first-line treatment?

- **There is no role for dose intense therapy with or without hematopoietic support or for intraperitoneal therapy as a standard arm in first-line treatment.**
- **Although there are randomized phase III clinical trials addressing the intraperitoneal route of cisplatin therapy in patients with minimal disease, interpretation of the results remains controversial, and therefore its use has not been widely adopted.**

Level of acceptance: 13/13

The conference believed that further investigations into the role of dose density are warranted. Trials evaluating dose intense therapies or intraperitoneal treatment require design improvement.

10. C3. Are there any subgroups defined by tumor biology who need specific treatment options/trials (and should not be included in 'mainstream trials')?

- **All subgroups of invasive epithelial ovarian cancer should be included in trials until specific studies are available.**

- **Patients with tumors of low malignant potential should not be included in future trials of invasive epithelial ovarian cancer.**

Level of acceptance: 13/13

The conference recognized that as more evidence becomes available, certain histological subtypes might show different biologic behaviors, particularly clear-cell and mucinous cancers. These subtypes may be further defined through molecular characterization. Currently, however, there are insufficient data to exclude any subtypes from trials. Different histological subtypes should be documented within phase III trials to allow subgroup analyses/meta-analyses.

11. C4. How to integrate new treatment modalities into studies?

- **It is currently unclear how to best integrate new treatment modalities into studies; however, identification and validation of predictors of response to new biological agents such as targeted therapies, vaccines and monoclonal antibodies should be a priority in such studies.**
- **Standard clinical end points should continue to be used in phase III studies.**

Level of acceptance: 13/13

The conference was aware of the problems of applying 'old methodology' to 'new approaches', but there was a strong feeling that the optimal use of these new agents is unknown and that it is premature to change study design.

12. C-5. How to integrate translational research in clinical trials in ovarian cancer?

- **Translational research should be considered in the planning of future clinical trials.**
- **Integration requires harmonization of consent processes and standardization of databases, including minimum datasets, and specimen banks, including central pathology review.**
- **Regulatory aspects of shared samples need facilitation.**
- **GCIG trials should have early consultation with GCIG translational research group.**

Level of acceptance: 13/13

The GCIG has gathered a large experience within its translational research and harmonization groups, with the latter having established uniform consent forms and defined regulatory issues associated with sample sharing. Both working groups could offer support if other study groups decide to include translational research in large randomized trials.

Conclusions

The 3rd International OCCC held by the Gynecological Cancer Intergroup in Baden-Baden, Germany, 5–9 September 2004 provided the first worldwide consensus on 12 important questions regarding the standards of care and future research in ovarian cancer. This was the first attempt to integrate a disparate variety of study groups from four continents and to represent each group's view through a structured consensus process. The process was so effective that the level of acceptance was high with unanimous decisions on 11 of 12 statements and only one minority report on a part of statement 7 (Table 1).

It is hoped that this high level of acceptance will help implementation of the consensus statements worldwide. The impact of this consensus conference on future studies will be evaluated in the next OCCC.

Table 1. Votes of each participating group on each consensus statement of the 3rd International Ovarian Cancer Consensus Conference 2004

Study group	Consensus statements											
	1	2	3	4	5	6	7	8	9	10	11	12
AGO-OVAR	y	y	y	y	y	y	y	y	y	y	y	y
ANZGOG	y	y	y	y	y	y	(n)	y	y	y	y	y
EORTC-GCG	y	y	y	y	y	y	y	y	y	y	y	y
GEICO	y	y	y	y	y	y	y	y	y	y	y	y
GINECO	y	y	y	y	y	y	y	y	y	y	y	y
GOG	y	y	y	y	y	y	y	y	y	y	y	y
JGOG	y	y	y	y	y	y	y	y	y	y	y	y
MRC/NCRI	y	y	y	y	y	y	y	y	y	y	y	y
NCI-US	y	y	y	y	y	y	y	y	y	y	y	y
NCIC-CTG	y	y	y	y	y	y	y	y	y	y	y	y
NSGO	y	y	y	y	y	y	y	y	y	y	y	y
RTOG	y	y	y	y	y	y	(n)	y	y	y	y	y
SGCTG	y	y	y	y	y	y	(n)	y	y	y	y	y
Level of acceptance	13/13	13/13	13/13	13/13	13/13	13/13	10/13	13/13	13/13	13/13	13/13	13/13

y, yes/agreed; (n), partial disagreement with minority report.

Acknowledgements

This document reports the final consensus statements as developed during the 3rd International OCCC, which was held by the GCIG from 5–9 September in Baden-Baden, Germany. This conference was supported by an unrestricted grant from Bristol-Myers-Squibb. Selection of participants, agenda and deliberations were not influenced by the financial support provider.

GCIG thanks AGO-OVAR for hosting this 3rd International OCCC; the scientific secretaries Monica Bacon (GCIG secretary, NCIC-CTG, Canada), Maret Bauer (AGO-OVAR Kiel, Germany), Karsten Gnauert (AGO-OVAR Wiesbaden, Germany) and Phillip Harter (AGO-OVAR Wiesbaden, Germany) for their support; Heidrun Bugl and Kinga Tahy (EMC Munich, Germany) for technical and organisational help, Bristol-Myers-Squibb

(BMS Munich, Germany) for supporting this conference with an unrestricted grant, and Glaxo Smithkline (GSK Munich, Germany) for supporting the publication of this supplement.

References

1. Neijt JP, Wiltshaw E, Lund B (eds). Advanced epithelial ovarian cancer: Where do we stand and where do we go? Ann Oncol 1993; 4 (Suppl 4): S1–S88.
2. Neijt JP, du Bois A, Williams C (eds). Advanced epithelial ovarian cancer – What do we know and what do we need? Ann Oncol 1999; 10 (Suppl 1): S1–S92.
3. Vermorken J, Avall-Lundqvist E, Pfisterer J et al. GCIG: History and current status. Ann Oncol viii39–viii42.
4. Stuart G, Avall-Lundqvist E, du Bois A et al. 3rd International Ovarian Cancer Consensus Conference: outstanding issues for future consideration. Ann Oncol 2005; 16 (Suppl 8): viii36–viii38.
5. Thigpen T, Stuart G, du Bois A et al. Clinical trials in ovarian carcinoma: requirements for standard approaches and regimens. Ann Oncol 2005; 16 (Suppl 8): viii13–viii19.
6. Vermorken J, Parmar M, Brady M et al. Clinical trials in ovarian carcinoma: study methodology. Ann Oncol 2005; 16 (Suppl 8): viii20–viii29.
7. Quinn M, Pfisterer J, Avall-Lundqvist E et al. Integration of new or experimental treatment options and new approaches to clinical trials. Ann Oncol 2005; 16 (Suppl 8): viii30–viii35.

Annals of Oncology 16 (Supplement 8) viii20–viii29, 2005
doi:10.1093/annonc/mdi963

Clinical trials in ovarian carcinoma: study methodology

J. B. Vermorken[1]*, M. K. B. Parmar[2], M. F. Brady[3], E. A. Eisenhauer[4], T. Hogberg[5], R. F. Ozols[3], J. Rochon[6], G. J. S. Rustin[2], S. Sagae[7] & R. H. M. Verheijen[1]

Gynecologic Cancer Intergroup (GCIG) and its member organizations: [1]EORTC (Europe); [2]MRC/NCRI (UK); [3]GOG (USA); [4]NCIC-CTG (Canada); [5]NSGO (Scandinavia); [6]AGO-OVAR (Germany); [7]JGOG (Japan)

Introduction

One of the major issues in the 3rd Ovarian Cancer Consensus Conference (OCCC) was to achieve consensus among study groups worldwide on appropriate requirements for entry criteria and end points for clinical trials in ovarian cancer. A 'clinical trial' was defined as a carefully designed, prospective medical study that attempts to answer a precisely defined set of questions with respect to the effects of a particular treatment or treatments [1]. The focus was primarily on phase II and III trials: the goal of the latter is to determine either the effectiveness of a treatment relative to the best current standard of care or whether a new treatment is as effective as a standard, but associated with less toxicity, cost or better quality of life. The design, execution and analysis of phase III trials not only should be based on sound scientific and ethical criteria, but it was agreed by all attendees that such trials must have sufficient statistical power to undertake an analysis of survival [2]. Historically, inadequately powered trials have undermined our ability to draw reliable conclusions on the values of different treatment approaches. As a result, several important questions remain the subject of continuing debate, despite randomized studies, including the exact role of chemotherapy in patients with high-risk early ovarian cancer after comprehensive surgical staging, the optimal number of treatment cycles in the treatment of advanced disease, the role of maintenance therapy and/or consolidation therapy, and the usefulness of dose-dense therapies and high-dose chemotherapy with autologous stem cell support. The future will be even more demanding with the evaluation of new drugs aimed at an ever increasing number of molecular targets [3]. For these reasons a worldwide consensus on standards for trials, particularly randomized studies, seems to be very timely.

Of the 12 questions that were addressed during the OCCC, three concerned study methodology and are the subject of this paper. These questions were as follows (Table 1).

Table 1. Consensus questions addressing the topic 'study methodology'

1. Which patient/disease characteristics should be considered as entry criteria or at least as strata for subgroup analysis in trials?
2. Which kind of phase III randomized study design can be recommended to the study groups to make future trials quicker, cheaper and more reliable?
3. Which are the recommended primary end points for future phase II and randomized phase III clinical trials in ovarian cancer?

Question 1. Inclusion criteria for ovarian cancer clinical trials: with focus at strict versus broad eligibility (ICON-like) criteria

In defining inclusion criteria for trials, one must consider whether certain baseline disease or patient factors lead to sufficiently different outcomes such that differing treatments or trials are appropriate. For these reasons, ovarian cancer studies have been conducted in three broad separate settings: front-line therapy in early disease, front-line therapy in advanced disease (as defined below) and therapy in recurrent disease. However, even within these categories, often clinical and pathological factors have been shown to have prognostic impact. For *advanced* ovarian cancer (FIGO stages IIB–IV) the 2nd OCCC (1998, Bergen aan Zee, The Netherlands) recommended that for adequate analysis the following details of known prognostic importance should be recorded on patients who were entered on front-line studies: age, performance status, histology, tumor grade (degree of differentiation), stage and residual disease (microscopic or none versus macroscopic) [4]. Entry criteria usually specify the limits of eligibility around these parameters.

In contrast, the two most important prognostic factors in patients with *early* ovarian cancer (FIGO stages I–IIA) are the degree of differentiation (grade of the disease) and the completeness of staging [5, 6]. However, stage, extracapsular growth, spontaneous rupture, the presence of ascites, DNA ploidy or DNA index (a quantitative pathology measure) and elevated CA 125 have also been identified as independent prognostic factors in some multivariate analyses and thus some of these are often specified as part of entry criteria [7, 8].

Most studies in patients with *recurrent* disease have written entry criteria based on those factors predictive of response to treatment rather than on survival. Time since last chemotherapy

Correspondence to: Dr J. B. Vermorken, Department of Oncology, University Hospital Antwerp, 2650 Edegem, Belgium.
Tel: +32 3-821-39-54. Fax: +32 3-825-05-64.
E-mail: Jan.b.Vermorken@uza.be

© 2005 European Society for Medical Oncology

has been the most commonly used measure to predict the likelihood of response to second-line therapy and many trials segment or restrict the population according to pre-specified time periods [7]. However, a large meta-analysis of several second-line chemotherapy studies (using data from >700 patients) indicated that other factors could also play a role in determining response, such as disease bulk, number of disease sites involved and histology [9]. Using the same dataset, significant factors at the start of second- or third-line therapy for longer subsequent survival were longer time since diagnosis, longer time since last chemotherapy, better performance status, low disease bulk, histology (non-mucinous), fewer disease sites involved and a normal hemoglobin level [10].

These data suggest that different prognostic groups can be identified even within the three traditional categories and that perhaps cohorts of patients defined by these criteria should be treated differently. This reasoning could be used as justification to use more restrictive eligibility criteria and to perform trials in multiple smaller subsets of patients. It should be understood, though, that even when trial entry is restricted, heterogeneity in the types of patients actually entered will take place and that, if too narrow a population is stipulated, the trial results may not be generalizable to the entire population. Furthermore, even if the plan is to be reasonably restrictive in patient entry for the purposes of being able to make comparisons across trials, there are problems in doing so. The following examples highlight these points.

(a) In *early* ovarian cancer several recent studies have focused on the role of chemotherapy in the so-called high-risk disease setting (ICON1, ACTION, GOG#157 [11–14]). Both ICON1 and ACTION compared platinum-containing adjuvant chemotherapy versus observation following surgery. In both studies the primary end point was survival. However, the entry criteria in both trials were different. In ICON1 these were quite liberal, i.e. any patient in whom the clinician was uncertain whether that patient should receive chemotherapy could enter the trial and surgery primarily consisted of total hysterectomy and bilateral salpingo-oophorectomy [13]. In the ACTION trial, however, the entry criteria were more restrictive, i.e. only patients with FIGO stages IA, IB (grades II and III), stages IC and IIA (all grades) and all clear cell carcinomas could enter the trial. Furthermore, ACTION had more strict guidelines for surgery consisting of total abdominal hysterectomy and bilateral salpingo-oophorectomy, followed by surgical staging, as indicated in the EORTC surgical guidelines [14]. The combined analysis on 925 patients showed an 8% improvement in survival for immediate chemotherapy versus observation, and despite differences in eligibility criteria (liberal or restricted), staging requirements and chemotherapy (more single-agent carboplatin in ICON1, more cisplatin-based combinations in ACTION), the individual results of the two trials were very similar, with the magnitude of the effect of chemotherapy being of very similar size. Subgroup analysis failed to demonstrate a different effect of chemotherapy in any of the subgroups that could be analyzed, i.e. age, tumor stage, histological cell type and grade of differentiation [11]. Unfortunately, the relationship between staging performance and the effect of chemotherapy could not be adequately analyzed. Only one-sixth of the population was optimally staged and, despite the more strict staging requirements in ACTION, only one-third of the patients received proper staging. A separate analysis of the ACTION trial suggested that the benefit of adjuvant chemotherapy was limited to patients with non-optimal staging [14]. However, there was not enough statistical power to deny a positive effect in patients who had been optimally staged. So, the interpretation remains unclear leading to a variety of attitudes in different countries. In GOG#157, a trial that studied the impact of longer versus shorter duration of adjuvant chemotherapy, staging was required as per GOG published guidelines [12]. However, of the 457 patients, only 70% met all eligibility criteria and 23% (107/457) were incompletely staged. These data suggest that in daily practice in non-specialized centers the percentage of patients with optimal staging will be substantially lower. So, optimal staging in early ovarian cancer remains problematic and the relevance of it to patients treated outside of clinical trials might be even more questionable.

(b) In the *advanced* disease setting similar difficulties have been encountered. Examples of this are the remarkable differences in outcome between protocol GOG#111 and protocol GOG#132, trials which were performed in sequence by the GOG in patients with the same eligibility criteria (suboptimal stages III and IV), and using the same treatment (paclitaxel 135 mg/m^2, 24 h, plus cisplatin 75 mg/m^2 every 3 weeks for six cycles). Progression-free and overall survival were 18 and 38 months, respectively, in GOG#111, and 14 and 27 months, respectively, in GOG#132 [15, 16]. So, even though the eligibility criteria were the same, there must have been a selection bias with worst prognosis patients in GOG#132. This means that other criteria besides stage and volume are important and need to be identified.

The assessment of the amount of residual disease is a particularly difficult item and open to much variation in subjective interpretation. There are now at least three large randomized trials showing that progression-free or overall survival are improved when cisplatin-based intraperitoneal (i.p.) chemotherapy is applied compared with intravenous administration of platinum-based chemotherapy [17–19]. The fact that in the first positive trial (the purest of all) no statistical significant advantage for intraperitoneal chemotherapy was found in the subset of patients with <0.5 cm disease is still puzzling and has led to a negative interpretation by some and has reopened the debate as to which patient population will ultimately benefit from i.p. therapy. How standard and objective are the methods of measuring the size of the largest residual peritoneal mass in the operating room? Maybe the distinction between suboptimally and optimally debulked disease should indeed be made on the basis of macroscopic versus no macroscopic disease left after surgery and i.p. therapies should be further studied in the latter category.

Clearly it seems very difficult, if not impossible, to draw reliable conclusions when comparing across different clinical trials, however similar they appear to be. This makes it even more important that comparisons within trials are as reliable as possible, which in turn emphasizes the overriding need for large-scale studies whenever possible.

(c) A further negative effect of having very strict eligibility criteria is that it may lead to slow accrual. A clear example of this is EORTC protocol 55875, a randomized phase III study in ovarian cancer patients with a pathologically complete remission after platinum-based intravenous chemotherapy [20]. The study evaluated the role of i.p. cisplatin versus no further treatment. It took 9 years to accrue 153 patients and the study was closed prematurely, and suffered from the inclusion of women who were either ineligible (10%) or had major protocol violations (11%), part of which were most likely related to limited experience with the technical aspects of i.p. therapy. Moreover, there was an awareness of a progressive change in 'standard' first-line intravenous chemotherapy, with paclitaxel–cisplatin progressively replacing cyclophosphamide–cisplatin. Nevertheless, the trial showed the same trend as the other i.p. trials, i.e. a superior outcome.

These examples gave sufficient room for discussion on how strict or how flexible one should be with respect to eligibility criteria. The less restrictive (ICON-like) eligibility criteria seem more in line with what is applicable to the general population; with this approach more patients can be entered and accrual is facilitated. The volume of the residuum in stage III disease may be a biased criterion when it concerns measurement of residual disease; however, as mentioned earlier, further studies in patients with no residual macroscopic disease may overcome this bias.

Therefore, working group B concluded that there are no hard and fast rules as to which types of patients should be entered into phase III clinical trials. However, some considerations need to be taken into account.

We noted that the first randomized trial of a new therapy is often carried out, partly as matter of expediency, in patients with stage IV or recurrent disease, where there is a high event rate and thus results come more quickly. If results are positive, this has sometimes led to further trials in earlier stages of the disease. We need to be aware that such a model might be appropriate, but might also be misleading. For example, 5-fluorouracil is not very active in advanced colon cancer, but has now become a mainstay of adjuvant treatment of this disease.

In any framework that considers inclusion and exclusion criteria for trials in ovarian cancer, it is important to consider not only who should be *included* in any given trial, but also whether any particular subgroup should be *excluded*. To address this it is useful to consider the following three questions:

a) Is the prognosis of the subgroup of patients sufficiently different to the group as a whole to conclude without further information that it is inappropriate to include this group of patients?

As an example, it is very unlikely that we would include stage IA grade I patients in the same trial as stage IV patients. This is mainly because the prognosis of these two groups of patients is so different that it is very unlikely that after surgery we would want to follow similar treatment strategies for them. However, the same argument may not hold for stage III and IV patients, whose prognosis although different, may not be different enough to *a priori* entertain different treatment strategies.

b) Is there 'good' biological, medical or statistical evidence that the treatment is going to be considerably more or less effective (or even ineffective) in a particular subgroup of patients?

As an example, we know that many therapies are likely to be more effective in patients with recurrent disease who have platinum-sensitive disease than in patients with platinum-refractory disease. Thus for most new therapies we would not include both groups of patients in the same trial. An exception would be, if there are strong preclinical data that an agent may be active only when a specific biologic marker is present, to include ovarian cancer patients with tumors that overexpress the marker and to exclude those whose tumors do not. An example of this is testing trastuzumab only in ovarian cancer patients with measurable persistent or recurrent epithelial ovarian cancer with 2+ or 3+ HER2 overexpression [21].

c) Is it likely that any result from the trial will be generally extrapolated to include a particular group of patients?

As an example: we know that if we perform a trial in stage IV disease, that in many cases any result is likely to be extrapolated to patients with stage III disease. In this case it would have been better to also include the stage III patients into the trial to assess whether there is evidence of a different size of effect in stage III and IV patients. Another common example is that of elderly patients who should most probably not be excluded from trials evaluating standard chemotherapy regimens, because it is very likely that results of such trials will be generalized to this patient population also.

To summarize, if the answers to questions (a) and (b) are 'no', then strong consideration should be given to *including* the subgroup of patients in the trial. Whatever the answers to (a) and (b), if the answer to (c) is 'yes' then, again, consideration should be given to *including* this group of patients into the trial. Thus, the answer to the first question on study methodology is as follows, which after discussion with the whole consensus panel obtained a unanimous acceptance (Table 2).

Table 2. Consensus statements in response to question 1

Which patient/disease characteristics should be considered as entry criteria or at least as strata for subgroup analysis in trials?
The following patient/disease characteristics should be formally considered for patients entry or as stratification factors:
Primary site, stage, prior treatment history, histological type, grade, residual disease, measurable or non-measurable disease, serum CA 125, performance status, age and co-morbidity and other validated prognostic factors. For post-recurrence/progression trials: disease-related symptoms and treatment-free interval.
Before exclusion of any particular patient group the following questions should be considered:
Is the prognosis of these patients sufficiently different to the group as a whole to conclude without further information that it is inappropriate to include this group of patients?
Is there good biological, medical or statistical evidence that the treatment is predicted to be considerably more or less effective (or even ineffective) in this group of patients?
Is the result from the trial likely to be applied to this group of patients?

Question 2. Trial design in ovarian cancer clinical trials: with focus at multi-arm versus single-question trials

With the increasing pace of drug development and the pressure to get answers more quickly, it is reasonable to consider the most optimal design(s) for large randomized trials to arrive at answers rapidly and efficiently. The sample size needed for such trials is substantial if modest, but real, overall or progression-free survival differences are to be detected (see later). Trials performed by GCIG groups have been able to accomplish this using the traditional two-arm trials: the median sample size in the five completed GCIG first-line ovarian cancer trials was 1300 patients [see Gynecologic Cancer Intergroup (GCIG): History and current status, this issue] and accrual time for these trials ranged from 2 to 3 years.

A complicating (but fortunate) factor in deliberating efficient trial design is that, at the present time, it is not unusual for *several* promising new agents or treatment regimens to be ready simultaneously for testing in a randomized phase III setting. It may be impractical and inefficient to test several new therapies in individual trials against a control arm by conducting multiple trials using a conventional parallel two-group design. For example, too few patients may be available given the required sample size of each of these trials, or the resources needed (for example, the costs) may be too great [22]. On the other hand, performing such trials sequentially would take too much time. Therefore, novel multi-arm designs in which a control regimen is compared with several new (experimental) therapies are worth considering. The issue of using multi-arm trials or single question studies was extensively debated in working group B and led to the recommendations that can be found at the end of this section in Table 4. These recommendations were in turn accepted unanimously by other Workshop representatives.

There are pros and cons to conducting a single multi-arm trial versus several two-arm studies. Multi-arm trials can be considerably more complex to design, conduct and analyze than two-arm, single-question trials. The additional complexities can be classified as arising from ethical, administrative or scientific/statistical considerations.

Ethical challenges

Obtaining the patient's informed consent for multi-arm trials is more challenging than the simpler two-arm trial when treatment arms include a broad range of agents. Since the patient's consent must be based on making an informed decision, additional care is required to ensure that prior to enrolling onto the study the patient understands the detailed information concerning the risks associated with each of the study regimens of which only one will ultimately be administered.

Administrative challenges

Multi-arm phase III clinical trials in gynecologic malignancies are likely to require collaboration among multiple cooperative groups. For example, the five-arm advanced ovarian cancer trial, GOG-182/ICON5, involved cooperative groups from Australia, New Zealand, Italy, UK and the USA. In this case, each of these groups had prior experience and established procedures for conducting phase III trials; however, collaboration requires standardization of these procedures. Each group makes concessions in order to develop uniform procedures for study development, conduct and monitoring. Standardizing the data monitoring process requires identifying those clinical observations that are necessary to meet the study objectives and developing a common set of data forms and data definitions that can be unambiguously implemented across all treatment centers and data centers involved in the study.

Multi-national studies introduce additional unique challenges. It is not uncommon for investigational agents to be available in some countries but not in others. Moreover, the regulatory procedures enforced within each country are not universal and individual investigators are often unable to make concessions in order to promote study-wide standards. Indeed, laws and regulations in each country are not static and therefore, procedures that are apparently sufficient at the initiation of the study may require modifications before the trial is completed. Furthermore, if all or several agents to be studied are investigational, competing pharmaceutical firms may not agree to have their agents studied in the same trial for business reasons, regardless of the scientific merit of the proposal. This latter situation may call for considerable negotiating skills.

It is reasonable to expect that future multi-arm trials will require even greater organizational efforts if they include investigational agents and require direct involvement of the industry. Trial sponsors from industry will typically impose additional study objectives and constraints on the conduct and administration of the study beyond those deemed appropriate for scientific reasons.

The eligibility criteria for multi-arm trials may also be more restrictive than for two-arm trials. Each additional treatment arm may either increase the requirement for restricting eligibility or reduce the patients' interest in participating in the study. For example, trials with a regimen containing an anthracycline may make it necessary to limit eligibility to patients who have not recently experienced congestive heart failure. Trials with a taxane regimen may eliminate patients experiencing peripheral neuropathy. These eligibility criteria that are considered justifiable for safety's sake have the unfortunate cumulative impact on reducing the number of patients who can participate in the trial. Moreover, some of the otherwise eligible patients may not be willing to accept randomization to all of the study treatments. For example, in a recent multi-arm trial evaluating tamoxifen and radiotherapy for the treatment of ductal carcinoma *in situ* of the beast, 46% of the eligible patients were willing to have either radiation or tamoxifen treatment randomly assigned, but not both [23]. To some extent these eligibility restrictions and patient preferences can be mitigated in multi-arm trials by using more complex randomization and analytic procedures [24].

Scientific and statistical challenges

The scientific challenges of multi-arm trials stem from the increased number of hypotheses that can be tested. In a two-arm

trial there is only one treatment comparison; however, in a clinical trial involving k different treatments, there are potentially $k(k-1)/2$ pair-wise treatment comparisons. That is, in a trial with five treatment arms there are potentially 10 distinct pair-wise treatment comparisons. Suppose that all five of the treatment regimens are truly equivalent and at the end of the study each pair-wise treatment comparison is tested at the traditional 0.05 significance level. In this case the probability of *incorrectly* declaring at least one treatment to be superior to another is 23% (Table 3). Such a high probability for this type of error is usually considered too great for a phase III trial. Typically, phase III trials control this error (called type I error) so that it does not exceed 5%.

There are several approaches that can be considered for limiting type I errors in multi-arm trials. The first approach is to require a greater level of evidence before declaring two treatment regimens different. For example, rather than requiring a P value <0.05 in order for a difference to be considered statistically significant, a trial could require P values to be $<0.05/m$, where m is the number of planned treatment comparisons. This adjustment is commonly called the Bonferroni procedure [25]. Therefore, in a five-arm trial in which all 10 pair-wise treatment comparisons are planned, requiring the P value to be $<0.05/10 = 0.005$ will limit the study-wide probability of type I error to no more than 5%. While the Bonferroni adjustment is easily applied, this procedure also reduces the chance of detecting differences between treatments when they truly exist (statistical power). There are other adjustment procedures that can be used to control type I errors in multi-arm trials that are slightly more complicated but preferable because they are more likely to detect differences when they truly exist [26]. All these adjustment procedures improve the specificity of the trial (reduce the probability of a type I error). However, without a corresponding increase in the size of the trial, these adjustment procedures also reduce the sensitivity of the trial for detecting differences between treatments when they truly exist. Therefore, multi-arm trials typically enrol more patients onto each treatment arm than a similarly designed two-arm trial in order to improve sensitivity while controlling overall specificity.

When there are fewer treatment comparisons made, there are fewer opportunities to make an error. This suggests a second approach for limiting type I errors within a multi-arm trial by limiting the number of planned treatment comparisons. This approach may not be as undesirable as it first appears. Consider a multi-arm trial in which one of the study treatments is the standard intervention. Also, suppose that there is an *a priori* preference for the standard treatment. In other words, the standard treatment will continue to be recommended unless the trial provides overwhelming evidence indicating that at least one of the experimental regimens is significantly better than the standard intervention. In this type k-arm trial, there are only $(k-1)$ comparisons between the standard treatment group and each of the experimental treatment groups that are of immediate interest. No comparisons between the pairs of experimental treatments are planned unless one experimental regimen is deemed superior to the control arm. Therefore, the Bonferroni-adjusted critical P value for this five-arm trial is $0.05/4 = 0.0125$, rather than 0.005 as in the previous five-arm trial described earlier. If all of the treatments approaches in this five-arm study are truly equivalent, this approach limits the probability of incorrectly accepting an experimental treatment as the new standard of care to no more than 5%. In order to maintain sensitivity this approach also requires increasing the number of patients enrolled. While the number of patients to be enrolled onto each treatment arm is still larger than that required for a two-arm trial, the increase is not as large as the multi-arm trial that does not restrict the number of treatment comparisons.

It is reasonable to wonder why an investigator who plans to compare several new experimental regimens to a standard treatment in a single multi-arm trial should use statistical considerations different from the investigator who decides to study the same regimens using several sequential two-arm trials. The difference between these two approaches arises from the dependence among the treatment comparisons when a single multi-arm trial is performed. Consider a single multi-arm trial in which the control group includes slightly more patients with a good prognosis than expected. In this trial all of the experimental regimens will tend to appear less beneficial than they truly are. Likewise, if the control regimen happens to include more patients with a poor prognosis than expected, then all of the experimental regimens will appear more active than they truly are. In other words, if one experimental arm in a multi-arm trial is deemed significantly better than the control arm, then it is more likely that another experimental regimen will also be deemed significantly better than the control [27]. There is a dependence among the estimated experimental treatment effects sizes introduced into the design and analysis when all of them are being compared with a single control arm. This dependence does not occur when each experimental regimen is compared to a different control arm, as when there are multiple sequential two-arm trials.

In summary, multi-arm clinical trials have ethical, administrative and scientific considerations that may not be present in two-arm trials. An ethical challenge can arise from the necessary information that a patient needs to understand regarding several experimental regimens in order to make an informed consent prior to enrolling onto the trial. Administrative challenges may arise from the need for greater resources required for conducting multi-arm trials. The greater scientific challenge in multi-arm trials is due to the proliferation of study objectives. There is no

Table 3. Number of possible comparisons and the probability of erroneously declaring one or more treatments different (type I error) in a multi-arm trial when $\alpha = 0.05$ for each test and there is no adjustment for multiple comparisons

Number of treatment groups	Number of possible pair-wise comparisons	Probability of at least one type I error in the entire study
2	1	0.050
3	3	0.113
4	6	0.178
5	10	0.234

Table 4. Consensus statements in response to question 2

Which kind of phase III randomized study design can be recommended to the study groups to make future trials quicker, cheaper and more reliable?
There is a continuing need to conduct large scale randomized trials requiring international collaboration through the GCIG.
The primary determinants for whether to use multi-arm or two-arm designs are study objectives, prioritization of the clinical questions and the availability of resources.
When questions to be answered are of similar priority, multi-arm trials may be preferable.

longer a single alternative hypothesis in multi-arm studies. Adjustments should be made for multiple correlated estimates.

Question 3. Relevant end points for clinical trials in ovarian cancer

The first and most important step in planning a clinical trial is to indicate clearly the primary and secondary objectives [2]. What questions is the trial being designed to answer? Once the objectives are known, this identifies the primary and secondary end points of the study. Trial end points can be classified as either 'true' or 'surrogate'. True end points have direct clinical relevance to the patient, such as symptoms improvement, survival duration, or cure rates. Surrogate end points assess events that are in the etiologic pathway to a true outcome [28]. The primary reason for using a surrogate end point instead of a true end point is either to reduce the duration (because this end point occurs earlier than the actual end point) and cost of a clinical trial or if it is believed that salvage therapies may obscure the effect of the study treatment on a true end point. As an example, progression-free survival has often been considered a surrogate end point for overall survival in trials including patients with advanced ovarian cancer. It is noteworthy that the justification for using a particular surrogate end point is frequently based on data suggesting a statistical correlation with a true end point. However, a correlation between a surrogate and true end point is a necessary, but not sufficient condition to justify a particular surrogate end point. The ideal surrogate end point for randomized trials is an intermediate event in the only causal pathway to the true end point, and the effect of an intervention (i.e. a treatment) on the true end point should be through its influence on the surrogate end point [2]. Reasons for failure of a surrogate end point could be explained in several ways: i.e. either (i) of several causal pathways of disease, the intervention only affects the pathway mediated through the surrogate, or (ii) the surrogate is not in the pathway of the intervention's effect, or is insensitive to its effect.

In ovarian cancer trials the traditional patient specific outcomes of interest often include: overall survival (or cure rate) and progression-free survival, response and toxicity, and symptom control/quality of life. Of these there is general agreement on overall survival (or cure rate) and symptom improvement/quality of life as primary meaningful end points [although quality of life is not (yet) used as such] and toxicity is considered a necessary measure (and primary end point for phase I studies). However, there is debate about the importance of response and progression-free survival as being meaningful end points due to the uncertainty as to whether the patient has any benefit from a longer time to tumor progression or from tumor regression itself. Regardless of the 'meaning' of these end points in and of themselves, if either or both were shown to be true surrogates of survival or quality of life, their use as primary trial end points is easily justified.

Phase II end points

In phase II trials of new agents (or combination) in ovarian cancer, where the primary objective is to determine early evidence of biologic effect of the new drug(s), historically objective response has been defined as the primary end point. It has the advantage of being non-invasive, subject to internationally recognized standardized criteria [29] and readily determined after a series of treatment cycles. Moreover, it is not influenced by salvage therapy. Its disadvantage, though, is the fact that by definition patients must have measurable disease at baseline to be evaluated. While this is usually the case in recurrent disease, it may not be so in the front-line setting. Furthermore, inter-observer variability in declaring response, even according to objective measures, has been documented [30]. Because ovarian cancer is often associated with elevation of the well-studied serum antigen, CA 125, and since the levels of the antigen correlated with disease burden, changes in the level of CA 125 seem a plausible substitute for objective tumor regression. Following on work originally conducted by Rustin where a set of CA 125 response criteria were suggested [31], the GCIG Response/Progression Working Group has defined modified Rustin (CA 125) criteria to be used prospectively as an addition to Response Evaluation Criteria in Solid Tumours (RECIST) as a method of defining response in relapsed ovarian cancer patients [29, 32]. The validity of the 50% response definition according to Rustin (later endorsed as the GCIG response criteria) as a substitute for objective response as assessed by RECIST was confirmed by the GINECO group in France in the setting of recurrent disease [33]. Prospective validation of these modified Rustin criteria (GCIG CA 125 definition) in recurrent disease is awaited, and several groups are using these in ongoing trials. For front-line trials, CA 125 response criteria also await validation and therefore cannot be used as such in that setting yet.

Phase III trials

In randomized trials there is no systematic evidence that objective response is a surrogate for overall survival. Furthermore, there are limited data on its surrogate value in assessing quality of life. Nevertheless, it is of interest that quality of life studies in relapsed ovarian cancer patients have indicated that quality of life scores improve in patients who respond to chemotherapy, confirming the palliative nature of chemotherapy [34]. There is obviously an inverse relationship between experienced toxicity and quality of life. This has been observed in randomized phase III trials in the front-line setting, e.g. in trials in which cisplatin was replaced by carboplatin in the combination with paclitaxel

[35]. Such differences have so far not been observed in randomized trials in the recurrent disease setting, but the information about it is scarce. Interestingly, although most consider quality of life an important primary end point for trials in incurable disease settings, it is seldom, if ever, a primary end point in randomized trials in recurrent ovarian cancer.

The GCIG members have accepted the definition of CA 125 progression, in contrast to CA 125 response, as an addition to objective disease progression in front-line randomized trials [36]. A patient may be declared to have progressive disease on the basis of either the objective RECIST criteria or the CA 125 progression criteria. The date of the progression will be the date of progression of the earlier of the two events if both are documented. Since it was recognized that the timing of investigations during first-line therapy and subsequent follow-up may also influence the assessment of progression-free survival in clinical trials, it was proposed that serum CA 125 levels would be obtained on day 1 of each chemotherapy cycle, 4 weeks after the last course, thereafter every 3–4 months for the first 36 months, every 6 months from months 37–60, and every year from 5 years after the primary diagnosis [36]. Although it is recommended that the date of progression is recorded according to both CA 125 and RECIST criteria, it is important to continue validation of CA125 progression by determining whether the trial outcome would be the same whether CA 125 was used or not.

End points: front-line phase III studies

The main issue in discussing randomized phase III front-line studies is whether progression-free survival (for advanced ovarian cancer) or relapse-free survival (for early ovarian cancer) can ever be considered meaningful primary end points. If one agrees that improvement in overall survival is the finding for which we would change our standard of care, then progression-free and relapse-free survival could be considered as alternative primary end points if the available data is strong enough to consider them valid surrogates for survival. What are these data?

For *early ovarian cancer*, there is only one adequately powered trial in the adjuvant setting, the combined ICON1/ACTION analysis [11]. Results from this trial showed that relapse-free survival differences were mirrored in the overall survival analysis. Data from other therapeutic areas such as breast cancer seem to support the use of relapse-free survival as a valid surrogate for survival in the adjuvant setting. Clinicians have changed practice and drugs do get approved for significant improvements in relapse-free survival without waiting for overall survival data [37]. Thus the use of relapse-free survival as a primary end point in randomized trials of adjuvant therapy in early ovarian cancer is justified not only by extrapolation from other solid tumor settings but also by data from the largest randomized trial in early ovarian cancer itself.

The data are also strong for recommendation of progression-free survival as a primary end point in front-line trials in *advanced ovarian cancer* on several counts. Recent adequately powered trials where progression-free and overall survival are known have shown concordant observations between progression-free survival differences and overall survival differences [15, 16, 38–41]. Buyse et al. [42] showed in a meta-analysis of advanced ovarian cancer trials (data from the Ovarian Cancer Meta-analysis Project [43]) that by applying a new method for validation of surrogate end points the treatment effects on the true end point (logarithm of survival) and the treatment effects on the surrogate end point (logarithm of time to progression) were highly correlated. Looking at the predictions of the effect of treatment on log (survival), based on the observed effect of treatment on log (time to progression), the authors concluded that time to progression could be used as a surrogate for survival in advanced ovarian cancer. The effect of treatment could be observed earlier if time to progression were used instead of survival and the effect was also somewhat more pronounced. Hence, a trial that used time to progression would require less follow-up time and fewer patients to establish the statistical significance of a truly superior treatment than a trial that used survival. The gains, however, would be modest because progression was followed by death within 1 year for most patients. Thus in the front-line setting both progression-free survival as a surrogate end point, and overall survival as a true end point are supported by evidence as reasonable primary end points. If progression-free survival is the primary end point, however, and an advantage to a new treatment is shown, information on the survival impact of that treatment will also be an important adjunct to trial results since, regardless of the historical weight of evidence supporting progression-free survival as a primary end point, clinicians will eventually want to know the survival outcome of a particular study. This may be even more important for phase III studies, in which new biological and targeted therapies are investigated, because it is not at all clear whether the relationship between progression-free survival (as a surrogate for overall survival) and overall survival (as the true end point), which is largely based on studies with chemotherapy, also applies for these newer and different forms of therapy. Furthermore, even if further therapy in the control arm at the time of progression dilutes the impact of the new treatment on the end point of overall survival, this would be important to know, because this may suggest that a policy of using the control therapy first, then using another therapy at the time of progression may be as good as using the new treatment in first-line. This all implies that trials should be adequately powered to address both end points with adequate follow-up. If progression-free survival is the primary end point, earlier reporting of data is possible and positive results may lead to earlier adoption of new treatments in some jurisdictions. Nevertheless, this should be followed by the reporting of overall survival data at some stage, to allow a full picture of the policy of using the two treatments to emerge.

End points: second-line phase III studies

For phase III trials in the second-line setting, progression-free survival does not seem to be a good surrogate for survival: there are several examples where progression-free survival was significantly improved, with no survival impact [44–47]. It can be argued that some of these studies were underpowered to detect survival improvements; however, the weight of evidence to consider progression-free survival a surrogate for survival,

and thus a primary end point in the second-line setting, is not strong as yet. In the recurrent disease setting, overall survival remains an important primary end point (particularly if more costly or toxic therapy is being offered). Progression-free survival data remain of interest but are unlikely to be sufficiently persuasive to shift practice patterns. Furthermore, since the rationale for treating patients with relapsed disease is a desire to improve symptoms and thus quality of life, an adequate measure of these factors would also be an appropriate primary end point for randomized trials. However, no universally acknowledged and standardized system of symptom measurement and analysis is readily available. GCIG will continue, through its working groups, to build a consensus on how meaningful improvements in disease-related symptoms can be quantified.

End points: maintenance/consolidation phase III studies

A special issue is maintenance and consolidation trials (see also the summary of Workshop C: Integration of new or experimental treatment options and new approaches to clinical trial, this issue). To date, randomized trials with both cytotoxic agents and biological agents are negative, both for progression-free and overall survival, with the exception of the SWOG/GOG trial, which showed a significant difference in progression-free survival in favor of the 12 versus 3 months of maintenance paclitaxel after complete response to platinum and paclitaxel-based chemotherapy [48]. This study was stopped early after a planned interim analysis based on progression-free survival outcomes. Because patients were informed and allowed to continue treatment for 12 months on the 3-month arm, this precluded any meaningful analysis of overall survival benefit. Since trials involving maintenance by definition have longer treatment on the experimental arm as compared with the control, it seems reasonable to expect that progression might be delayed: the real question is whether the prolonged therapy improves survival. Thus, overall survival is the primary end point that should be selected for trials of this design. Interestingly, the next trial in the USA employing prolonged consolidation will randomize patients to no further therapy after front-line chemotherapy versus taxane and will consider overall survival as the primary end point.

End points for interim analysis

The example of the SWOG/GOG trial also raises the issue of early stopping/interim analysis of randomized trials. All such analyses must be pre-specified in the protocol (which was in fact the case in the example cited). However, early stopping for extreme differences (*benefit*) should be based on the primary end point, not an intermediate or surrogate end point since, as was the case for the SWOG/GOG study, to do otherwise may forever impair the ability to perform an analysis of the primary study end point [49]. Therefore, if the primary end point is not overall survival (but, for example, progression-free survival) we suggest that early stopping guidelines for benefit should include both the primary end point and overall survival for the reasons described in the previous section. In cases when the stopping rule is geared to halt the study for reasons of *lack of benefit*, the end point for the analysis may reasonably be either the primary end point or a valid intermediate/surrogate end point.

End points for studies of non-cytotoxic agents

The use of agents that target novel molecular changes in malignancy (as opposed to the usual cytotoxic targets of DNA and tubulin) has raised some interesting questions about study design and end points. Thus, data on non-cytotoxics in ovarian cancer, and in some other tumor types, do not suggest that end points being used in phase I or II trials are any different from those used in trials with cytotoxic agents [21, 50]. While some novel end points, particularly for phase II trials, such as non-progression or imaging measures have been proposed, these are not yet validated and should await this step before application except on an experimental basis.

Once non-cytotoxic drugs are in phase III evaluation, there is no reason to consider end points other than those described above. It will still be important to determine, before changing practice, what the impact of the new agent is on overall, relapse-free or progression-free survival, depending on the

Table 5. Consensus statements in response to question 3

Which are the recommended primary end points for future phase II and randomized phase III clinical trials in ovarian cancer?
The recommended primary end points for future clinical trials in ovarian cancer are:
Phase II screening for activity: response[a] (objective RECIST or GCIG defined CA 125: to be specified in each protocol)
Phase III
Early ovarian cancer: recurrence-free survival (note: recurrence = recurrent disease + death from any cause)
Advanced first-line: both progression-free survival (PFS) and overall survival (OS) are important end points to understand the full impact of any new treatment. Thus either may be designated as the primary end point. Regardless of which is selected, the study should be powered so both PFS and OS can be appropriately evaluated.
Maintenance following first-line: OS[1 minority statement]
Post-recurrence/progression trials: The choice of the primary end point needs to be fully justified with appropriate power calculations. Symptom control/quality of life (for early relapse) and OS (for late relapse) may be the preferred primary end point although PFS should still be used in the assessment of new treatments. Whatever the primary end point, the ability of the study design to detect important differences in survival should be formally addressed.
Interim analysis: end points
Time points for all efficacy analyses should be pre-specified in the protocol
Early stopping/reporting for benefit
Primary end point
If OS is not the primary end point then it is highly recommended that any stopping guidelines include specific criteria for stopping separately for both the primary end point and OS
Early stopping for lack of benefit (in phase III or phase II–III)
Primary or intermediate end points

[a]For non-cytotoxic or biologic agents, other end points such as non-progression, immune response, etc., are being investigated, but are not yet validated.

phase III setting. Thus far, investigators continue to design phase III trials of non-cytotoxic agents using traditional clinical end points [51, 52].

Summary of end point recommendations

With all the above-mentioned considerations working group B formulated the recommendations listed in Table 5.

Acknowledgements

This document summarizes the discussions held in the group B section during the 3rd Ovarian Cancer Consensus Conference, organized by the GCIG and held in Baden-Baden, Germany, 3–5 September 2004. None of the participants in this section indicated any conflict of interest related to the sponsor of this event (Bristol-Myers Squibb).

References

1. Sylvester R. Planning cancer clinical trials. In Buyse ME, Staquet MJ, Sylvester RJ (eds): Oxford Cancer Clinical Trials – Methods and Practice. 47–63. Oxford University Press 1982.
2. Brady MF, Thigpen JT, Vermorken JB, Parmar MKB. Randomised trials in ovarian cancer: trial design considerations. Ann Oncol 1999; 10 (Suppl 1): S75–S82.
3. Cannistra SA. Cancer of the ovary. N Engl J Med 2004; 351: 2519–2529.
4. Berek JS, Bertelsen K, du Bois A et al. Advanced epithelial ovarian cancer: 1998 consensus statements. Ann Oncol 1999; 10 (Suppl 1): S87–S92.
5. Vergote I, De Brabanter J, Fyles A et al. Prognostic importance of degree of differentiation and cyst rupture in stage I invasive epithelial ovarian carcinoma. Lancet 2001; 357: 176–182.
6. Zanetta G. The accuracy of staging: an important prognostic determinator in stage I ovarian carcinoma. A multivariate analysis. Ann Oncol 1998; 9: 1097–1101.
7. Eisenhauer EA, Gore M, Neijt JP. Ovarian cancer: should we be managing patients with good and bad prognostic factors in the same manner? Ann Oncol 1999; 10 (Suppl 1): S9–S15.
8. Tropé C, Kaern J, Hogberg T et al. Randomized study on adjuvant chemotherapy in stage I high-risk ovarian cancer with evaluation of DNA-ploidy as prognostic instrument. Ann Oncol 2000; 11: 281–288.
9. Eisenhauer EA, Vermorken JB, van Glabbeke M. Predictors of response to subsequent chemotherapy in platinum pretreated ovarian cancer: an analysis of 704 patients. Ann Oncol 1997; 8: 963–968.
10. van Glabbeke M, Vermorken J, Eisenhauer E. Predictors of survival (OS) and time to progression (TTP) following 2nd and 3rd line treatment in advanced ovarian cancer (OVCA): An analysis of 744 patients (pts) treated with paclitaxel, docetaxel, or high dose epirubicin. Proc Am Soc Clin Oncol 1997; 16: 254a (Abstr 1265).
11. ICON1 and EORTC-ACTION Collaborators. International Collaborative Ovarian Neoplasm trial 1 and Adjuvant Chemotherapy In Ovarian Neoplasm trial: two parallel randomized phase III trials of adjuvant chemotherapy in patients with early-stage ovarian carcinoma. J Natl Cancer Inst 2003; 95: 105–112.
12. Bell J, Brady M, Lage J et al. A randomized trial of three versus six cycles of carboplatin and paclitaxel as adjuvant treatment in early stage ovarian epithelial carcinoma: a Gynecologic Oncology Group study. Gynecol Oncol 2003; 88: 156 (Abstr 1).
13. International Collaborative Ovarian Neoplasm (ICON1) Collaborators. International Collaborative Ovarian Neoplasm Trial 1: a randomized trial of adjuvant chemotherapy in women with early-stage ovarian cancer. J Natl Cancer Inst 2003; 95: 125–132.
14. Trimbos JB, Vergote I, Bolis G et al. Impact of adjuvant chemotherapy and surgical staging in early-stage ovarian carcinoma: European Organization for Research and treatment of Cancer – Adjuvant Chemotherapy in Ovarian Neoplasm Trial. J Natl Cancer Inst 2003; 95: 113–125.
15. McGuire WP, Hoskins WJ, Brady MF et al. Cyclophosphamide and cisplatin compared with paclitaxel and cisplatin in patients with stage III and IV ovarian cancer. N Engl J Med 1996; 334: 1–6.
16. Muggia FM, Braly PS, Brady MF et al. Phase III randomized study of cisplatin versus paclitaxel versus cisplatin and paclitaxel in patients with suboptimal stage III or IV ovarian cancer: a Gynecologic Oncology Group study. J Clin Oncol 2000; 18: 106–115.
17. Alberts DS, Liu PY, Hannigan EV et al. Intraperitoneal cisplatin plus intravenous cyclophosphamide versus intravenous cisplatin plus intravenous cyclophosphamide for stage III ovarian cancer. N Engl J Med 1996; 335: 1950–1955.
18. Markman M, Bundy BN, Alberts DS et al. Phase III trial of standard-dose intravenous cisplatin plus paclitaxel versus moderately high-dose carboplatin followed by intravenous paclitaxel and intraperitoneal cisplatin in small-volume stage III ovarian carcinoma: an intergroup study of the Gynecologic Oncology Group, Southwestern Oncology Group and Eastern Cooperative Oncology Group. J Clin Oncol 2001; 19: 1001–1007.
19. Amstrong DK, Bundy BN, Baergen R et al. Randomized phase III study of intravenous (IV) paclitaxel and cisplatin versus IV paclitaxel, intraperitoneal (IP) cisplatin and IP paclitaxel in optimal stage III epithelial ovarian cancer (OC): a Gynecologic Oncology Group trial (GOG 172). Proc Am Soc Clin Oncol 2002; 21: 201a (Abstr 803).
20. Piccart MJ, Floquet A, Scarfone G et al. Intraperitoneal cisplatin versus no further treatment: 8-year results of EORTC 55875, a randomized phase III study in ovarian cancer patients with a pathologically complete remission after platinum-based intravenous chemotherapy. Int J Gynecol Cancer 2003; 13 (Suppl 2): 196–203.
21. Bookman MA, Darcy KM, Clarke-Pearson D et al. Evaluation of monoclonal humanized anti-HER2 antibody, trastuzumab, in patients with recurrent or refractory ovarian or primary peritoneal carcinoma with overexpression of HER2: a phase II trial of the Gynecologic Oncology Group. J Clin Oncol 2003; 21: 283–290.
22. Royston P, Parmar MKB, Qian W. Novel designs for multi-arm clinical trials with survival outcomes with an application in ovarian cancer. Stat Med 2003; 22: 2239–2256.
23. UK Coordinating Committee on Cancer Research (UKCCCR) Ductal Carcinoma In Situ (DCIS) Working Party. Radiotherapy and tamoxifen in women with completely excised ductal carcinoma in situ of the breast in the UK, Australia and New Zealand: randomized controlled trial. Lancet 2003; 362: 95–102.
24. Law MG, Emery S. Selective exclusion of treatment arms in multi-arm randomized clinical trials. Stat Med 2003; 22: 19–30.
25. Hochberg Y, Tamhane AC. Multiple Comparison Procedures. New York: John Wiley and Sons 1987.
26. Dunnett CW, Tamhane AC. A step up multiple test procedure. J Am Stat Assoc 1992; 87: 162–170.
27. Proschan MA, Follman DA. Multiple comparisons with control in a single experiment: Why do we feel differently? Am Stat 1995; 49: 144–149.
28. Fleming TR, DeMets DL. Surrogate endpoints in clinical trials: Are we being misled? Ann Int Med 1994; 125: 605–613.

29. Therasse P, Arbuck SG, Eisenhauer EA et al. New guidelines to evaluate the response to treatment in solid tumors. J Natl Cancer Inst 2000; 92: 205–216.
30. Thiesse P, Olivier L, Di Stefano-Louineau D et al. Response rate accuracy in oncology trials: reasons for interobserver variability. J Clin Oncol 1997; 15: 3507–3514.
31. Rustin GJ. Use of CA125 to assess response to new agents in ovarian cancer trials. J Clin Oncol 2003; 21 (Suppl 10): 187–193.
32. Rustin GJS, Quinn M, Thigpen T et al. New guidelines to evaluate the response to treatment in solid tumors (ovarian cancer). J Natl Cancer Inst 2004; 96: 487–488.
33. Guastalla J, Vincent P, Le Rol A et al. CA-125 evaluation of chemotherapy response in patients with recurrent ovarian cancer (ROC). Proc Am Soc Clin Oncol 2002; 21: 204a (Abstr 815).
34. Lakusta CM, Atkinson MJ, Robinson JW et al. Quality of life in ovarian cancer patients receiving chemotherapy. Gynecol Oncol 2001; 81: 490–495.
35. du Bois A, Luck HJ, Meier W et al. A randomized clinical trial of cisplatin/paclitaxel versus carboplatin/paclitaxel as first-line treatment of ovarian cancer. J Natl Cancer Inst 2003; 95: 1320–1329.
36. Vergote I, Rustin GJS, Eisenhauer EA et al. New guidelines to evaluate the response to treatment in solid tumors (ovarian cancer). J Natl Cancer Inst 2000; 92: 1543–1545.
37. Nabholtz JM, Vannetzel JM, Llory JF, Boufette P. Advances in the use of taxanes in the adjuvant therapy of breast cancer. Clin Breast Cancer 2003; 4: 187–192.
38. Piccart MJ, Bertelsen K, James K et al. Randomized trial of cisplatin–paclitaxel versus cisplatin–cyclophosphamide in women with advanced epithelial ovarian cancer: three-year results. J Natl Cancer Inst 2000; 92: 699–708.
39. International Collaborative Ovarian Neoplasm (ICON) Group. Paclitaxel plus carboplatin versus standard chemotherapy with either single-agent carboplatin or cyclophosphamide, doxorubicin, and cisplatin in women with ovarian cancer: the ICON3 randomised trial. Lancet 2002; 360: 505–515.
40. du Bois A, Combe J, Rochon C et al. Epirubicin/paclitaxel/carboplatin (TEC) vs paclitaxel/carboplatin (TC) in first-line treatment of ovarian cancer (OC) FIGO stage IIB–IV. An AGO-GINECO Intergroup phase III trial. J Clin Oncol 2004; 22: 450 (Abstr 5007).
41. De Placido S, Scambia G, Di Vagno G et al. Topotecan compared with no therapy after response to surgery and carboplatin/paclitaxel in patients with ovarian cancer: Multicenter Italian Trials in Ovarian cancer (MITO-1) randomized study. J Clin Oncol 2004; 22: 2635–2642.
42. Buyse M, Molenberghs G, Burzykowski T et al. The validation of surrogate endpoints in meta-analyses of randomized experiments. Biostatistics 2000; 1: 49–67.
43. Ovarian Cancer Meta-analysis Project. Cyclophosphamide plus cisplatin versus cyclophosphamide, doxorubicin, and cisplatin chemotherapy of ovarian carcinoma: a meta-analysis. J Clin Oncol 1991; 9: 1668–1674.
44. Eisenhauer EA, ten Bokkel Huinink WW, Swenerton KD et al. European–Canadian randomized trial of paclitaxel in relapsed ovarian cancer: high-dose versus low-dose and long versus short infusion. J Clin Oncol 1994; 12: 2654–2666.
45. Pfisterer J, Plante M, Vergote I et al. Gemcitabine/carboplatin (GC) vs carboplatin (C) in platinum sensitive recurrent ovarian cancer (OVCA). Results of a Gynecologic Cancer Intergroup randomized phase III trial of the AGO OVAR, the NCIC CTG and the EORTC GCG. J Clin Oncol 2004; 22: 450 (Abstr 5005).
46. du Bois A, Meier W, Luck HJ et al. Chemotherapy versus hormonal treatment in platinum- and paclitaxel-refractory ovarian cancer: a randomized trial of the German Arbeitsgemeinschaft Gynaekologische Onkologie (AGO) Study Group Ovarian Cancer. Ann Oncol 2002; 13: 251–257.
47. ten Bokkel Huinink W, Gore M, Carmichael J et al. Topotecan versus paclitaxel for the treatment of recurrent epithelial ovarian cancer. J Clin Oncol 1997; 15: 2183–2193.
48. Markman M, Liu PY, Wilczynski S et al. Phase III randomized trial of 12 versus 3 months of maintenance paclitaxel in patients with advanced ovarian cancer after complete response to platinum and paclitaxel-based chemotherapy: a Southwest Oncology Group and Gynecologic Oncology Group trial. J Clin Oncol 2003; 21: 2460–2465.
49. Cannistra S. The eithics of early stopping rules: who is protecting whom? J Clin Oncol 2004; 22: 1542–1545.
50. Perulekar WR, Eisenhauer EA. Phase I trial design for solid tumor studies of targeted, non-cytotoxic agents: theory and practice. J Natl Cancer Inst 2004; 96: 990–997.
51. Hall GD, Brown JM, Coleman RE et al. Maintenance treatment with interferon for advanced ovarian cancer: results of the Northern and Yorkshire gynaecology group randomised phase III study. Br J Cancer 2004; 91: 621–626.
52. Biagi JJ, Eisenhauer EA. Systemic treatment policies in ovarian cancer: the next 10 years. Int J Gynecol Cancer 2003; 13 (Suppl 2): 231–240.