ORIGINAL ARTICLE

An adaptive population enrichment phase III trial of TRC105 and pazopanib versus pazopanib alone in patients with advanced angiosarcoma (TAPPAS trial)

C. R. Mehta[1,2*], L. Liu[1] & C. Theuer[3]

[1]Department of Biostatistics, Cytel Inc, Cambridge; [2]Department of Biostatistics, Harvard TH Chan School of Public Health, Boston; [3]Clinical Department, TRACON Pharmaceuticals, San Diego, USA

Correspondence to: Dr Cyrus R. Mehta, Department of Biostatistics, Cytel Inc., 675 Massachusetts Avenue, Cambridge, MA 02139, USA. Tel: +1-617-448-3849;
E-mail: mehta@cytel.com
Note: This study was previously presented as a poster at ASCO-2017, 2018.

Background: Major challenges in clinical trials of ultra-orphan oncology diseases include limited patient availability and paucity of reliable prior data for estimating the treatment effect and, therefore, determining optimal sample size. Angiosarcoma (AS), a particularly aggressive form of soft tissue sarcoma with an incidence of about 2000 cases per year in the United States and Europe is poorly addressed by current systemic therapies. Pazopanib, an inhibitor of vascular endothelial growth factor receptor (VEGFR) is approved for the treatment of AS, with modest benefit. TRC105 (carotuximab) is a monoclonal antibody to endoglin, an essential angiogenic target highly expressed on proliferating endothelium and both tumor vessels and tumor cells in AS, that has the potential to complement VEGFR tyrosine kinase inhibitors. In a phase I/II study of soft tissue sarcoma, TRC105 combined safely with pazopanib and the combination demonstrated durable complete responses and encouraging progression-free survival (PFS). In addition, there was a suggestion of superior benefit in patients with cutaneous lesions versus those with the non-cutaneous lesions.

Patients and methods: This article describes the design of a recently initiated phase III trial of TRC105 And Pazopanib versus Pazopanib alone in patients with advanced AngioSarcoma (TAPPAS trial). Given the ultra-orphan status of the disease and the paucity of reliable prior data on PFS or overall survival (end points required for regulatory approval as a pivotal trial), an adaptive design incorporating population enrichment and sample size re-estimation was implemented. The design incorporated regulatory input from the Food and Drug Administration (FDA) and European Medicines Agency and proceeded following special protocol assessment designation by the FDA.

Conclusions: It is shown that the benefit of the adaptive design as compared with a conventional single-look design arises from the learning and subsequent improvements in power that occur after an unblinded analysis of interim data.

Registered on Clinicaltrials.gov: NCT02979899.

Key words: population enrichment, adaptive design, angiosarcoma, sample size re-estimation, promising zone

Introduction

One dilemma of clinical development is whether to open enrollment to all patients regardless of biomarker status or restrict enrollment to a subgroup carrying a specific biomarker. Restricting enrollment to the targeted subgroup without sufficient empirical evidence of lack of efficacy in the non-targeted subgroup may deny a large segment of the population access to a potentially beneficial treatment. In contrast, running a large trial in a heterogeneous population may place many patients for whom the drug is ineffective at unnecessary risk and dilute the treatment effect. Adaptive clinical trial designs provide an opportunity to improve the manner in which new agents having empirical or biologic plausibility of benefiting specific tumor types within the same disease class are tested. Specifically one can refine at an interim analysis of the ongoing trial the right population and the right sample size to take

forward, thereby improving the probability of a successful outcome. This paper describes just such a design for an ongoing study of angiosarcoma (AS), an ultra-orphan tumor-type and aggressive form of soft tissue sarcoma with an incidence of approximately 1000 patients per year in the United States and a similar number in Europe.

Standard regimens for unresectable advanced AS include chemotherapy (taxanes, anthracyclines, and gemcitabine) as well as pazopanib, an inhibitor of multiple receptor tyrosine kinases including vascular endothelial growth factor receptors (VEGFR). Tumor control of metastatic disease with these therapies is short-lived with median progression-free survival (PFS) ranging from 3.0 to 6.6 months, and median overall survival (OS) of approximately 8–11 months [1]. The short time to progression following current treatment of metastatic disease emphasizes that AS is a disease in need of more effective and well-tolerated treatment options. TRC105 (carotuximab) is a monoclonal antibody to endoglin (CD105), an essential angiogenic target densely expressed on proliferating endothelium and tumor cells in AS that is distinct from the VEGFR [2]. Endoglin is up-regulated following VEGFR inhibition (VEGFRI) through hypoxia inducible factor alpha expression, and TRC105 inhibits angiogenesis, tumor growth, and metastases in preclinical models and complements the activity of multi-kinase VEGFRI [3, 4]. In a phase I/II study of soft tissue sarcoma patients, TRC105 was combined safely with pazopanib and the combination demonstrated durable complete responses and median PFS of 7.8 months in AS patients [5]. These data compared favorably with the largest series of AS patients treated with single-agent pazopanib ($n = 40$), which reported no complete responses, median PFS of 3.0 months, and OS of 9.9 months [6]. The benefit in patients with cutaneous lesions appeared to be superior to that of patients with non-cutaneous disease: of 12 chemotherapy refractory and VEGFR inhibitor naive patients, 4 of 6 cutaneous patients had longer time on treatment with TRC105 + pazopanib compared with prior chemotherapy versus 3 of 6 patients with non-cutaneous AS; only VEGFR refractory patients with cutaneous AS benefitted from TRC105 + pazopanib and both durable CRs (lasting over 2 years each) were seen in cutaneous patients.

The randomized, multinational, multicenter, open-label, parallel group, adaptive phase III trial of TRC105 And Pazopanib versus Pazopanib alone in patients with advanced AngioSarcoma (TAPPAS) trial was designed following regulatory discussions with the European Medicines Agency (EMA) and Food and Drug Administration (FDA). Given the ultra-orphan status of the disease and the paucity of reliable prior data on PFS or OS (end points required for regulatory approval as a pivotal trial), an adaptive design with an unblinded interim analysis to permit the following mid-course corrections was implemented:

1. Sample size re-estimation: It was proposed to start out with a small study initially—adequately powered to detect a somewhat larger but still realistic treatment effect—with the option to increase the sample size if deemed necessary.
2. Population enrichment: Since there was some indication of greater tumor sensitivity to TRC105 in the cutaneous subgroup the design included the option to restrict future enrollment to the cutaneous subgroup.

The study was launched in January 2017 and is registered on Clinicaltrials.gov (NCT02979899).

Methods

The primary objective of the TAPPAS trial is to demonstrate superior PFS of the combination of TRC105 and pazopanib (arm B) versus single-agent pazopanib (arm A). Patients are stratified by AS subtype (cutaneous versus non-cutaneous) and the number of lines of prior systemic therapy for AS (none versus 1 or 2) and randomized in a 1 : 1 ratio. Patients are eligible for treatment until disease progression, unacceptable toxicity, withdrawal of consent, or other reasons. Efficacy is assessed by RECIST every 6 weeks following randomization, including computed tomography or magnetic resonance imaging of non-cutaneous sites of disease and standardized photography of cutaneous sites of disease. Survival is assessed on study and every 3 months following study discontinuation.

The adaptive strategy

On the assumption that the median PFS is 4.0 months for arm A, it is reasonable to design the study to detect an improvement to 7.27 months for arm B, which corresponds to a hazard ratio (HR) of 0.55 for arm B to arm A on the exponential scale. These assumed median PFS values are based on the reported experience of pazopanib as a single agent and when combined with TRC105 in single-arm trials of AS patients [6]. However, due to the small sample size of these studies and due to possible differences in treatment effect in the cutaneous and non-cutaneous AS subgroups, the accumulating data will be analyzed by an independent data monitoring committee (DMC) at an interim analysis time point and the future course of the trial may be adapted on the basis of the interim results. The DMC will recommend one of three possible actions: continue as planned with the full population, continue with the full population and an increase in sample size and PFS events, continue with the cutaneous subgroup, thereby enriching the study population. These options are displayed schematically in Figure 1 where the interim result is partitioned into one of four zones – favorable, promising, enriched, or unfavorable. Classification of the interim results into each zone is based on conditional power, which is defined as the probability that the trial will achieve statistical significance at the final analysis given the data already observed at the interim analysis. (Refer to supplementary Section S2, available at *Annals of Oncology* online for a technical definition of conditional power.)

The rationale for the classification and re-design strategy depicted in Figure 1 is as follows. In the favorable zone, the conditional power is sufficiently high so that an increase in resources is not needed. In the promising zone, the conditional power is moderately high and can be elevated to a desirable level by an appropriate increase of sample size and PFS events. The unfavorable zone is characterized by a conditional power so low that an increase in resources cannot be justified for the full population. However, since there is an a priori expectation that arm B might differentially benefit the cutaneous subgroup the conditional power of this subgroup is computed whenever the interim results for the full population fall in the unfavorable zone. If the conditional power of the cutaneous subgroup, *after allowing for an increase of sample size and PFS events up to a pre-specified maximum,* is sufficiently high, the interim results are said to fall in the enrichment zone and the study population is enriched by restricting future enrollment to the cutaneous subgroup. Thus enrichment may be seen as a last resort for improving the chances of a successful trial. Only if the conditional power falls in the unfavorable zone for the full population would we even consider enrichment, and in that case, the conditional power of the cutaneous subgroup must be sufficiently high, possibly with an increase in sample and PFS events, to justify the enrichment. The conditional power cut-points that determine the different zones are shown in Figure 2 along with the corresponding allowable increases in sample size and PFS events. The rationale for these choices is discussed in 'Selecting the design parameters' section. Although not formally part of the TAPPAS design, there is also an informal futility zone in which the DMC has the flexibility to exercise its clinical judgement and terminate the trial for futility.
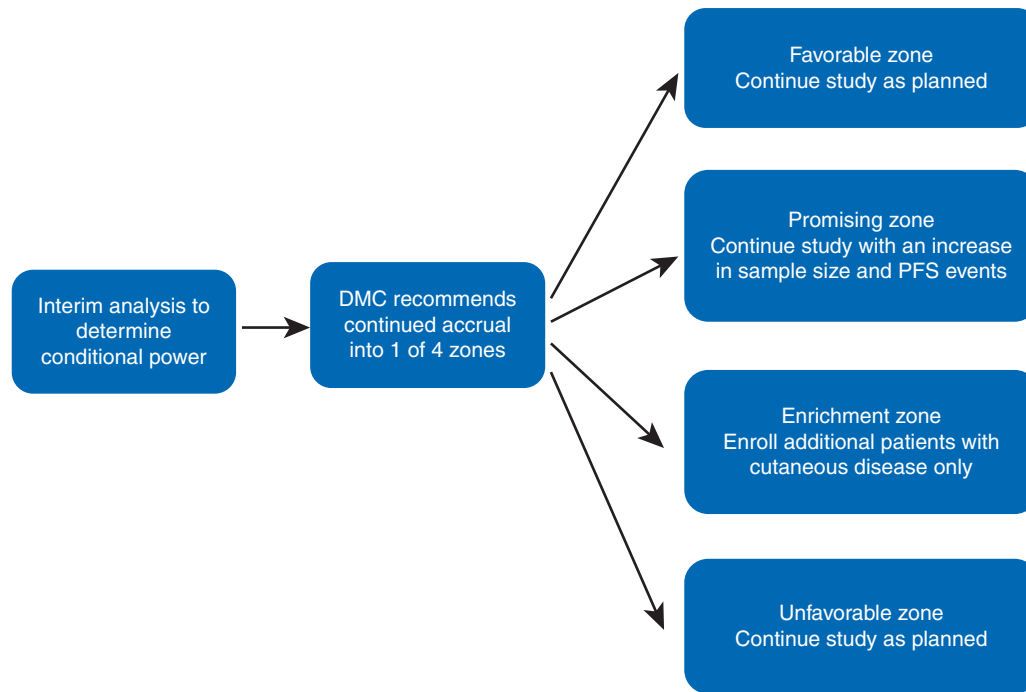
Figure 1. Schematic representation of the adaptive enrichment design.

Special considerations for event-driven enrichment trials

TAPPAS is an event-driven trial. This means that the number of PFS events and not the sample size is the primary driver of power for the study. Sample size does of course play an important role in the sense that the larger the sample size, the sooner the required number of PFS events will arrive. There is thus a trade-off between sample size, study duration and number of PFS events. While this trade-off exists for all event-driven trials, it is especially complex for trials in which there is the possibility of an adaptive enrichment after the interim analysis. The statistical validity of making adaptive changes to an ongoing trial based on an unblinded interim analysis hinges on keeping the dataset that was utilized for the interim analysis independent of the data that will result from the adaptive change. This can be especially problematic for studies with a time to event end point like PFS, because some subjects who are administratively censored in the dataset available at the time of the interim analysis might subsequently experience a PFS event. If one were to avail of data in the patient profiles of these subjects (e.g. their toxicities, lab values, or extent of tumor shrinkage) that might be correlated with PFS to assist in making the enrichment decision, the type-1 error could be inflated [7]. On the other hand, the DMC is charged with examining all available data and utilizing clinical judgement before making a recommendation to the sponsor. Recognizing this dilemma, Jenkins et al. [8] suggested a novel approach that would permit full use of all data available at the time of the interim analysis, including from patients who are censored for PFS. In this approach, the study population is split in two subsequent cohorts whose initial sample sizes and PFS events are pre-specified before any unblinding of the data. Cohort 1 is enrolled first followed by Cohort 2. At an interim analysis, which occurs while Cohort 1 is still accruing the pre-specified number of patients and PFS events, the data are unblinded to an independent DMC which then recommends one of the four zone-wise decisions shown in Figure 1. The DMC's recommendation regarding sample size, number of PFS events and patient population, *applies only to cohort 2*. Cohort 1 continues to accrue patients and events after the interim analysis without any change until the pre-specified numbers are obtained. This ensures complete independence of the *P*-values from the two cohorts, which can then be combined to produce a statistically valid final analysis.

Selecting the design parameters

The operating characteristics by which one assesses the effectiveness of a time-to-event design are power, average study duration and average sample size. For the proposed adaptive population enrichment design, these operating characteristics depend on the following design parameters: the initial specification of the number of patients and number of PFS events assigned to the two cohorts; the timing of the interim analysis; the conditional power cut-offs for determining the zone in which to classify the interim results; the formulae for increasing sample size and PFS events in the promising and enrichment zones. Choosing the right configuration for these design parameters is rather complicated and is achieved by a combination of theory, graphical display, and simulation. The final design is displayed in Figure 2 in the form of a decision tree. A detailed discussion of the process by which this decision tree was constructed is provided in supplementary Section S1, available at *Annals of Oncology* online.

Implementing the decision rules and performing the final analysis

An independent DMC will be responsible both for monitoring patient safety and implementing the above adaptive decision rules. In order to maintain strict confidentiality and minimize the possibility of operational bias, the DMC will access the interim analysis reports through a fully validated, 21 CFR Part 11 compliant, web-based database and document management system that provides role-based access to sensitive documents and has audit tracking capabilities to record who saw what and when.

The final analysis will be based on combining the *P*-values from the two cohorts. The fact that the sample size and PFS events for Cohort 1 are pre-specified and do not change after the data are unblinded at the interim analysis ensures that these *P*-values will be independent, a necessary condition for them to be combined in a statistically valid manner. In order to preserve family-wise Type 1 error, a statistical adjustment must be made to account for the multiplicity due to selecting either the full population or the cutaneous subgroup at the interim analysis. An additional statistical adjustment must be made to account for possible data-dependent changes to the sample size and number of events for Cohort 2. The mathematical details are available in supplementary Section S2,
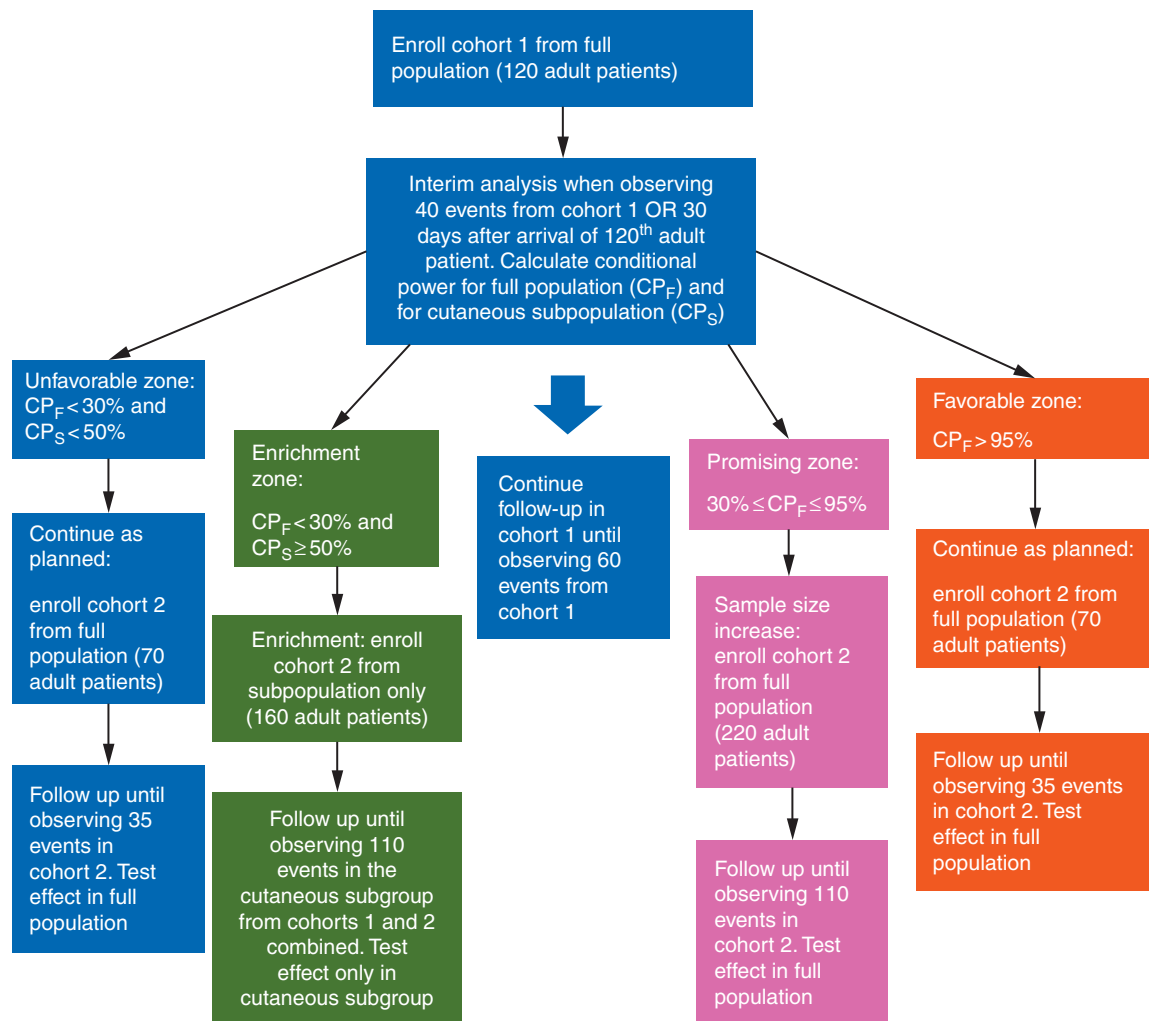
Legend: Cohort 1 will enroll a total of 120 patients and follow them until 60 PFS events are obtained. At an interim analysis based on the first 40 PFS events an independent DMC will compute CP_F, the conditional power for the full population and CP_S, the conditional power for the cutaneous subgroup. The formulae for these conditional powers are given in the Supplementary Appendix. (a) If $CPF < 0.3$ and $CPS < 0.5$, the results are in the unfavorable zone. The trial will enroll 70 patients to cohort 2 and follow them until 35 PFS events are obtained. (b) If $CPF < 0.3$ and $CPS \geq 0.5$, the results are in the enrichment zone. The trial will enroll 160 patients with cutaneous disease to cohort 2 and follow them until 110 PFS events have been obtained from the combined patients in both cohorts with cutaneous disease only. (c) If $0.3 \leq CPF \leq 0.95$, the results are in the promising zone. The trial will enroll 220 patients to cohort 2 and follow them until 110 PFS events are obtained. (d) If $CPF > 0.95$, the results are in the favorable zone. The trial will enroll 70 patients to cohort 2 and follow them until 35 PFS events are obtained. |

Figure 2. Decision tree for sample size re-estimation and enrichment.

available at *Annals of Oncology* online. The operating characteristics of this adaptive design are evaluated by simulation. They are described in detail in the 'Discussion' section and are displayed in Table 1.

Discussion

Given the rarity of AS, the absence of randomized data from a reference trial, the heterogeneous tumor type, and the requirement to establish efficacy through a time to event end point, it would have been a challenge to adopt a conventional fixed sample design for TAPPAS. Therefore, an adaptive design was adopted that incorporated regulatory input from the FDA and EMA. TAPPAS being an event-driven design was powered by PFS events rather

than sample size. Sample size nevertheless had a crucial role to play since the rate of arrival of PFS events, hence the study duration, depended on enrollment rates and drop-out rates. The initial guesses at these quantities were updated by a blinded review of the accumulated data 16 months after study initiation, resulting in a larger, more realistic sample size than initially assumed. The lesson learned from that experience causes us to recommend that such a blinded review be conducted routinely in future adaptive trials, since it will not affect the type-1 error but could affect study duration and power significantly.

Simulations were carried out to assess the operating characteristics of the adaptive enrichment design, and compare them with those of a non-adaptive, single-look trial having the same expected sample size and expected number of events. The results

Table 1. Operating characteristics of adaptive and fixed sample designs based on 10 000 simulations.

HR for cutanous/ non-cutaneous		Zone	Prob of zone	Power		Average study duration (months)		Average PFS events		Average sample size (number of patients)	
				Adaptive enrichmnt	Fixed sampl design	Adaptive enrichmnt	Fixed sampl design	Adaptive enrichment	Fixed sample design	Adaptive enrichmnt	Fixed sample design
0.55	0.55	Enrch	5%	81%		56		140		277	
		Unfav	14%	32%		40		95		190	
		Prom	38%	94%		52		170		339	
		Fav	42%	92%		40		95		190	
		Total	100%	84%	92%	45	40	126	126	252	252
0.55	0.65	Enrch	9%	80%		56		139		275	
		Unfav	19%	24%		39		95		190	
		Prom	40%	89%		50		170		338	
		Fav	32%	88%		39		95		190	
		Total	100%	76%	83%	45	40	129	129	257	257
0.55	0.75	Enrch	14%	81%		57		139		275	
		Unfav	22%	18%		38		95		190	
		Prom	40%	84%		50		170		338	
		Fav	24%	86%		38		95		190	
		Total	100%	70%	71%	45	40	131	131	261	261
0.55	0.85	Enrch	19%	79%		57		138		274	
		Unfav	24%	15%		37		95		190	
		Prom	38%	79%		49		170		336	
		Fav	19%	80%		38		95		190	
		Total	100%	64%	58%	45	40	131	131	261	261
0.55	01.0	Enrch	27%	80%		57		138		272	
		Unfav	29%	10%		37		95		190	
		Prom	33%	76%		49		170		335	
		Fav	12%	75%		36		95		190	
		Total	100%	58%	40%	46	39	131	131	260	260

Both designs have the same average number of events and average sample size.

are presented in Table 1. For the adaptive design, Table 1 displays overall power, average study duration, average events and average sample size, both zone-wise and overall. For the fixed sample design, zone-wise entries are not applicable. The fixed design has higher overall power than the adaptive design when the subgroups have similar HRs, but lower overall power than the adaptive design when the HRs become disparate. Also the fixed design has about a 5 month shorter average study duration than the adaptive design across all scenarios. However, the real benefit of the adaptive design arises from the learning and subsequent improvements in power that occur after the interim analysis. For example, if the HR is 0.55 for the cutaneous subgroup and 0.75 for the non-cutaneous, then the adaptive and non-adaptive designs have almost the same overall power of around 70%. But the power of the adaptive design is boosted to 81% in the enrichment zone, and boosted to 84% in the promising zone thereby greatly improving the chances of a successful outcome, should the interim results fall in one of these zones. To be sure these increases in power come at the cost of corresponding increases in sample size and PFS events, but the latter are only called up after an informed examination of the interim results by the DMC. In contrast the fixed design with the same average sample size and PFS events as the corresponding adaptive design has 70% overall

power and no opportunity to make changes and gain additional power. The greater the heterogeneity in the two subgroups the more advantageous is the adaptive design. Thus if the HR is 0.55 for the cutaneous subgroup and 1 for the non-cutaneous, the adaptive design still has a reasonable chance for a successful outcome with 80% power conditional on being in the enrichment zone and over 75% power in the promising and favorable zones. On the other hand, the fixed design has only 40% overall power and no possibility of improving on it.

When there is limited heterogeneity the fixed design is a more efficient option for it achieves greater overall power for the same expected sample size and PFS events. This is because the adaptive design faces the statistical burden of closed hypothesis testing (see supplementary Section S2, available at *Annals of Oncology* online) simply to allow the option for enrichment even if no enrichment is implemented. This underscores the importance of having a strong a priori clinical or biological basis for expecting heterogeneity in the subgroups. If there is no such basis, it might be preferable to not build the possibility for enrichment into the design but rely on sample size adaptations alone to improve power. Sample size adjustments when not accompanied by subgroup or dose selection carry a minimal penalty for they do not require closed testing [9].

The typical application of adaptive designs is for seamlessly expanding a randomized phase II trial to phase III. In such designs the phase II portion may consist of multiple treatment arms versus a common control or it may be a two-arm proof-of-concept (POC) trial. In either case, an interim analysis is carried out with pre-specified decision rules for either terminating the trial for futility, completing it as planned, or expanding it to phase III by dropping poorly performing arms and expanding the sample size of the remaining arms. For the final analysis, the data from both phases are combined with appropriate statistical adjustments. A recent example was proposed by Chen et al. [10] in which some additional flexibility was built in to allow the use of a short-term end point for decision making and a long-term end point for the final analysis. The example was applied to a hypothetical non-small-cell lung cancer trial. The uniqueness of the TAPPAS trial is that in addition to sample size re-estimation, it permits population enrichment. Moreover, TAPPAS is an example of an actual ongoing pivotal trial. We are not aware of any other pivotal population enrichment trial that has actually been implemented in oncology. For a general discussion of adaptive designs that includes several case studies of trials that were implemented in the respiratory and cardiovascular therapeutic areas, refer to Bhatt and Mehta [11].

Funding

Disclosure

The authors have declared no conflicts of interest.

References

1. ESMO/European Sarcoma Network Working Group. Soft tissue and visceral sarcomas: eSMO Clinical Practice Guidelines for diagnosis, treatment and follow-up. Ann Oncol 2014; 25: iii102–iii112.
2. Rosen LS, Gordon MS, Robert F et al. Endoglin for targeted cancer treatment. Curr Oncol Rep 2014; 16(2): 365.
3. Duffy AG, Ma C, Ulahannan SV et al. Phase I and preliminary phase II study of TRC105 in combination with sorafenib in hepatocellular carcinoma. Clin Cancer Res 2017; 23(16): 4633–4641.
4. Tian H, Huang JJ, Golzio C et al. Endoglin interacts with VEGFR2 to promote angiogenesis. Faseb J 2018; https://doi.org/10.1096/fj.201700867RR.
5. Sankhala K, Riedel R, Conry R et al. Every other week dosing of TRC105 (endoglin antibody) in combination with pazopanib in patients with advanced soft tissue sarcoma. In Connective Tissue Oncology Society Annual Meeting, November 2017, Maui, Hawaii.
6. Kollar A, Jones R, Silva S et al. Pazopanib in advanced vascular sarcomas: An EORTC Soft Tissue and Bone Sarcoma Oncology Group Retrospective Analysis. In Connective Tissue Oncology Society Annual Meeting, 2016, Maui, Hawaii.
7. Bauer P and Posch M. Modification of the sample size and the schedule of interim analyses in survival trials based on data inspections. Letter to the Editor. Stat Med 2004; 23: 1333–1335.
8. Jenkins M, Stone A, Jennison CJ. An adaptive seamless phase II/III design for oncology trials with subpopulation selection using correlated survival endpoints. Pharm Stat 2010; doi: 10.1002/pst.472.
9. Mehta C and Pocock S. Adaptive increase in sample size when interim results are promising: a practical guide with examples. Stat Med 2011; 30(28): 3267–3284.
10. Chen C, Anderson K, Mehrotra DV et al. A 2-in-1 adaptive phase 2/3 design for expedited oncology drug development. Contemp Clin Trials 2018; 64: 238–242.
11. Bhatt D and Mehta C. Adaptive designs for clinical trials. N Engl J Med 2016; 375(1): 65–74.