

RESEARCH ARTICLE

Open Access



Patient recruitment strategies for adaptive enrichment designs with time-to-event endpoints

Ryuji Uozumi^{1*} , Shinjo Yada² and Atsushi Kawaguchi³

Abstract

Background: Adaptive enrichment designs for clinical trials have great potential for the development of targeted therapies. They enable researchers to stop the recruitment process for a certain population in mid-course based on an interim analysis. However, adaptive enrichment designs increase the total trial period owing to the stoppage in patient recruitment to make interim decisions. This is a major drawback; it results in delays in the submission of clinical trial reports and the appearance of drugs on the market. Here, we explore three types of patient recruitment strategy for the development of targeted therapies based on the adaptive enrichment design.

Methods: We consider recruitment methods which provide an option to continue recruiting patients from the overall population or only from the biomarker-positive population even during the interim decision period. A simulation study was performed to investigate the operating characteristics by comparing an adaptive enrichment design using the recruitment methods with a non-enriched design.

Results: The number of patients was similar for both recruitment methods. Nevertheless, the adaptive enrichment design was beneficial in settings in which the recruitment period is expected to be longer than the follow-up period. In these cases, the adaptive enrichment design with continued recruitment from the overall population or only from the biomarker-positive population even during the interim decision period conferred a major advantage, since the total trial period did not differ substantially from that of trials employing the non-enriched design. By contrast, the non-enriched design should be used in settings in which the follow-up period is expected to be longer than the recruitment period, since the total trial period was notably shorter than that of the adaptive enrichment design. Furthermore, the utmost care is needed when the distribution of patient recruitment is concave, i.e., when patient recruitment is slow during the early period, since the total trial period is extended.

Conclusions: Adaptive enrichment designs that entail continued recruitment methods are beneficial owing to the shorter total trial period than expected in settings in which the recruitment period is expected to be longer than the follow-up period and the biomarker-positive population is promising.

Keywords: Adaptive enrichment designs, Time-to-event endpoints, Patient recruitment, Interim analysis, Targeted therapies

*Correspondence: uozumi@kuhp.kyoto-u.ac.jp

¹Department of Biomedical Statistics and Bioinformatics, Kyoto University Graduate School of Medicine, 54 Kawahara-cho, Shogoin, Sakyo-ku, 606-8507 Kyoto, Japan

Full list of author information is available at the end of the article



Background

The development of targeted therapies is highly attractive to numerous investigators in pharmaceutical research owing to the rapid advancement of personalized medicine [1]. In the clinical phase, biomarker information based on gene expression contributes to identification of subgroups. Generally, gene expression profiles are analyzed by integrated omics technologies [2]. In confirmatory clinical trials, there are numerous examples of therapies that showed a benefit for only a specified subgroup. For instance, gefitinib significantly improved patients with advanced pulmonary adenocarcinoma, particularly those in the subgroup that consists of biomarker-positive patients for epidermal growth factor receptor (EGFR) inhibitor with pulmonary adenocarcinoma, whereas no improvement was observed in patients who tested negative [3]. This trial was conducted for both the biomarker-positive and -negative patients, although this was not pre-specified to demonstrate the benefit for only the biomarker-positive patients. In contrast, as another example, crizotinib significantly improved patients in the subgroup under a clinical trial in which only biomarker-positive patients with advanced anaplastic lymphoma kinase (ALK)-positive nonsquamous non-small cell lung cancer (NSCLC) were set as the targeted population [4]. In order to develop targeted therapies, a complex clinical trial design considering the existence of a promising biomarker to formulate the subgroup is required. Like in the gefitinib example [3], we aimed to set up a clinical trial targeting both the biomarker-positive and -negative patients.

Adaptive enrichment designs could be useful for the development of targeted therapies when the overall population (O), which consists of both biomarker-positive (P) and biomarker-negative (N) populations, is set as the target and particularly when P is promising, but not certain, based on the early phase clinical trial results. These designs provide an opportunity to stop the recruitment process for patients in N in mid-course and continue enrolling patients in only P on the basis of an interim analysis, whereas the classical group sequential design solely provides the opportunity to stop the trial early for efficacy or futility reasons. Accordingly, it is possible to curtail the duration in which patients are exposed to a fruitless treatment if an experimental treatment is overwhelmingly worse than the standard treatment in N , but is a benefit in P .

We consider an adaptive enrichment design where the endpoint is time-to-event. In this context, there are various advantages with respect to stopping the investigation for N in mid-course depending on the recruitment situation, i.e., whether the interim analysis is performed after accumulating the targeted number of patients or not. The advantages can be illustrated with two phase III clinical

trial examples from CLEOPATRA and REGATTA studies [5–8]. First, pertuzumab was evaluated in patients with anti-human epidermal growth factor receptor 2 (HER2)-positive metastatic breast cancer in a CLEOPATRA study that incorporated interim analyses [5–7]. The first interim analysis showed a significantly improved progression-free survival [5], and the second interim analysis showed a significantly improved overall survival as well [6]. With respect to patient recruitment, the first interim analysis included 808 patients, satisfying the targeted sample size (TSS) of 800 patients. Thus, the goal of interim analyses was not to decrease the number of recruited patients, which already exceeded the TSS at the time of the first interim analysis. If the adaptive enrichment design is applied in this case, the only advantage is the potential for a reduced follow-up period for patients in N .

In contrast, in the REGATTA study, an interim analysis was conducted to investigate the efficacy of the addition of gastrectomy to chemotherapy in advanced gastric cancer patients with a single non-curable factor when the targeted number of patients was not yet reached [8]. The required sample size in the REGATTA study was set to 330 and there was an interim futility criterion, i.e., a hazard ratio of greater than 1.00. Consequently, this trial was terminated owing to futility at the interim analysis based on a hazard ratio (95% confidence interval) for overall survival of 1.09 (0.78 to 1.52). By stopping the trial early, the total number of patients was reduced by approximately half, since the interim analysis included 175 patients. Accordingly, in this case, the adaptive enrichment design is advantageous with respect to decreasing the number of recruited patients as well as curtailing the follow-up period for N . However, patient recruitment is halted until the interim decision is made. This increases the duration of the trial and delays the submission of experimental treatments for market approval. In practice, a longer trial duration is a particularly serious drawback of the adaptive enrichment design for pharmaceutical companies.

In this paper, we explore patient recruitment methods for the development of targeted therapies with an adaptive enrichment design. In the setting of the adaptive enrichment design with time-to-event endpoints, data are handled depending on the stage in which each patient is recruited. For example, a patient recruited in the first stage and observed as an event in the second stage will be handled as first-stage data. We consider patient recruitment in the context of a two-stage adaptive enrichment design. To the best of our knowledge, patient recruitment within the adaptive enrichment design setting typically needs to be halted pending an interim decision, and is resumed after the interim decision has been made. Here, we focus on patient recruitment that is continued for patients in O , or only for patients in P , even during the period pending interim decisions after the first stage. We

performed a simulation study to demonstrate the major drawback of using the typical recruitment method with an adaptive enrichment design as well as the advantages of incorporating continued recruitment methods into this design.

Methods

Review of the adaptive enrichment design with time-to-event endpoints

A number of adaptive designs in confirmatory clinical trials have been investigated, such as the adaptive treatment selection design, adaptive sample size modification design, and adaptive enrichment design [9]. In this paper, we consider the adaptive enrichment design for the development of targeted therapies. In this design, an interim analysis is conducted to identify whether *O* or only *P* should be included in the subsequent stage. For time-to-event endpoints, constructing test statistics for the final analysis by handling data depending on the stage at which each patient is recruited is needed to control the type I error rate [10]. Let $p_k^{(g)}$ denote the p-value for stage *k* in population $g \in \{O, P\}$. The final analysis with a one-sided significance level of α under a null hypothesis $H_0^{(g)}$ will be conducted following an inverse combination method [11] as follows:

$$C(p_1^{(g)}, p_2^{(g)}) = w_1 \Phi^{-1}(1 - p_1^{(g)}) + w_2 \Phi^{-1}(1 - p_2^{(g)}) > \Phi^{-1}(1 - \alpha), \tag{1}$$

where $\Phi(\cdot)$ denotes the cumulative distribution function of the standard normal distribution and w_k denotes the pre-specified weight for each stage, defined as $w_k = \sqrt{I_{O,k} / \sum_j I_{O,j}}$ with information level $I_{O,k}$ at stage *k* in the calculation of $C(p_1^{(O)}, p_2^{(O)})$ under the assumption that the p-values from each stage are independent. Note that $I_{O,k}$ is set as the pre-specified number of events that corresponds to information from *O* in the context of time-to-event endpoints. Here, w_k is also used in the calculation of $C(p_1^{(P)}, p_2^{(P)})$ under the assumption that the pre-specified number of events for *P* is the same as $I_{O,k}$. The Hochberg [12] procedure is applied to control the familywise type I error rate at the final analysis. Let $p_k^{(O,P)} = \min[2 \cdot \min(p_k^{(O)}, p_k^{(P)}), \max(p_k^{(O)}, p_k^{(P)})]$ denote the p-value for the intersection hypothesis $H_0^{(O,P)} = H_0^{(O)} \cap H_0^{(P)}$. Following the closure principle [13], $H_0^{(O)}$ is rejected if $C(p_1^{(O,P)}, p_2^{(O,P)}) < \alpha$ and $C(p_1^{(O)}, p_2^{(O)}) < \alpha$ are satisfied and $H_0^{(P)}$ is rejected if $C(p_1^{(O,P)}, p_2^{(O,P)}) < \alpha$ and $C(p_1^{(P)}, p_2^{(P)}) < \alpha$ are satisfied in cases where *O* is determined to be continued, whereas $H_0^{(P)}$ is rejected if

$C(p_1^{(O,P)}, p_2^{(P)}) < \alpha$ and $C(p_1^{(P)}, p_2^{(P)}) < \alpha$ are satisfied in case where only *P* is determined to be continued and *N* is terminated at the interim analysis.

In contrast, typical hazard ratios are often used to conduct interim analyses for the sake of simplicity, although several approaches have been developed to make interim decisions. In this paper, we consider the interim decision rule using the hazard ratios $HR_{int}^{(g)}$ with the thresholds $\eta^{(g)}$ for population $g \in \{O, P\}$ at the interim as follows: (i) stop the trial for futility if $HR_{int}^{(O)} > \eta^{(O)}$ and $HR_{int}^{(P)} > \eta^{(P)}$, (ii) continue *O* if $HR_{int}^{(O)} < \eta^{(O)}$ regardless of the value of $HR_{int}^{(P)}$, (iii) continue only *P* if $HR_{int}^{(O)} > \eta^{(O)}$ and $HR_{int}^{(P)} < \eta^{(P)}$, where the experimental treatment is beneficial against the control if the hazard ratio is less than 1.00.

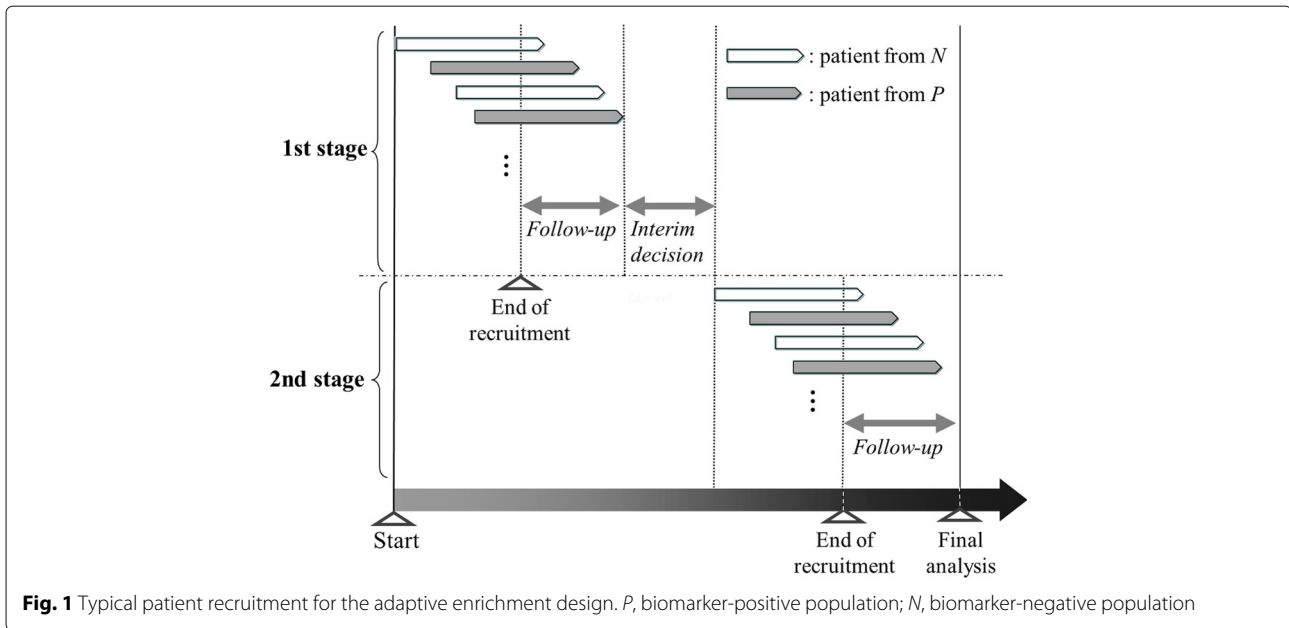
Patient recruitment strategies in the adaptive enrichment design

In the context of time-to-event endpoints, it is necessary to consider the recruitment and follow-up periods to accumulate the targeted number of events. To establish a randomized parallel-group clinical trial with equal allocation comprising the experimental (*E*) and control (*C*) groups, it is widely popular to estimate the TSS to achieve $100 \cdot (1 - \beta)\%$ power under a one-sided test with a significance level of α according to the Schoenfeld [14] formula as follows:

$$TSS = \frac{4 \{ \Phi^{-1}(1 - \alpha) + \Phi^{-1}(1 - \beta) \}^2}{\delta^2} \times \frac{2}{S_E(t) + S_C(t)}$$

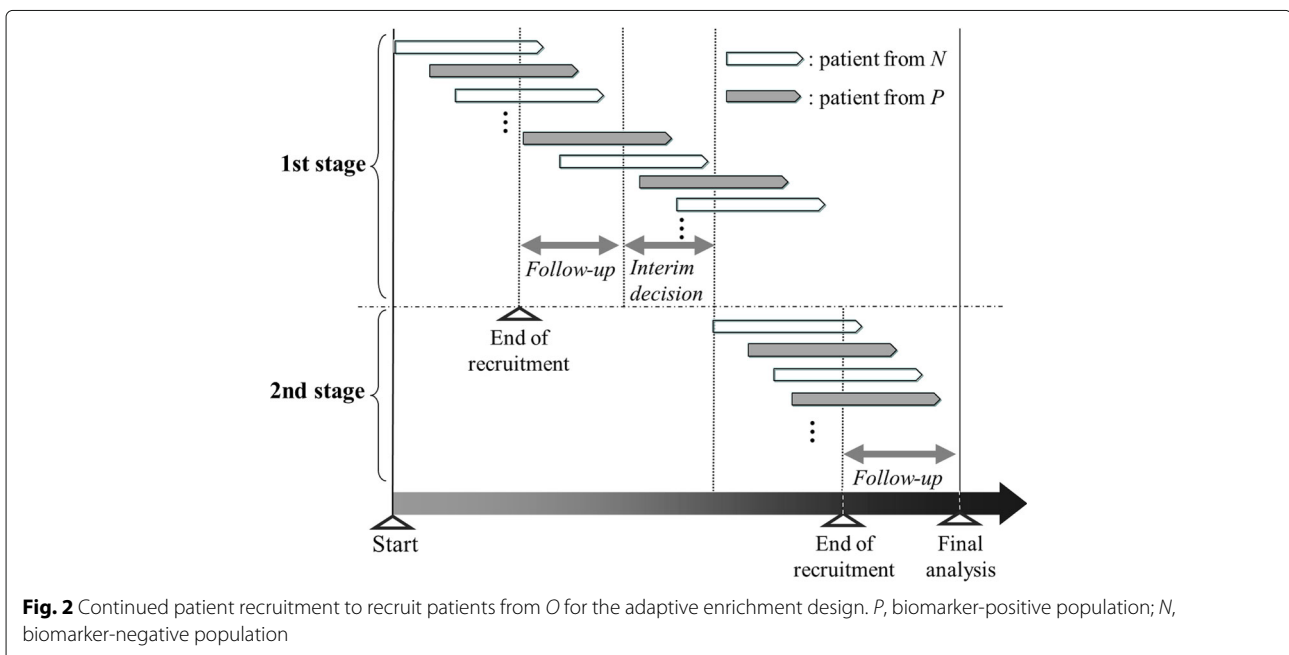
where δ is the log-transformed hazard ratio and $S_E(t)$ and $S_C(t)$ are the survival probabilities for *E* and *C* groups at time *t*. In this formula, the TSS is estimated by considering the expected event rates. The TSS varies depending on the recruitment and follow-up periods during the estimation of the expected event rates, in contrast with the TSS calculation in the setting in which the number of events is considered a binary endpoint.

As described in the Background, whether the required number of patients has been recruited before the interim analysis is the key issue in the adaptive enrichment design. In this paper, we consider a method in which patient recruitment is halted until the interim decision regarding the continuation of the population is made, as supposed in most methodology papers considering adaptive enrichment designs, e.g., Jenkins et al. [10], Brannath et al. [15], Friede et al. [16], Krisam and Kieser [17], and Uozumi and Hamada [18]. In contrast to cases where the required number of patients has been recruited before the interim analysis, such as in the CLEOPATRA study [5], this is expected to result in longer total trial periods owing to the pause in recruitment to make the interim decision. As illustrated in Fig. 1, the typical recruitment



method in the adaptive enrichment design has to be halted from the end of recruitment at the first stage until recruitment is resumed at the beginning of the second stage. In contrast, the alternative recruitment methods enable continued recruitment, even during the period from the end of recruitment at the first stage to the interim decision based on the interim analysis results, as illustrated in Figs. 2 and 3. The first method shown in Fig. 2 allows continued recruitment for patients in O during the interim decision period. In the second method

described in Fig. 3, recruitment is continued for patients in P with the belief that P is promising whereas N is harmful. A benefit of the continued recruitment method for P is that it provides the opportunity for investigators to stop N early owing to futility, under the assumption that the experimental treatment in P is relatively promising. Stopping N early reduces the actual number of patients recruited and curtails the follow-up period for biomarker-negative patients enrolled before the interim analysis. Note that no modification to the statistical methods used



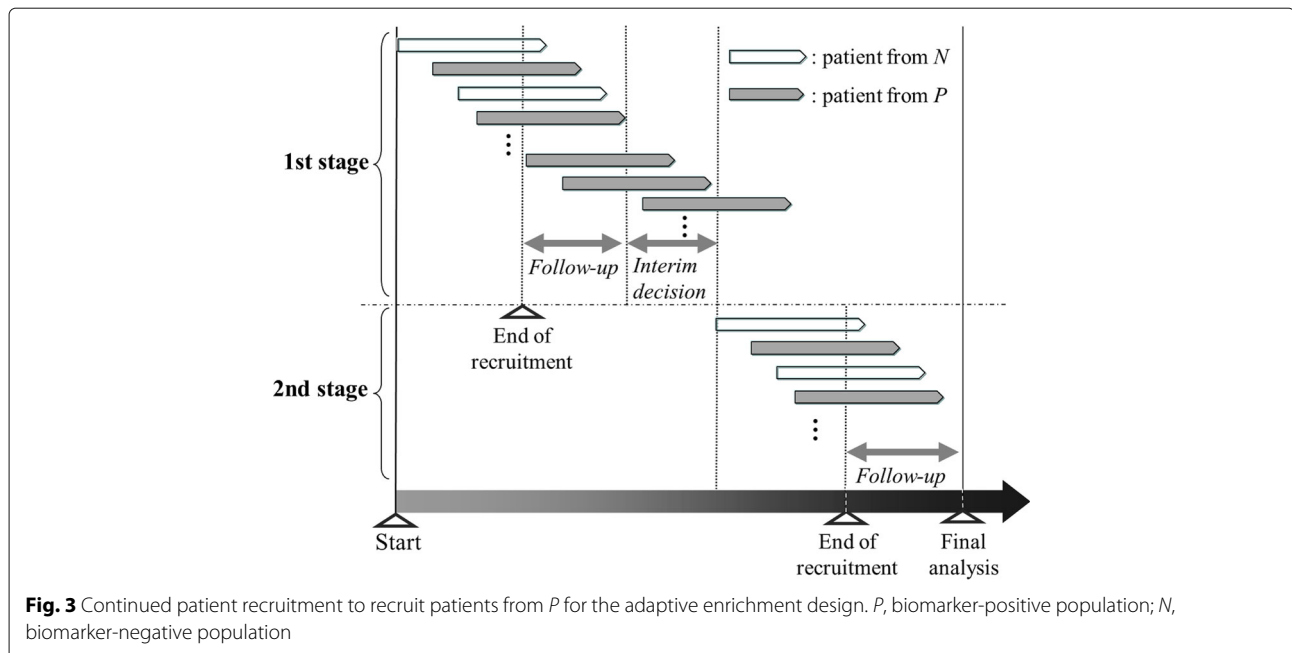


Fig. 3 Continued patient recruitment to recruit patients from P for the adaptive enrichment design. P , biomarker-positive population; N , biomarker-negative population

for the adaptive enrichment design is required when the continued recruitment methods illustrated in Figs. 2 and 3 are applied.

Results

Simulation study

We conducted simulations to investigate the operating characteristics of three types of patient recruitment strategy using an adaptive enrichment design with respect to the expected numbers of patients and total trial periods. The two alternative methods that allow continued recruitment for O or only P are assessed. In our simulations, we consider a randomized, parallel-group clinical trial with two arms, comprising E and C groups.

Simulation setting

With respect to the population, O includes 50% of P in a setting wherein a promising biomarker exists and in which the targeted therapy is beneficial only for P . We set the survival probability in C at the end of follow-up to 20%. The survival probability in E at the end of follow-up was derived on the basis of the assumed hazard ratio between E and C . With respect to the assumed hazard ratios, we consider five scenarios for applications to the development of targeted therapies. The assumed hazard ratios for P are set to 0.50 across every scenario, whereas those for N are set to (Scenario 1) 0.50, (Scenario 2) 0.90, (Scenario 3) 1.00, (Scenario 4) 1.11, and (Scenario 5) 1.43. We set the sample size for O to 330 following a previous example [8]. When the pre-specified number of events set to 270 occurs, the final analysis is conducted with a one-sided significance level of 2.5% following the closure principle

[13] and using the Hochberg [12] procedure with respect to the multiplicity between O and P . An interim analysis is performed after half of the pre-specified events occur in order to determine whether subsequent recruitment includes O or is restricted to P . We do not consider early termination if E is deemed promising for efficacy. With respect to the thresholds in the interim decision rule, $\eta^{(O)}$ and $\eta^{(P)}$ are each set to 1.00 for the sake of simplicity.

Durations of patient recruitment R and follow-up F are set to $(R, F) = (2, 4), (4, 2), (6, 2)$, respectively. We consider both uniform and non-uniform recruitment patterns over the interval 0 to R . When it comes to the non-uniform recruitment pattern, the truncated exponential distribution is assumed with the cumulative distribution function as follows:

$$F(t) = \frac{1 - \exp(-\gamma t)}{1 - \exp(-\gamma R)},$$

where γ denotes the shape parameter, with $\gamma \neq 0$ [19]. The distribution of recruitment is convex if $\gamma > 0$ and concave if $\gamma < 0$. For instance, for $\gamma < 0$, the recruitment speed is slow at the early period. In this simulation setting, the slow recruitment pattern, with $\gamma < 0$ and using the continued recruitment methods, also shows that recruitment accelerates during the interim decision. In contrast, the fast recruitment pattern, with $\gamma > 0$ and using the continued recruitment methods, implies that recruitment slows down during the interim decision period. For convex and concave recruitment patterns, we set γ to 2 and -2 . Furthermore, we consider the duration for data lock between stages. The duration from the data lock of

the first stage to the beginning of the second stage, i.e., interim decision period, is set to 0.2.

The timing of the interim analysis is mainly assumed as 50% of the pre-specified events based on the example [8]. We further evaluate the trade-off between timing of the interim analysis and the accuracy of the interim results. The timing of the interim analysis is set to 25%, 50%, and 75% of the pre-specified events. The accuracy of the interim hazard ratio for P is assessed.

Simulation results

We evaluated the operating characteristics of the two recruitment methods in the adaptive enrichment designs in terms of the expected numbers of patients and expected total trial periods. The characteristics were quantified based on 10,000 simulations.

Figure 4 shows the expected numbers of patients from O using each recruitment method within the adaptive enrichment design assuming uniform, convex, and concave recruitments. In general, the expected number of patients was less than the reference number of 330 when the non-enriched design was employed. The non-enriched design corresponds to the group sequential design that includes early termination only for futility. As shown in Fig. 4, the minimum number of patients is 165 under typical recruitment owing to the possibility of no extra recruitment after recruiting half of the targeted number of patients. The minimum number of patients using continued recruitment with P in the setting

of $F > R$ is calculated as 247 which consists of half number of patients before the interim analysis and patients in P after the interim analysis. In particular, the number of patients was decreased in Scenario 5, in which the E is notably harmful in N , as adaptive enrichment for P is expected. In the continued recruitment methods, patients are continually enrolled, even during the period in which the interim decision is made, as described in Figs. 2 and 3. The expected number of patients is similar among the recruitment methods in Scenarios 1 to 4, regardless of recruitment patterns, when F is expected to be longer than R . In Scenario 5, under the setting of $F > R$, the expected number of patients using the continued recruitment method to recruit patients from O is largest among the three methods. However, the difference is apparent when R is expected to be longer than F . The number of patients using the continued recruitment methods is expected to be lower, especially when $(R, F) = (6, 2)$. In particular, the advantage of the continued recruitment methods is observed in Scenario 5 in terms of the similar distributions to the typical recruitment method. The distribution obtained using the continued recruitment method to recruit patients from P is more similar to that obtained using the continued recruitment method to recruit patients from O . Both continued recruitment methods require more patients than the typical recruitment method owing to the continued recruitment during the interim decision period. The continued recruitment

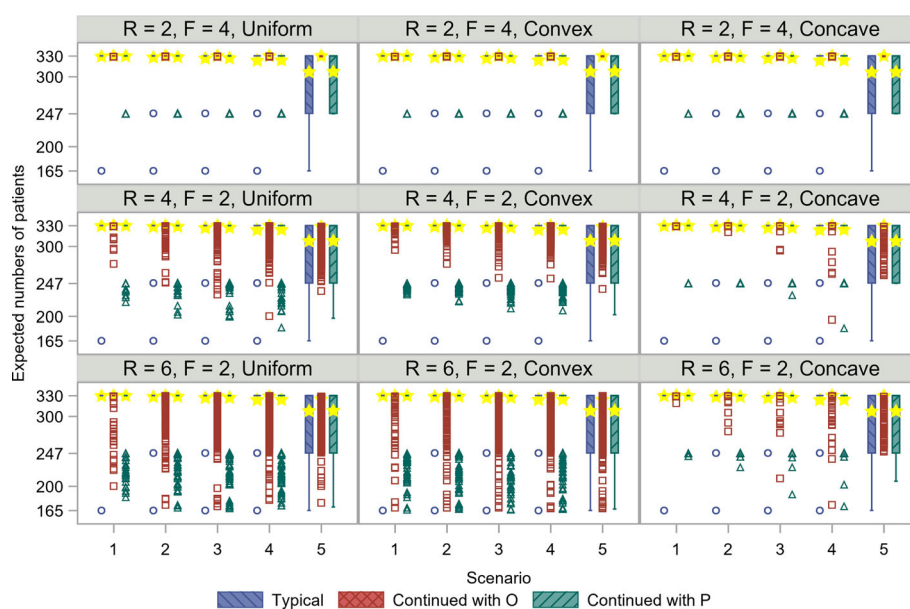
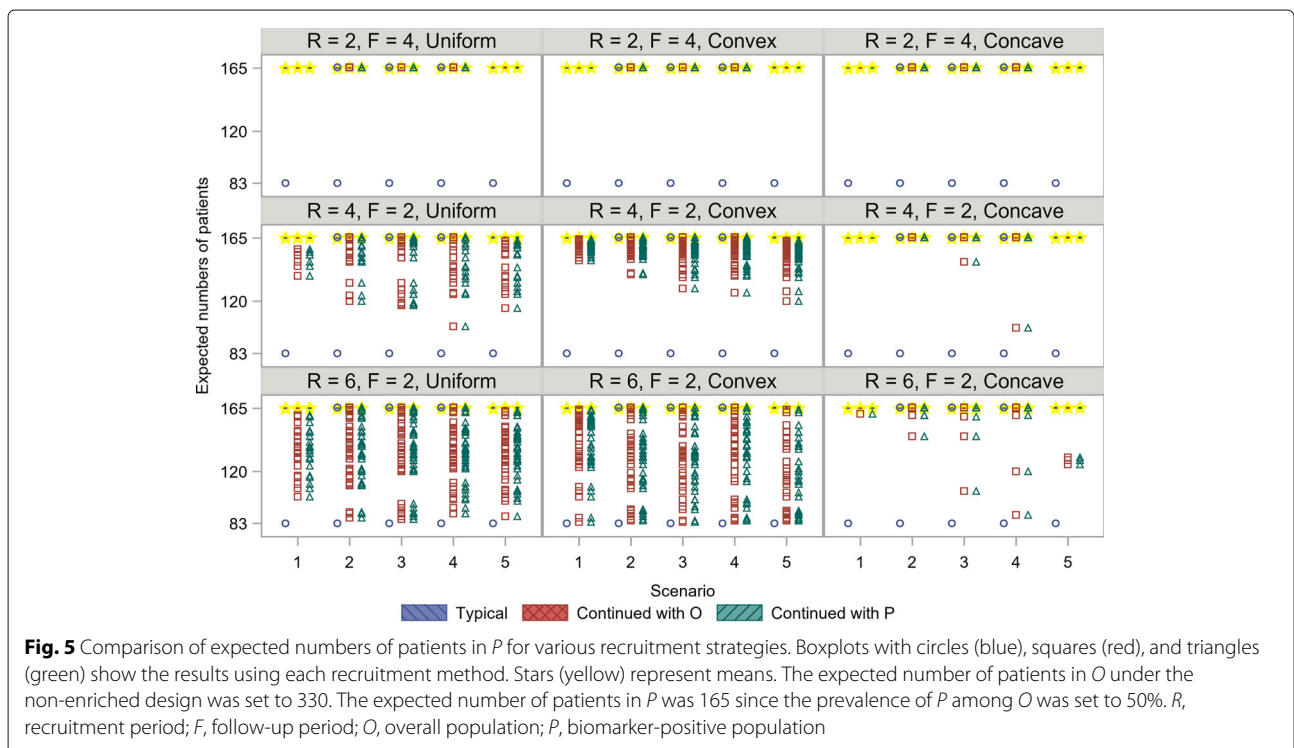


Fig. 4 Comparison of expected numbers of patients in O for various recruitment strategies. Boxplots with circles (blue), squares (red), and triangles (green) show the results using each recruitment method. Stars (yellow) represent means. The expected number of patients under the non-enriched design was set to 330. R , recruitment period; F , follow-up period; O , overall population; P , biomarker-positive population

method to recruit patients from O particularly needs far more patients than the continued recruitment method to recruit patients from P since the continued recruitment is not limited to patients from P . Moreover, the difference in the minimum number of patients was notable, even in Scenarios 1 to 4. Figure 4 shows the differences in expected number of patients for the various recruitment patterns. Although the expected and maximum number of patients are similar regardless of the recruitment patterns, the minimum number of patients with the concave recruitment pattern, i.e., $\gamma < 0$, is higher than with the uniform recruitment pattern. On the other hand, the minimum number of patients with the convex recruitment pattern, i.e., $\gamma > 0$, is closer to that with the uniform recruitment pattern. Next, Fig. 5 shows the expected numbers of patients from P using each recruitment method. The expected number of patients in P is half of 330 since the prevalence of P among O is set to 50%. Owing to the possibility of early termination for futility after recruiting half of the targeted number of patients, the numbers of patients in P using the continued recruitment methods with O or P were distributed from 83 to 165, especially in the setting of $R > F$.

Figure 6 summarizes the expected total trial periods for each recruitment method. Uniform and truncated exponential distributions were assumed to investigate the performance of the recruitment methods. As a reference, the expected total trial period is $R + F$ when the non-enriched design is employed. In comparison, the expected

total trial period was shorter in Scenario 5, since the probability of selecting only P is highest among the five scenarios. With respect to the recruitment and follow-up periods (R, F), the non-enriched design should be used when F is expected to be longer than R , since the total trial period using the adaptive enrichment design is much longer than that using the non-enriched design. By contrast, the adaptive enrichment design is applicable when R is expected to be longer than F , since the total trial period using the adaptive enrichment design does not substantially differ from that of the non-enriched design. These results are particularly apparent when $(R, F) = (6, 2)$, i.e., when R is expected to be much longer than F . Despite the similar expected numbers of patients for all of the recruitment methods, the expected total trial period is notably shorter for the continued recruitment methods than for the typical recruitment method; hence, the application of the adaptive enrichment design is attractive when using the continued recruitment methods. In particular, the expected total trial period is shorter using the continued recruitment method with O , despite more patients being required. Note, however, that the minimum number of patients is larger when the adaptive enrichment design is applied. In addition, situations in which patient recruitment pattern is concave during R should be handled with the utmost care, since a longer total trial period is expected to achieve the pre-specified number of events than the setting in which recruitment follows a uniform or convex pattern.



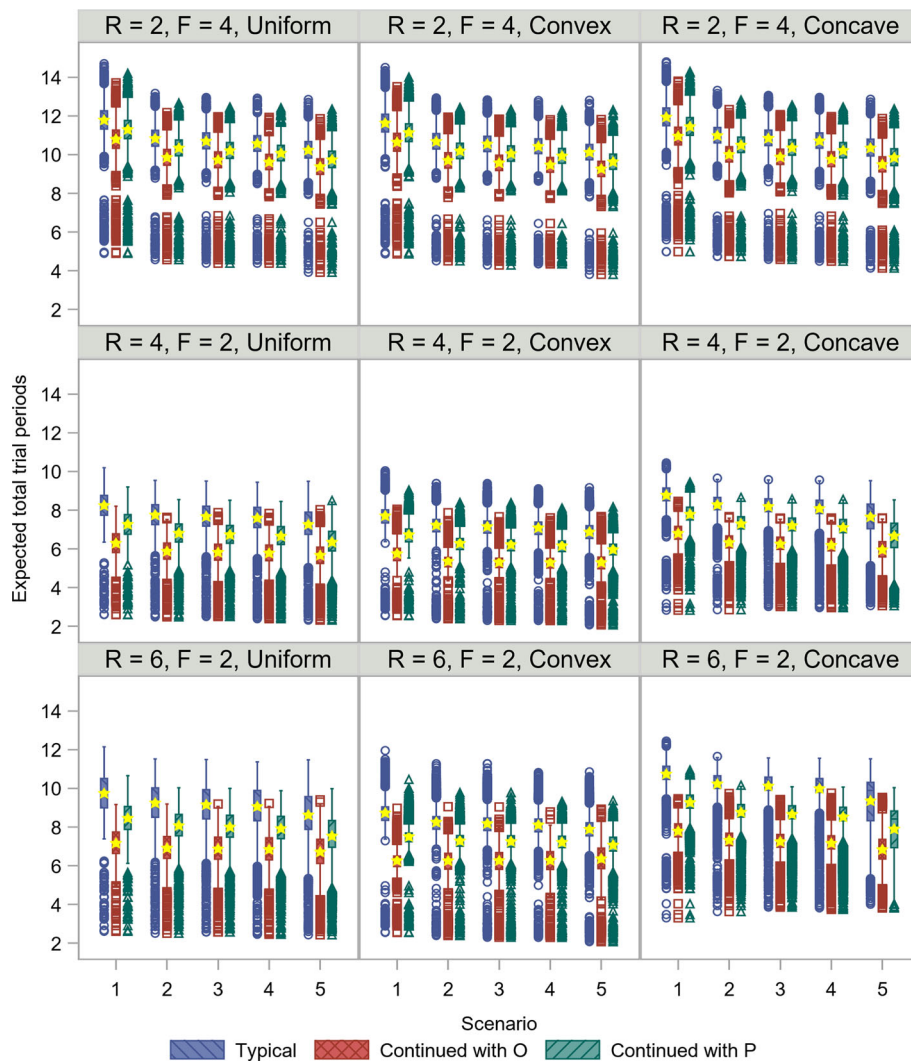


Fig. 6 Comparison of expected total trial periods among patient recruitment strategies. Boxplots with circles (blue), squares (red), and triangles (green) show the results using each recruitment method. Stars (yellow) represent means. R , recruitment period; F , follow-up period; O , overall population; P , biomarker-positive population

Finally, Fig. 7 illustrates the boxplots assessing trade-off between timing of the interim analysis and the accuracy of interim hazard ratio for P while providing the expected number of patients assuming uniform, convex, and concave recruitment patterns under the setting of $(R, F) = (6, 2)$ and Scenario 5. Early interim analysis with enrichment obviously leads to a beneficial result in terms of smaller number of patients required; however, the accuracy is severely decreased owing to the lack of information. This implies that the probability of selecting an inappropriate population in mid-course is increased with early interim analysis. In contrast, accuracy is improved when a late interim decision is made, but the benefit of adaptive enrichment designs is reduced owing to a large number of patients required.

Discussion

Enrichment designs are generally applied without stopping recruitment; however, we focus on adaptive enrichment designs with emphasis on whether patient recruitment is halted. In this paper, we assessed patient recruitment methods in the context of adaptive enrichment designs for the development of targeted therapies in the setting in which recruitment is halted until an interim decision regarding the continuation of O or P is made. The adaptive enrichment design with the typical recruitment method results in a longer total trial period than that of the non-enriched design. As an alternative, the continued recruitment methods resolve this issue, i.e., they enable a shorter total trial period. Despite this benefit, the expected number of patients required does

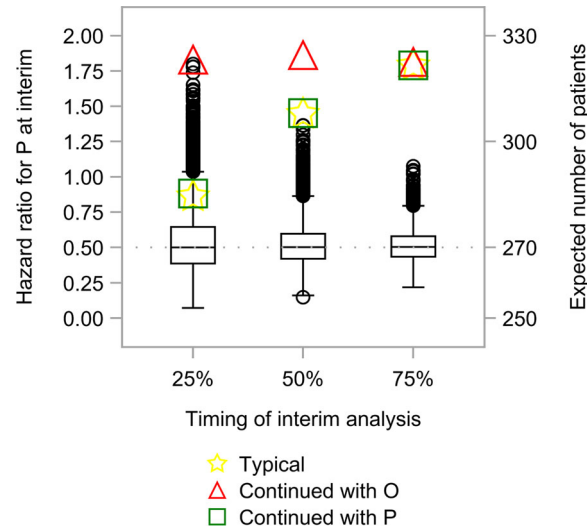


Fig. 7 Comparison between timing of the interim analysis and the accuracy of interim hazard ratio for P . Boxplots with circles (black) show trade-off between timing of the interim analysis and the accuracy of interim hazard ratio for P . Stars (yellow), triangles (red), and squares (green) represent the expected number of patients using patient recruitment strategies. R , recruitment period; F , follow-up period; O , overall population; P , biomarker-positive population

not differ between the three recruitment methods. In the continued recruitment methods, recruitment is continued even after the first stage of recruitment. In particular, the continued recruitment method to recruit patients from O enables the total trial period to be shortened. However, this method has the corresponding possibility of increasing the anticipated minimum number of patients required. On the other hand, the continued recruitment method to recruit patients only from P requires the assumption that P is promising; however, it is more attractive in terms of the lower anticipated minimum number of patients. Using the continued recruitment method to recruit patients from P is preferred if reducing sample size is more desirable than reducing the trial period. This approach resulted in a similar number of patients to the typical recruitment method. The number of patients differed from that obtained using the continued recruitment to recruit from O , particularly under the setting in which R is expected to be longer than F . With emphasis on total trial period, it is preferable to use the continued recruitment method to recruit patients from O should be employed, taking into account the trade-off between sample size and total trial period.

In our simulation study, we considered that the assumed hazard ratios are consistent between stages. According to the draft guidance by the Food and Drug Administration [20], adaptive designs can lead to potential operational biases that result in inconsistency between stages. If potential operational bias is introduced after the interim analysis, we might face a situation in which N is stopped at the interim analysis, but in which both P and N are found

to be promising after the interim decision. To ensure trial integrity, access to interim analysis results should be limited to the data monitoring committee members who are independent of the investigators being involved in the trial. The potential operational bias introduced in this scenario will be assessed in future work.

For the sake of simplicity, we did not consider early termination if the experimental treatment was deemed promising with respect to efficacy. In the context of adaptive enrichment designs, there are several types of assumptions for clinical trial phases, such as phase II/III or phase III settings. When early termination for each population is considered in a phase III clinical trial, the O'Brien and Fleming [21] stopping boundary at the time of the interim analysis is calculated with the use of a Lan and DeMets [22] alpha-spending function on the basis of the number of events observed at the data cutoff date. Moreover, we set the proportion of P among O to 50% for simplicity. The continued recruitment methods reduce the length of the trial period when this proportion is higher than 50%. By contrast, the benefit of these methods is reduced when the proportion is less than 50%. In practice, the thresholds for the hazard ratios or other measures for each population in the interim decision should be carefully set based on thorough simulations for various scenarios, as Gallo et al. [23] pointed out when using simple group sequential designs.

Furthermore, the timing of the interim analysis in the adaptive enrichment design should be considered, similar to the context of a group sequential design

[24, 25], because early decisions have the potential risk for over-estimation of treatment effect [26, 27]. Our simulation results indicate that early interim analysis spoils the accuracy of interim results that leads to an incorrect decision. To supplement immature information in mid-course, the potential use of biomarkers or other endpoints associated with the primary endpoint of interest would be useful to drive the adaptive enrichment decision. When overall survival is assumed as a primary endpoint, other time-to-event endpoints such as progression-free survival would be beneficial to strengthen the information level in mid-course in the situation in which progression-free survival is highly correlated with overall survival, as illustrated in Uozumi and Hamada [18]. Although we assumed performing the interim decision with a simple rule, more investigation for the optimal timing should be required when applying the adaptive enrichment design in practice since performance greatly depends on the interim decision rule as shown in Benner and Kieser [28].

Note that the adaptive enrichment design is not encouraged for the development of targeted therapies in some conditions in which the non-enriched design is more appropriate. For instance, the adaptive enrichment design with more than two stages would be much more inefficient in terms of halting the recruitment several times to make interim decisions. Similarly, this is true even using the adaptive treatment selection designs when multiple experimental treatment groups are set as a randomized parallel-group clinical trial.

Our simulation study showed that the adaptive enrichment design is beneficial when R is expected to be longer than F , while the non-enriched design is recommended when F is expected to be longer than R . For instance, the adaptive enrichment design is applicable in the setting in which progression-free survival, rather than overall survival, is set as a primary endpoint. Furthermore, the total trial period is particularly long when the distribution of recruitment is concave. In practice, the interim analysis is performed when a pre-specified number of events occurs, regardless of the recruitment pattern, and prolonged recruitment and follow-up periods are considered until the pre-specified number of events is obtained when the recruitment speed is slow. Hence, the application of the adaptive enrichment design is not encouraged when the interim analysis hinders recruitment until the interim decision is made in mid-course, regardless of whether the trial periods is prolonged.

Conclusions

We evaluated patient recruitment methods in the context of an adaptive enrichment design for developing targeted therapies. The adaptive enrichment design typically requires a longer total trial period than that of

the non-enriched design. As an alternative, the continued recruitment methods provide the option to continue recruiting patients from O or only from P even during the interim decision period when recruitment from O is halted. The performance of the continued recruitment methods demonstrates that they may enable adaptive enrichment designs to be applied to a broader range of clinical trials, particularly when R is expected to be longer than F .

Abbreviations

ALK: Anaplastic lymphoma kinase; C: Control; E: Experimental; EGFR: Epidermal growth factor receptor; F: Follow-up period; HER2: Human epidermal growth factor receptor 2; NSCLC: Non-small cell lung cancer; N: biomarker-negative population; O: Overall population; P: biomarker-positive population; R: Recruitment period; TSS: Targeted sample size

Acknowledgements

The authors thank the reviewers, Associate Editor, and Editor for their helpful comments and suggestions that have improved the paper.

Authors' contributions

RU devised the study and wrote the first draft. SY and AK provided overall guidance, including revisions and critical comments. All authors contributed to the interpretation of the results and approved the final version.

Funding

This research was partially supported by a Grant-in-Aid for Scientific Research (17K12649) from the Ministry of Education, Culture, Sports, Science and Technology of Japan. AK was funded partly by the Intramural Research Grant (27-8) for Neurological and Psychiatric Disorders of NCNP. The funders had no role in the design of the study and collection, analysis, and interpretation of data and in writing the manuscript.

Availability of data and materials

The datasets used and/or analysed during the current study are available from the corresponding author on request.

Ethics approval and consent to participate

Not applicable.

Consent for publication

Not applicable.

Competing interests

The authors declare that they have no competing interests.

Author details

¹Department of Biomedical Statistics and Bioinformatics, Kyoto University Graduate School of Medicine, 54 Kawahara-cho, Shogoin, Sakyo-ku, 606-8507 Kyoto, Japan. ²Biostatistics Department I, Data Science Division, A2 Healthcare Corporation, 1-4-12, Utsubohommachi, Nishi-ku, 550-0004 Osaka, Japan. ³Faculty of Medicine, Saga University, 5-1-1 Nabeshima, 849-8501 Saga, Japan.

Received: 27 July 2018 Accepted: 7 July 2019

Published online: 22 July 2019

References

- Aggarwal S. Targeted cancer therapies. *Nat Rev Drug Discov.* 2010;9(6):427–8.
- Kamel HFM, Al-Amodi HSAB. Exploitation of gene expression and cancer biomarkers in paving the path to era of personalized medicine. *Genomics Proteomics Bioinforma.* 2017;15(4):220–35.
- Mok TS, Wu YL, Thongprasert S, Yang CH, Chu DT, Saijo N, Sunpaweravong P, Han B, Margono B, Ichinose Y, Nishiwaki Y, Ohe Y, Yang JJ, Chewaskulyong B, Jiang H, Duffield EL, Watkins CL, Armour AA, Fukuoka M. Gefitinib or carboplatin–paclitaxel in pulmonary adenocarcinoma. *N Engl J Med.* 2009;361(10):947–57.

4. Solomon BJ, Mok T, Kim DW, Wu YL, Nakagawa K, Mekhail T, Felip E, Cappuzzo F, Paolini J, Usari T, Iyer S, Reisman A, Wilner KD, Tursi J, Blackhall F. First-line crizotinib versus chemotherapy in ALK-positive lung cancer. *N Engl J Med*. 2014;371(23):2167–77.
5. Baselga J, Cortés J, Kim SB, Im S-A, Hegg R, Im YH, Roman L, Pedrini JL, Pienkowski T, Knott A, Clark E, Benyunes MC, Ross G, Swain SM. Pertuzumab plus trastuzumab plus docetaxel for metastatic breast cancer. *N Engl J Med*. 2012;366(2):109–19.
6. Swain SM, Kim SB, Cortés J, Ro J, Semiglazov V, Campone M, Ciruelos E, Ferrero JM, Schneeweiss A, Knott A, Clark E, Ross G, Benyunes MC, Baselga J. Pertuzumab, trastuzumab, and docetaxel for HER2-positive metastatic breast cancer (CLEOPATRA study): overall survival results from a randomised, double-blind, placebo-controlled, phase 3 study. *Lancet Oncol*. 2013;14(6):461–71.
7. Swain SM, Baselga J, Kim SB, Ro J, Semiglazov V, Campone M, Ciruelos E, Ferrero JM, Schneeweiss A, Heeson S, Clark E, Ross G, Benyunes MC, Cortés J. Pertuzumab, trastuzumab, and docetaxel in HER2-positive metastatic breast cancer. *N Engl J Med*. 2015;372(8):724–34.
8. Fujitani K, Yang HK, Mizusawa J, Kim YW, Terashima M, Han SU, Iwasaki Y, Hyung WJ, Takagane A, Park DJ, Yoshikawa T, Hahn S, Nakamura K, Park CH, Kurokawa Y, Bang YJ, Park BJ, Sasako M, Tsujinaka T. Gastrectomy plus chemotherapy versus chemotherapy alone for advanced gastric cancer with a single non-curable factor (REGATTA): a phase 3, randomised controlled trial. *Lancet Oncol*. 2016;17(3):309–18.
9. Bauer P, Bretz F, Dragalin V, König F, Wassmer G. Twenty-five years of confirmatory adaptive designs: opportunities and pitfalls. *Stat Med*. 2016;35(3):325–47.
10. Jenkins M, Stone A, Jennison C. An adaptive seamless phase II/III design for oncology trials with subpopulation selection using correlated survival endpoints. *Pharm Stat*. 2011;10(4):347–56.
11. Lehmacher W, Wassmer G. Adaptive sample size calculations in group sequential trials. *Biometrics*. 1999;55(4):1286–90.
12. Hochberg Y. A sharper Bonferroni procedure for multiple tests of significance. *Biometrika*. 1988;75(4):800–2.
13. Marcus R, Peritz E, Gabriel KR. On closed testing procedures with special reference to ordered analysis of variance. *Biometrika*. 1976;63(3):655–60.
14. Schoenfeld D. The asymptotic properties of nonparametric tests for comparing survival distributions. *Biometrika*. 1981;68(1):316–9.
15. Brannath W, Zuber E, Branson M, Bretz F, Gallo P, Posch M, Racine-Poon A. Confirmatory adaptive designs with Bayesian decision tools for a targeted therapy in oncology. *Stat Med*. 2009;28(10):1445–63.
16. Friede T, Parsons N, Stallard N. A conditional error function approach for subgroup selection in adaptive clinical trials. *Stat Med*. 2012;31(30):4309–20.
17. Krisam J, Kieser M. Optimal interim decision rules based on a binary surrogate outcome for adaptive biomarker-based trials in oncology. *Stat Biopharm Res*. 2017;9(4):321–32.
18. Uozumi R, Hamada C. Interim decision-making strategies in adaptive designs for population selection using time-to-event endpoints. *J Biopharm Stat*. 2017;27(1):84–100.
19. Johnson NL, Kotz S, Balakrishnan N. *Continuous Univariate Distributions*, vol 1, 2nd ed. New York: Wiley; 1994.
20. Food and Drug Administration. *Guidance for industry: adaptive designs for clinical trials of drugs and biologics*, draft. Silver Spring, MD: U.S. Food and Drug Administration; 2018.
21. O'Brien PC, Fleming TR. A multiple testing procedure for clinical trials. *Biometrics*. 1979;35(3):549–56.
22. Lan KG, DeMets DL. Discrete sequential boundaries for clinical trials. *Biometrika*. 1983;70(3):659–63.
23. Gallo P, Mao L, Shih VH. Alternative views on setting clinical trial futility criteria. *J Biopharm Stat*. 2014;24(5):976–93.
24. Togo K, Iwasaki M. Optimal timing for interim analyses in clinical trials. *J Biopharm Stat*. 2013;23(5):1067–80.
25. Freidlin B, Othman M, Korn EL. Information time scales for interim analyses of randomized clinical trials. *Clin Trials*. 2016;13(4):391–9.
26. Counsell N, Biri D, Fraczek J, Hackshaw A. Publishing interim results of randomised clinical trials in peer-reviewed journals. *Clin Trials*. 2017;14(1):67–77.
27. Wang H, Rosner GL, Goodman SN. Quantifying over-estimation in early stopped clinical trials and the "freezing effect" on subsequent research. *Clin Trials*. 2016;13(6):621–31.
28. Benner L, Kieser M. Timing of the interim analysis in adaptive enrichment designs. *J Biopharm Stat*. 2018;28(4):622–32.

Publisher's Note

Springer Nature remains neutral with regard to jurisdictional claims in published maps and institutional affiliations.

Ready to submit your research? Choose BMC and benefit from:

- fast, convenient online submission
- thorough peer review by experienced researchers in your field
- rapid publication on acceptance
- support for research data, including large and complex data types
- gold Open Access which fosters wider collaboration and increased citations
- maximum visibility for your research: over 100M website views per year

At BMC, research is always in progress.

Learn more biomedcentral.com/submissions

