



Published in final edited form as:

Int J Gynecol Cancer. 2012 January ; 22(1): 63–69. doi:10.1097/IGC.0b013e31822e29aa.

Identifying Clinical Improvement in Consolidation Single-arm Phase II Trials in Ovarian Cancer Patients in Second or Greater Clinical Remission

Alexia Iasonos, PhD¹, Paul Sabbatini, MD², David R. Spriggs, MD², Carol A. Aghajanian, MD², Roisin E. O’Cearbhaill, MD², Martee L. Hensley, MD², and Howard T. Thaler, PhD¹

¹Department of Epidemiology and Biostatistics, Memorial Sloan Kettering Cancer Center, New York, NY 10065

²Gynecologic Medical Oncology Service, Department of Medicine, Memorial Sloan Kettering Cancer Center, New York, NY 10065

Abstract

OBJECTIVE—Estimates of Progression-Free Survival (PFS) from single-arm Phase II consolidation/maintenance trials for recurrent ovarian cancer are usually interpreted in the context of historical controls. We illustrate how the duration of second-line therapy (SLT), the time on the investigational therapy (IT) and patient enrollment plan can affect efficacy measures from maintenance trials and might result in underpowered studies.

METHODS—Efficacy data from three published single-arm consolidation therapies in second remission in ovarian cancer were used for illustration. The studies were designed to show an increase in estimated median PFS from 9 to 13.5 months. We partitioned PFS as the sum of the duration of SLT, treatment-free interval (TFI), and duration of IT. We calculated the statistical power when IT is given concurrently with SLT or following SLT by varying the start of IT. We compared the sample sizes required when PFS includes the time on SLT vs PFS that starts following SLT at initiation of IT.

RESULTS—Required sample sizes varied with duration of SLT. If IT starts with initiation of SLT, only 34 patients are needed to provide 80% power to detect a 33% hazard reduction. In contrast, 104 patients are required for a single arm study for 80% power, if IT begins 7.5 months after SLT initiation.

CONCLUSIONS—Designs of non-randomized consolidation trials that aim to prolong PFS must consider the effect of the duration of SLT on the endpoint definition and on required sample size. If IT is given concurrently with SLT, and following SLT, then SLT duration must be restricted per protocol eligibility, so that a comparison with historical data from other single-arm Phase II studies is unbiased. If IT is given following SLT, duration of SLT should be taken into account in the design stage since it will affect statistical power and sample size.

Keywords

maintenance; consolidation; ovarian cancer; second line chemotherapy; endpoint; design

INTRODUCTION

While over 80% of patients with advanced-stage ovarian cancer (OC) will demonstrate a clinical response to first-line platinum-based chemotherapy, the majority will recur and ultimately succumb to their disease, with 5-year overall survival ranging from 5 – 30% (1,2). Significant effort has been dedicated to avoid subsequent recurrences following primary therapy in patients who relapse and return to remission, including continuation of second line therapy in the form of either consolidation or maintenance therapy (3, 4, 5). Traditionally, consolidation therapy refers to short term strategies using the same or a different treatment in order to consolidate the response to therapy. Maintenance therapy generally refers to using the same treatment over a longer time period continued until progression rather than for a fixed time period. (6) The aim of either approach is to prolong the disease free period. Since patients are typically in clinical remission (CR) at the start of consolidation/maintenance therapy, Progression-Free Survival (PFS) is the primary endpoint used in consolidation studies. (7) However, guidance for defining clinical improvement for patients in remission trials remains sparse. (8, 9) In particular, how to measure the effect of the consolidation therapy independent of the effect of the therapy that has achieved the CR in the setting of non-randomized trials in patients who are in second or greater clinical remission?

The second remission population is ideal to evaluate consolidation strategies since nearly all patients will have disease progression over a short time period of 9–15 months. (7) The different disease states in recurrent ovarian cancer patients who are receiving consolidation therapy after second remission have been described previously. (10) Historical estimates of PFS for this population from phase II consolidation trials often include variable duration of second-line therapy (SLT), possibly a treatment free interval (TFI) and the time on investigational therapy (IT). For example, common practice does not distinguish a consolidation strategy that consists of 10 months on SLT followed by 3 months on IT, versus a strategy that consists of 3 months on SLT followed by 10 months on IT. If both strategies have a median PFS of 13 months, does either strategy warrant a phase III trial? Was the investigational therapy received long enough to allow effect?

Furthermore, the patient populations across phase II consolidation trials in second line non-randomized setting have not been selected in a consistent way in terms of previous line of treatments, and whether they are enrolled and start IT along with SLT at the time of primary recurrence or after they have achieved complete response. Furthermore, some trials enroll patients in strict second CR or greater, while other trials enroll patients who have achieved a partial response (PR) or stable disease (SD). The heterogeneity in patient population and lack of randomization in single arm studies make it difficult to compare consolidation regimens on the basis of PFS since the results of one strategy might not be generalizable to another study. The key question is how to decide if a single-arm consolidation trial shows enough promise to move forward to a randomized Phase III trial.

We hypothesize that the design and analysis of consolidation trials should take into account the starting point of consolidation therapy, with regard to SLT, in order to be able to identify a promising PFS duration for further randomized study. The primary focus of this paper is to consider consolidation strategy designs in patients in second or greater CR. We define clinical improvement in a non randomized setting in a consistent way so that comparisons with historical estimates are valid and decisions whether a single-arm study is promising and worthy of further study are reliable. We discuss eligibility criteria so that the appropriate patient population is enrolled consistently in future trials.

METHODS

For the historical estimates we use median PFS from published single-arm consolidation studies in patients in second or subsequent complete clinical remission in ovarian cancer^(11–13). We included three cohorts of 35 patients each: two prospective consolidation clinical trials and one untreated population in their second remission who was followed-up for observation (13). The Phase II trials evaluated the efficacy of imatinib and the combination of goserelin with bicalutamide with median PFS of 12.1 (11) and 11.8 months (12) respectively. Eligibility criteria and criteria for response were consistent in all three cohorts. Details regarding the combined analysis of patients who were in second or subsequent remission have been described previously (10).

We calculated the sample sizes required to have 80% power in order to show an increase in estimated median PFS from 9 to 13.5 months, which corresponds to a 33% hazard reduction, using two different starting points for PFS definition. We partitioned PFS in three intervals by expressing it as the sum of the duration of SLT, TFI and IT (i.e., $PFS=SLT+TFI+IT$, Figure 1) and calculated the sample size under different values of SLT, TFI and starting point of IT. We assumed that PFS follows exponential distribution from the start of SLT with a drop in hazard at the start of IT, that the magnitude of the drop does not depend on starting time, and that patients are followed until failure. All tests are one-sided, single arm comparisons at $P \leq 0.05$. Details regarding the calculations can be found in the appendix.

We evaluated two endpoints: PFS from SLT was defined as the time from the start of SLT to disease progression or death (the traditional definition); PFS from IT was defined as the time from the start of IT to disease progression or death. Power calculations used intent-to-treat paradigm by including all eligible patients at the start of SLT for the first endpoint. However, for the calculation of PFS from IT, patients must be in clinical remission at the time of IT, i.e., patients who progress prior to initiation of IT are excluded. The assumptions for the respective designs using the above endpoints are summarized in Table 1. We also evaluated power and sample size requirements for three treatment strategies, as shown in Figure 2. Consider the start and end dates of second-line therapy (SLT) as points A and B respectively, and the start of protocol/investigational therapy as point C.

- *Strategy 1* enrolls patients and starts IT concurrently with the start of SLT; patients in CR at the end of SLT continue on the IT therapy.
- *Strategy 2* enrolls patients who are in CR immediately at the end of SLT and thus IT is given sequentially without any delay.
- *Strategy 3* allows for a treatment-free interval between the two treatments, which is a more realistic representation of clinical practice. That is, patients enroll in a consolidation trial after the completion of SLT, but a variable period of delay exists while the patient is screened and begins IT.

The effect of these strategies on sample size can be described by varying the duration of SLT and the starting point of IT, relative to SLT, as follows: a starting point for IT of zero is equivalent to starting IT concurrently with SLT (strategy 1). Starting IT immediately after the end of SLT, assuming SLT is given for 6 cycles every 3 weeks, implies a starting point for IT of 4.5 months (strategy 2). A starting point of 6 or 7.5 months allows for a TFI interval of 6 and 12 weeks, respectively, after the end of SLT (strategy 3).

RESULTS

Based on the completed single-arm consolidation trials, the median duration of SLT for the combined population of the three cohorts that received treatment was 4.5 mos (IQR 3.6–

5.9), median TFI: 2.5 months, and the duration of IT varied from 4 to 7.5 months. For the hypothetical data, we used the same parameters as the ones obtained in our completed consolidation trials, but we varied the duration on SLT and starting point of IT, and calculated corresponding power estimates. Figure 3 shows the Kaplan Meier estimates of PFS of five simulated trials, compared with a survival curve with the historical median PFS of 9 months. All trials were simulated to have an increase in median PFS from 9 to 13.5 months, and samples of 34 or 100 patients. When IT is given concurrently with SLT (i.e., start time for IT is 0), the trials have higher PFS estimates compared with the historical estimate regardless of the sample size. However, as the initiation of IT delays to 6 or 7.5 months after SLT, then the curves may cross the historical estimate in a 34 patient study, and statistically non-significant results are likely, unless the sample size increases to 100 patients.

Table 2 shows sample sizes required to achieve 80% power, and a 33% hazard improvement from a historical estimate of PFS of 9 months at varying starting points for IT. For example, a study of 34 patients will provide 80% power only if the IT starts concurrently with SLT, after recurrence from primary treatment. If IT starts 4.5 months after SLT, which corresponds to 6 cycles of chemotherapy every 3 weeks, then 67 patients are required for a single-arm trial. If IT starts after completing 6 months on SLT, 84 patients are required. If one accrues only 34 patients at start of SLT with planned 33% hazard reduction and the duration of SLT delays the start of IT to 7.5 months, the power drops to 50% (Figure 1. Supplemental Material). To maintain 80% power in this scenario, either the IT would have to reduce the hazard by 50%, or the sample size would have to increase to approximately 104 events. Note that only 81 out of 104 patients will receive IT, since some patients would progress before 7.5 months. This calculation uses the traditional definition of PFS (i.e., start of chemotherapy to progression). The advantage of this design is that the results are generalizable to the patient population observed right after first progression, and historical estimates of PFS are available since this design uses the traditional definition of PFS. However, all patients must be followed up from initiation of SLT, although a smaller number of patients will respond and receive the IT, hence longer follow-up and more resources are required.

Using Design 2 and the respective endpoint, the starting point of PFS is the start of IT, regardless when IT starts. Assuming the effect size is 33% in hazard reduction and seeking 80% power, the sample size needed at the start of the IT is 34 patients. The later the investigational therapy begins, the more patients need to be screened since patients might become ineligible due to progression prior to initiation of IT (Table 3). For example, if IT starts 7.5 months after the start of SLT, then we expect 10 ineligible patients due to progression before 7.5 months; thus, 44 patients need to be screened in order to enroll 34 eligible patients at 7.5 months.

The advantages of Design 2 are as follows: 1) the increase in sample size is minimal compared to using PFS from start of SLT since the power of the study is not affected by SLT duration, which occurred before initiation of IT; 2) The PFS endpoint starting from initiation of IT focuses only on the time period during which patients are benefiting by the IT. However, one of the major limitations of using this endpoint is the lack of historical estimates, since it uses a non-traditional definition of PFS which does not include the chemotherapy treatment interval. Moreover, PFS estimates may not be generalizable to the patient population after first progression since eligibility criteria at the start of IT require patients in CR.

DISCUSSION

We demonstrated that consolidation trials in second line non-randomized setting, designed to show an improvement in PFS over a historical control, can be underpowered for the primary endpoint or can provide biased estimates which cannot be compared with results from other studies. The reason that a single arm consolidation trial might be underpowered, is that estimates of efficacy such as PFS include the duration of second line therapy which dilutes the effect of the investigational treatment. We showed that the study power is affected by the duration of second-line therapy and starting time of IT both of which can vary widely in practice. The longer the time on SLT, the larger the sample size or the greater the clinical benefit must be to show improvement. We recommend that designs of consolidation trials take into account the duration of SLT, by either defining PFS from the start of IT or restricting SLT duration per protocol. This is not a purely statistical decision, since both approaches raise clinical and logistical issues.

It is acknowledged that the question of whether IT is efficacious can be best answered in a Phase III definitive trial of comparing two randomized arms, namely SLT alone and SLT with consolidation therapy added, ie SLT+IT. If power is reduced, randomized Phase II trials would also provide a head to head comparison with a concurrent control and the lack of historical estimates would be eliminated (14). However, in order to design randomized studies we need meaningful PFS estimates for the control arm and the expected improvement in which to base the sample size required. These estimates are always based on smaller Phase II trials. Furthermore, when a larger randomized Phase II trial is not feasible, single-arm consolidation trials remain a viable option in identifying agents with activity before committing to move into a larger confirmatory trial.

Our focus has been second line therapy, but the question of what is considered a clinically meaningful improvement and when PFS should start applies to consolidation trials in other lines of treatments. In primary therapies these issues are less critical because the duration of first line therapy is typically uniform, averaging from 6–8 cycles whereas the duration of second line therapy can be more variable. For example the Phase III trial known as SWOG S9761/GOG 178 (15) in which advanced stage OC patients with complete response to platinum/taxane therapy were randomized to receive either 3 or 12 cycles of monthly paclitaxel showed a significant improvement in PFS favoring 12 cycles (median PFS 22 vs 14 months; p value=0.006) when PFS was measured from the start of first line therapy and front line therapy was restricted to 5–6 cycles. On the other hand, the Oregovomab trial (16) which randomized advanced OC patients to maintenance immunotherapy or placebo after 4 to 12 weeks of front line therapy showed no improvement with median PFS of 10.3 (oregovomab) vs 12.9 (placebo) p value=0.2, when PFS was measured from randomization 4 to 12 weeks *after* the end of front line therapy. The estimate from GOG 178 includes the time of front line therapy, while it correctly restricts it per protocol, while the Oregovomab trial excludes the time of front line therapy by starting PFS at randomization and allowing a TFI of 4 to 12 weeks prior to randomization. While different approaches of reporting PFS are used here, the results may be compared because the duration of primary therapy is relatively consistent. However, in non randomized consolidation trials in the setting of second line treatment, the starting point of PFS is not uniformly defined and duration of SLT is not restricted and can be variable. This limits the ability to compare different studies.

In order to minimize this variability, we propose eligibility restrictions for non-randomized trials evaluating agents in the consolidation, second line setting. One approach would be to restrict the time on SLT and the TFI. The duration of SLT cannot be absolutely restricted as patients may achieve CR at variable time points, but we suggest a design allowing 5–6 cycles of SLT. In addition, if starting IT after SLT, the TFI should be similarly restricted and

allow a TFI of up to 2 months from the completion of SLT to the start of IT. If these restrictions are not feasible, another approach would be to exclude SLT from the definition of PFS by calculating PFS from the start of IT, and we have shown that the benefits in terms of sample size and resources are clear in this setting. However, comparisons with historical data must be cautious. When PFS is calculated from the start of SLT, the estimates are valid for all patients enrolled after primary recurrence. When the duration of SLT is excluded from PFS definition, the estimates are less prone to bias since they measure the efficacy of the investigational treatment alone, but these estimates are valid only to patients who have achieved CR after completion of SLT and the literature is less robust in this regard.

Our study addresses the effect of the duration of SLT on the final PFS estimates under specific assumptions. Our sample size and power calculations considered a specific difference in median PFS based on our experience and the estimates reported in the literature. We assumed PFS follows exponential distribution and the hazard is constant within each treatment interval. While this assumption may not be justified when analyzing real data, it has appeal for sample size calculations due to its interpretability and simplicity and it is typically used (17). Power estimates may differ under other distributions, and such evaluation is beyond the scope of this paper. However, our conclusions about the importance of defining the starting times for IT and PFS apply in general.

We evaluated various treatment strategies and endpoints currently used in consolidation trials and examined the effect of duration of second-line therapy on power and sample size requirements. The appropriate selection of patient population and the endpoint to be examined are the two major challenges in the design of consolidation trials so that comparisons with historical estimates are valid. We recommend that the individual intervals, namely, time on second-line therapy, treatment-free interval, and time on investigational therapy, be reported in future trials so that historical estimates can be obtained and used in the design of single-arm consolidation trials. An informative, unbiased comparison with results of other single-arm Phase II studies will depend on increased uniformity of SLT.

Supplementary Material

Refer to Web version on PubMed Central for supplementary material.

Acknowledgments

Research Support: Grant Support: CA138738-01, PO1 CA052477 (D.R. Spriggs)

Abbreviations

CR	clinical remission
IT	investigational therapy
OC	Ovarian cancer
PD	progressive disease
SLT	second-line therapy
TFI	Treatment-free interval
PFS	Progression free interval

References

1. Bonnefoi H, A'Hern RP, Fisher C, Macfarlane V, Barton D, Blake P, Shepherd JH, Gore ME. Natural history of stage IV epithelial ovarian cancer. *J Clin Oncol*. 1999; 17:767–775. [PubMed: 10071265]
2. McGuire WP, Hoskins WJ, Brady MF, et al. Cyclophosphamide and cisplatin compared with paclitaxel and cisplatin in patients with stage III and stage IV ovarian cancer. *N Engl J Med*. 1996; 334:1–6. [PubMed: 7494563]
3. Dearnley DD, McMeekin DS. Consolidation therapy in ovarian cancer: where do we stand? *Curr Opin Obstet Gynecol*. 2006; 18:3–7. [PubMed: 16493252]
4. Gadducci A, Cosio S, Conte PF, Genazzani AR. Consolidation and maintenance treatments for patients with advanced epithelial ovarian cancer in complete response after first-line chemotherapy: a review of the literature. *Crit Rev Oncol Hematol*. 2005; 55:153–66. [PubMed: 15890524]
5. Foster T, Brown TM, Chang J, et al. A review of the current evidence for maintenance therapy in ovarian cancer. *Gynecol Oncol*. 2009; 115:290–301. [PubMed: 19717182]
6. Sabbatini P, Spriggs DR. Consolidation for ovarian cancer in remission. *J Clin Oncol*. 2006; 24:537–539. [PubMed: 16446325]
7. Sabbatini P. Consolidation therapy in ovarian cancer: a clinical update. *Int J Gynecol Cancer*. 2009; 19(Suppl 2):S35–S39. [PubMed: 19955912]
8. Bast RC, Thigpen JT, Arbuck SG, et al. Clinical trial endpoints in ovarian cancer: report of an FDA/ASCO/AACR Public Workshop. *Gynecol Oncol*. 2007; 107:173–176. [PubMed: 17950384]
9. Markman M, Markman J, Webster K, et al. Duration of response to second-line, platinum-based chemotherapy for ovarian cancer: implications for patient management and clinical trial design. *J Clin Oncol*. 2004; 22:3120–3125. [PubMed: 15284263]
10. Sabbatini P, Spriggs D, Aghajanian C, et al. Consolidation strategies in ovarian cancer: observations for future clinical trials. *Gynecol Oncol*. 2010; 116:66–71. [PubMed: 19836827]
11. Juretzka M, Hensley ML, Tew WA, et al. A phase 2 trial of oral imatinib in patients with epithelial ovarian, fallopian tube, or peritoneal carcinoma in second or greater remission. *Eur J Gynaecol Oncol*. 2008; 29:568–572. [PubMed: 19115680]
12. Levine D, Park K, Juretzka M, et al. A phase II evaluation of goserelin and bicalutamide in patients with ovarian cancer in second or higher complete clinical disease remission. *Cancer*. 2007; 110:2448–2456. [PubMed: 17918264]
13. Harrison ML, Gore ME, Spriggs D, et al. Duration of second or greater complete clinical remission in ovarian cancer: exploring potential endpoints for clinical trials. *Gynecol Oncol*. 2007; 106:469–475. [PubMed: 17614127]
14. Rubinstein LV, Korn EL, Freidlin B, Hunsberger S, Ivy SP, Smith MA. Design issues of randomized phase II trials and a proposal for phase II screening trials. *J Clin Oncol*. 2005; 23:7199–206. [PubMed: 16192604]
15. Markman M, Liu PY, Moon J, et al. Impact on survival of 12 versus 3 monthly cycles of paclitaxel (175 mg/m²) administered to patients with advanced ovarian cancer who attained a complete response to primary platinum-paclitaxel: follow-up of a Southwest Oncology Group and Gynecologic Oncology Group phase 3 trial. *Gynecol Oncol*. 2009; 14(2):195–198. [PubMed: 19447479]
16. Berek J, Taylor P, McGuire W, et al. Oregovomab maintenance monoimmunotherapy does not improve outcomes in advanced ovarian cancer. *Journal of Clinical Oncology*. 2009; 27:418–425. [PubMed: 19075271]
17. Broglio A, Berry D. A Detecting an overall survival benefit that is driven from progression free survival. *JNCI*. 2009; 101:1642–1649. [PubMed: 19903805]

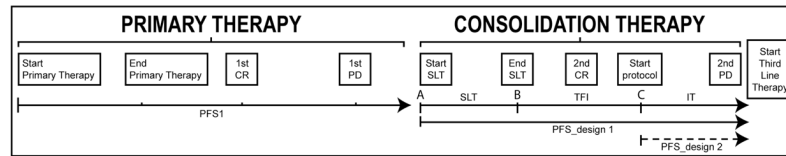
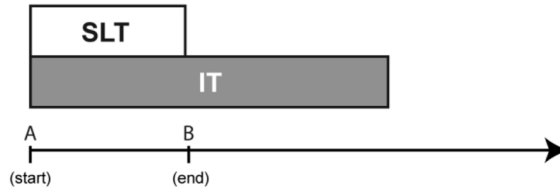


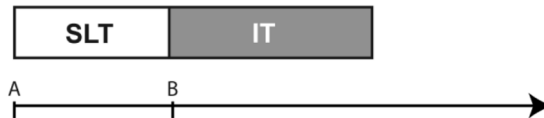
Figure 1. Consolidation treatment in Ovarian Cancer received in Second Line Setting following primary recurrence: definition of treatment intervals. Abbreviations: CR (complete response); PD (progressive disease); SLT (second-line therapy); TFI (Treatment-free interval); IT (investigational therapy).

STRATEGY 1:

SLT and IT are given concurrently

**STRATEGY 2:**

IT is given immediately after the end of SLT

**STRATEGY 3:**

IT is given after a treatment free interval (TFI) has elapsed after the end of SLT

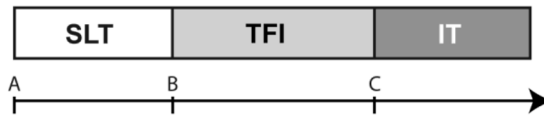
**Figure 2.**

Illustration of the starting point of investigational therapy (IT) in relation to second-line therapy (SLT) under three potential treatment strategies. Abbreviations: TFI: Treatment-free interval.

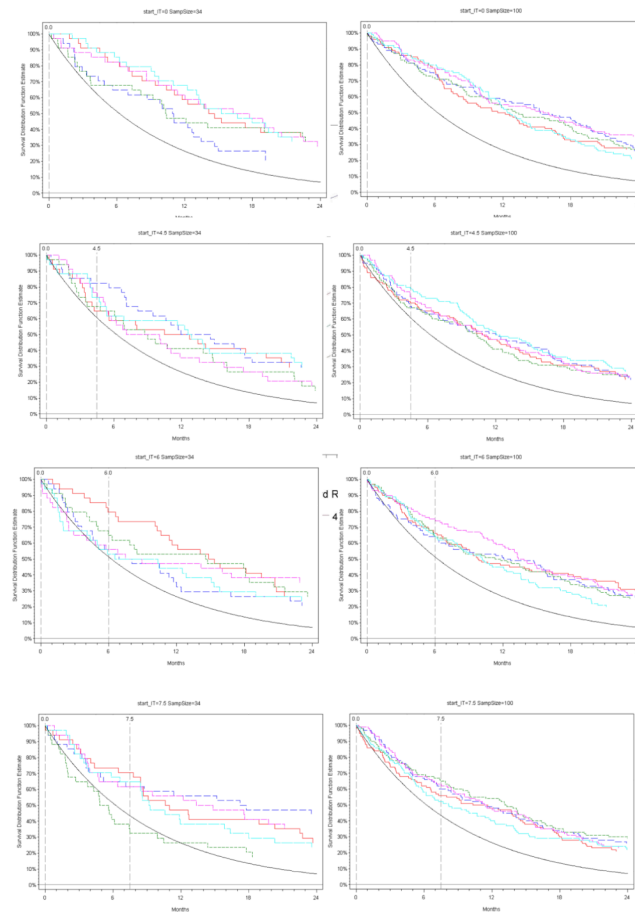


Figure 3. Simulated Kaplan Meier estimates of PFS relative to a reference survival curve with median PFS of 9 months (solid black line), with hazard reduction of 33.3% and sample sizes of 34 (left panel) or 100 (right panel) patients respectively, at varying times of initiating investigational therapy (IT).

Table 1

Assumptions for designs and respective endpoints

Design 1	Design 2
Endpoint PFS starts at the start of SLT. PFS includes SLT, TFI, and IT.	Endpoint PFS starts at the start of IT; it includes time on protocol/ investigational therapy (IT) only.
Null Hypothesis: PFS follows exponential distribution with median of 9 months	
Alternative Hypothesis: With added treatment at the initiation of investigational therapy, there is a new exponential distribution with a larger median (ie lower but constant hazard).	Alternative Hypothesis: An increase in median PFS from 9 to 13.5 months is equivalent to 33% hazard reduction.
Power calculations are based on intent to treat analysis since PFS starts at the start of SLT;. It includes all eligible patients at the start of SLT, although patients who progress before initiation of IT would not receive IT.	Power calculations are based on conditional analysis at the initiation of IT. The eligibility criteria at the start of IT include patients who are in CR/PR/SD (ie exclude pts who had early PD).

Table 2

Sample size required under PFS endpoint where PFS is calculated from the start of second line therapy (equivalent to Design 1)

Time of initiating investigational therapy (IT)	33% Hazard Reduction Sample Size (Number of patients entered at the start of SLT)	Average number of patients treated at the end of SLT
0 mos (at start SLT)	34	34
4.5 mos	67	58
6 mos	84	69
7.5 mos	104	81

Table 3

Sample size required under PFS endpoint where PFS is calculated from the start of investigational therapy (equivalent to Design 2)

Time of initiating investigational therapy (IT)	33% Hazard Reduction Sample Size (Number of patients entered at the start of IT)	Average number of patients ineligible at the start of IT (number of patients to be screened prior to initiation of IT)
0 mos (at start SLT)	34	0 (34)
4.5 mos	34	6 (40)
6 mos	34	8 (42)
7.5 mos	34	10 (44)