



## What is the expected benefit of patient-centric clinical development in oncology?

Alessandra Serra, Pavel Mozgunov, Thomas Jaki & Fabio Rigat

To cite this article: Alessandra Serra, Pavel Mozgunov, Thomas Jaki & Fabio Rigat (2022) What is the expected benefit of patient-centric clinical development in oncology?, Journal of Biopharmaceutical Statistics, 32:3, 414-426, DOI: [10.1080/10543406.2022.2065506](https://doi.org/10.1080/10543406.2022.2065506)

To link to this article: <https://doi.org/10.1080/10543406.2022.2065506>



© 2022 The Author(s). Published with license by Taylor & Francis Group, LLC.



[View supplementary material](#)



Published online: 18 Jul 2022.



[Submit your article to this journal](#)



Article views: 1576



[View related articles](#)



[View Crossmark data](#)

# What is the expected benefit of patient-centric clinical development in oncology?

Alessandra Serra <sup>a</sup>, Pavel Mozgunov <sup>a</sup>, Thomas Jaki <sup>a,b</sup>, and Fabio Rigat<sup>c</sup>

<sup>a</sup>MRC Biostatistics Unit, University of Cambridge, Cambridge, UK; <sup>b</sup>Department of Mathematics and Statistics, Lancaster University, Lancaster, UK; <sup>c</sup>Janssen Statistics & Decision Sciences, Janssen Pharmaceuticals, UK

## ABSTRACT

The identification and quantification of predictive biomarkers characterize personalized medicine approaches and patient-centric clinical development. In practice, the sponsor needs evaluating whether biomarker-informed clinical development strategies are more likely to benefit current and future patients. To this end, a simple metric is proposed and assessed here quantifying the expected clinical benefit (ECB) of clinical development programmes. Using simulation scenarios and endpoints relevant to oncology, the ECB of a simple biomarker-informed strategy is shown to be specific and sensitive. Also, the ECB difference is shown to increase in the biomarker-driven incremental efficacy and with the population prevalence of biomarker-positive study participants.

## ARTICLE HISTORY

Received 23 July 2021

Accepted 15 February 2022

## KEYWORDS

Biomarker; drug development; precision oncology; sequence of trials

## 1. Introduction

Progress in the measurement and understanding of the molecular landscape of cancer underpins the ever-evolving classification of tumour types (Kleihues and Leslie 2000) and motivates strategies investigating the safety and efficacy of novel agents in molecularly defined patient populations (Freidlin and Simon 2005; Freidlin et al. 2010; Beckman et al. 2011; Beckman and Chen 2013; Kelloff and Sigman 2012; Graf et al. 2015; Biankin et al. 2015; Fontes Jardim et al. 2015; Perez-Gracia et al. 2017; Wilson and Altman 2017; Liu et al. 2018; Huang and Hobbs 2019; Polley et al. 2019; Hu and Dignam 2019; Janiauda et al. 2019; Joshi et al. 2020). Use of predictive and prognostic markers in clinical development is correspondingly addressed in regulatory guidances, which stress the need for robust biomarker development and for statistically sound biomarker-informed design (FDA US Food and Drug Administration 2014, 2018a, 2018b, EMA European Medicine Agency 2017).

Notwithstanding the overwhelming momentum backing up clinical biomarker development, more work is needed to develop well-understood, specific and sensitive metrics comparing biomarker-informed versus biomarker-naïve strategies to inform rational and transparent investment decisions prior to and during clinical development. To this end, an expected clinical benefit (ECB) metric is defined and assessed here, building on the approaches of Beckman et al. (2011) and Biankin et al. (2015). The main goal of the ECB is to combine into one utility score the size of the target population, the efficacy gains and the technical probability of success of a candidate clinical development plan. For instance, if a drug is expected to be very promising only for a small group of patients, then a high probability of success of a study conducted in this group is not necessarily a useful measure of benefit as it does not reflect the societal value of drug development for the whole patient population. Also, the

**CONTACT** Alessandra Serra  [alessandra.serra@mrc-bsu.cam.ac.uk](mailto:alessandra.serra@mrc-bsu.cam.ac.uk)  MRC Biostatistics Unit, University of Cambridge, Cambridge CB2 0SR

 Supplemental data for this article can be accessed online at <https://doi.org/10.1080/10543406.2022.2065506>

© 2022 The Author(s). Published with license by Taylor & Francis Group, LLC.

This is an Open Access article distributed under the terms of the Creative Commons Attribution License (<http://creativecommons.org/licenses/by/4.0/>), which permits unrestricted use, distribution, and reproduction in any medium, provided the original work is properly cited.

probability of success alone does not measure development cost components, such as the expected number of trial participants who are not likely to benefit from the drug and the total number of patients enrolled in the programme.

As a parameter informing prospectively clinical development investment decisions, the ECB is fundamentally different from utility scores meant to evaluate the benefit-risk profile of candidate treatments for health technology assessment (Saint-Hilary et al. (2019)). Also, rather than deriving a utility score based on comparing a treatment's clinical benefit to its risk of toxicity, the decision rules implemented here ensure that the ECB measures the benefit of investigational treatments conditionally on their safety being no worse than that of a fixed comparator.

This paper proceeds as follows. An example of an oncology clinical trial setting is introduced in [Section 2](#) before a detailed description of the clinical development strategies that are considered in this study is provided in [Section 3](#). The ECB is proposed here to measure the expected clinical benefit of these strategies. We illustrate a performance evaluation of the proposed ECB in [Section 4](#). Future developments and final remarks are provided in [Section 5](#).

## 2. Example of a clinical trial setting

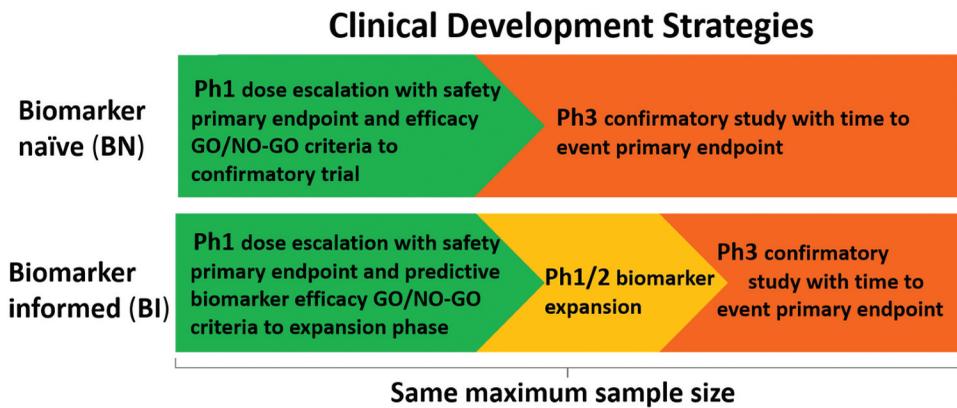
Clinical trial NCT01975519<sup>1</sup> was a combined phase Ib and IIa study of an anti-angiogenesis monoclonal antibody (TRC105) in combination with a tyrosine kinase inhibitor (pazopanib) in patients with advanced soft tissue sarcoma. The phase Ib study's objective was to evaluate the safety and tolerability of the investigational drug combination in dose-escalation. The objectives of phase IIa were to estimate the progression-free survival (PFS) of patients with advanced soft tissue sarcoma and the overall response rate in a cohort of patients with angiosarcoma. Trial NCT01975519 was then followed by an enrichment phase III trial in patients with advanced angiosarcoma (the TAPPAS trial, NCT02979899, Mehta et al. 2019) to demonstrate the superiority of TRC105 in combination with pazopanib versus pazopanib alone based on PFS. Patients were randomized with equal probability to each treatment group with stratification by angiosarcoma subgroups (cutaneous and non-cutaneous) and by the prior number of systematic therapies. An interim analysis was planned to assess whether to continue recruitment in the full population or in the cutaneous subgroup only.

These TRC105 studies provide a practical example of a biomarker-informed strategy with patient stratification and selection based on histology markers. This strategy can be compared with a biomarker-naïve development plan targeting the whole advanced soft tissue sarcoma population using the ECB definition proposed below. The strategy associated to the largest ECB would guide data generation and lead to the best expected outcome for clinical trial participants and for future patients. In the next section, the biomarker-naïve and biomarker-informed strategies are defined in more general terms and their associated ECBs are derived.

## 3. Clinical development strategies

A clinical development strategy (or plan) is a sequence of clinical trials, which aims to establish whether an investigational drug (ID) is safe and effective in a target population. Optimal strategies expose the smallest number of participants to unsafe or ineffective doses and maximize the probability of licensure of safe and effective drugs. For instance, for the clinical trial setting described in [Section 2](#), the sequence of clinical trials consists of a phase I study in a broad soft tissue sarcoma population, followed by a phase II study focusing on angiosarcomas and a phase III study focusing on angiosarcoma subtypes.

Following this example, the two development strategies illustrated in [Figure 1](#) are considered here. The top strategy is biomarker-naïve (BN), consisting of two trials in the overall patient population: a dose-escalation study (phase I) to establish the safety and clinical activity of the ID and a confirmatory study (phase III) to establish the incremental efficacy of the investigational compound with respect to standard of care. The bottom strategy in [Figure 1](#) is biomarker-informed (BI). This



**Figure 1.** Clinical development strategies compared in expected clinical benefit and expected sample size of development for each simulation scenario.

strategy consists of a phase I study assessing the safety and activity of the investigational compound in biomarker positive and biomarker negative subjects, followed by a phase I expansion phase establishing the biomarker-defined patient subgroup where efficacy is maximized, and a phase III study confirming the incremental efficacy of the investigational compound within this population subgroup versus standard of care. Note that biomarker status here is not thought of as being necessarily determined by a single marker, but rather as dichotomous summary of a potentially predictive biomarker signature.

### 3.1. Expected clinical benefit of biomarker-naïve and biomarker-informed development strategies

From a sponsor's perspective, benefit of development of an investigational compound ought to include safety and efficacy aspects relevant to licensure as well as the critical cost components involved in running the trials. For instance, Chen and Beckman (2009) used a benefit-cost ratio function to find the optimal go-no go criteria in late-stage drug development. Graf et al. (2015) proposed a utility function to quantify the achieved gains resulting from different outcomes of a trial. Huang and Hobbs (2019) proposed an approach to estimate the local benefit of biomarker-informed strategies using treatment response surfaces, selection rules and prognostic determinants. Saint-Hilary et al. (2019) proposed a benefit-risk metric overcoming the limitations of linear multi-criteria decision analysis.

The ECB proposed here covers factors that are both relevant to informing the sponsor's clinical development decisions and that also define the likelihood of benefit to patients: disease epidemiology, effect size of the investigational treatment and probability of licensure enabling access to treatment. This built-in alignment between the sponsor's and the patient utility components captured by the ECB is key to ensuring that investment decisions based on ECB rankings address genuine unmet medical needs.

Let  $b+$  and  $b-$  be the predictive baseline biomarker status and let  $P(b+)$  and  $P(b-)$  be, respectively, the prevalence of biomarker positive and of biomarker negative subjects in the patient population. In practice, robust estimates of biomarker prevalence in relevant target populations are available for clinically validated markers present in sufficiently large populations. Hence, we limit assessments of the ECB to scenarios with biomarker prevalence of at least 20%. Finally,  $P(L)$ ,  $P(L|b+)$ ,  $P(L|b-)$  are, respectively, the technical probabilities of licensure in the whole patient population, in its  $b+$  subset, or its  $b-$  subset. These probabilities may be thought of simply in terms of statistical significance – with no reference to effect size – or they may be quantified as conditional probabilities of exceeding a predefined minimal clinical effect. In this study, these probabilities are equal to zero if no safe

dose can be selected in phase I. Otherwise, the probability of success is the expected power of a randomized parallel arm phase III confirmatory study testing superiority of the investigational drug at the selected safe dose against standard of care. The ECB combining these three factors is

$$ECB = (\text{population coverage}) \times (\text{effect size}) \times (\text{probability of licensure}) \quad (1)$$

In Eqn. (1) the probability of licensure and the biomarker population incidence may be thought as weights of the treatment effect and they are indirectly relevant and less interpretable to patients. Instead, the magnitude of the effect size is the only factor measuring utility units directly relevant and meaningful to the patient. A more detailed discussion of alternative metrics is provided in the discussion section.

Motivated by the TAPPAS trial, we consider PFS as primary endpoint and its associated dRMST (Pak et al. 2017; Zhao et al. 2016) as effect size. The dRMST measures the difference in expected PFS between the confirmatory study arms and within the study follow-up. Hence, the dRMST is an optimal measure of effect size under quadratic loss. A natural alternative effect size estimator here is the difference in response rate (dORR) to the investigational drug and to standard of care. In what follows, we use the dRMST due to the relevance of standard time-to-event registrational endpoints in oncology. Under (1), the ECBs of the BN and of the BI strategies illustrated in Figure 1 are respectively,

$$ECB_{BN} = E(dRMST|L) \times P(L) \quad (2)$$

$$ECB_{BI} = E(dRMST|L, b+) \times P(L|b+) \times P(b+) + E(dRMST|L, b-) \times P(L|b-) \times P(b-). \quad (3)$$

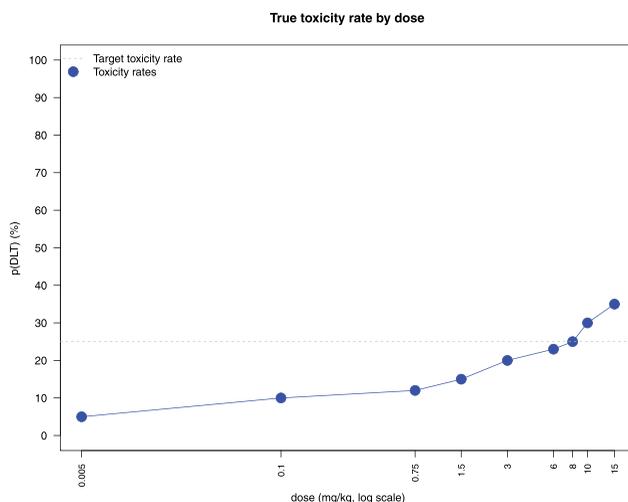
## 4. Simulation study

The goal of this simulation study is to assess the specificity and sensitivity of (2) and (3) under data generation scenarios selected to favour  $ECB_{BN}$  or  $ECB_{BI}$  respectively. In each scenario, the choice of the parameters is informed by historical clinical trial data and simulations are used to make decisions and robustness assessment.

Full details about the statistical models and parameter values underpinning the data simulations are presented in Appendix A.1 and A.2.

### 4.1. Simulation setup

Motivated by the phase Ib study NCT01975519 and by similar immuno-oncology studies, dose-escalation data were simulated for a set of nominal dose levels (0.005, 0.1, 0.75, 1.5, 3, 6, 8, 10, 15 mg/kg) with true toxicity rates depicted in Figure 2 using a maximum of 40 subjects enrolled in dose cohorts of four each. A standard modified continual reassessment method with overdose control (mCRM EWOC; Sweeting et al. 2013) was used to guide dose escalation with target toxicity rate of 25%. Following Sweeting et al. (2013), the prior mean values for the logistic regression  $\beta_1$  and  $\beta_2$  were set, respectively, to  $-1.09$  and  $0.25$ , their variances were  $1.26$  and  $3.92$  and their prior correlation was null. Also, the probability of dose-dependent toxicity was assumed independent of biomarker status. Data were simulated under an approximate 1:1 stratification of the  $b+$  and  $b-$  subjects. Efficacy responses were generated using a logistic regression with dose as explanatory variable for the BN strategy and biomarker status, dose level and their interaction for the BI strategy. Efficacy was assumed to be absent in  $b-$  participants and, depending on the scenario, absent or increasing monotonically in dose in  $b+$  participants.



**Figure 2.** True toxicity rates by doses. The target DLT rate used here is 25%.

The maximum dose that is safe by mCRM-EWOC and shows an estimated response rate greater than 33% was selected to proceed to the next phase of development. For the BI strategy, this consists of an expansion phase establishing efficacy at the selected dose in the b+ subgroup only. The decision of whether to progress to the final confirmatory trial is based on a one-sided chi-square test p-value at threshold 0.025.

In the final stage of the simulation study, time to progression data were simulated using an exponential regression model with exponential link function and expected value proportional to biomarker status. When efficacy was present in b+ participants, their expected time to progression was thus proportionally longer. Maximum sample sizes of 100, 200, 300, 400 and 500 patients were considered for the confirmatory trial and patients are randomised with equal probability to investigational treatment or to standard of care. A maximum of 40% of the confirmatory study sample sizes was reserved for the biomarker expansion phase in the BI strategy. Approximately 30% events were censored at random at the time of analysis, which was assumed to occur 36 months into the phase III study. To define the licensure probability  $P(L)$ , we use a log-rank test p-value with significance threshold 0.05.

Simulations were run for different scenarios using biomarker prevalences  $P(b+) = 20\%$ , 50%, 80%, reflecting a range where biomarker-specific development would be realistically appealing.

Strong, weak, null or inferior biomarker-mediated efficacy simulation scenarios are depicted in Figures S.1-S.8 and described in detail in Table S.1 in the Supplementary Material. The true response rates at all doses are, respectively, all below or equal to the 33% comparator threshold under the inferior and null scenarios. The expected time to progression is, respectively, shorter or identical to that of the comparator in the inferior and null scenarios. The true dRMST, the difference in median PFS and the hazard ratio between the investigational drug and the comparator are, respectively,  $-2.1$ ,  $-1.5$ ,  $2.3$  under the inferior scenario and 0, 0 and 1 under the null scenario.

Under the weak and strong efficacy scenarios, response rates increase in dose in the biomarker positive population prevalence, while they are constant in dose and below 33% in biomarker negative subjects. Tables S.2-S.3 in the Supplementary Material describe the true dRMST, the corresponding difference in median PFS and the hazard ratio between the investigational drug and the comparator in the b+ subpopulation under the weak and strong efficacy scenarios for each dose level.

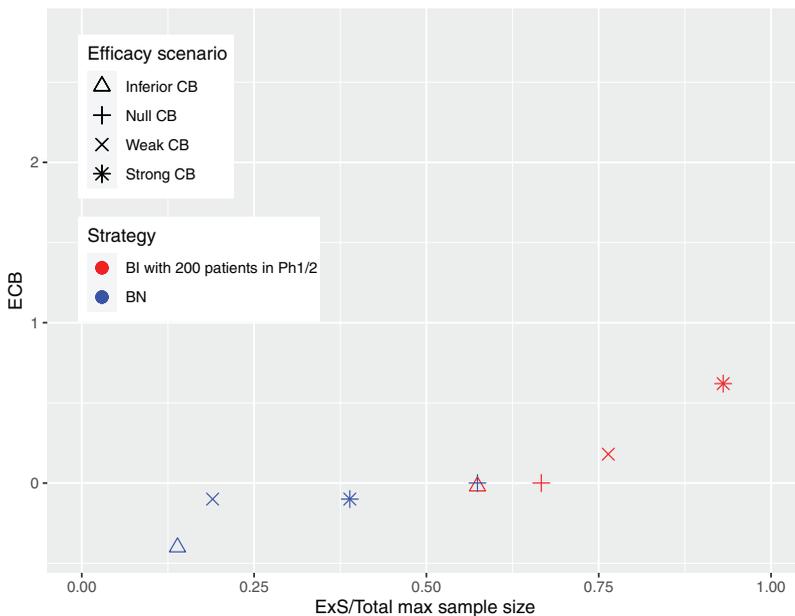
For each strategy, we measured the cost of development as the expected proportion of patients (ExS) enrolled over the total common sample size, and the ECB as defined in (2) and (3). The ExS can be seen here as a measure evaluating the feasibility of the study. All simulations were run using the software R (<https://cran.r-project.org/>).

### 4.2. Numerical results

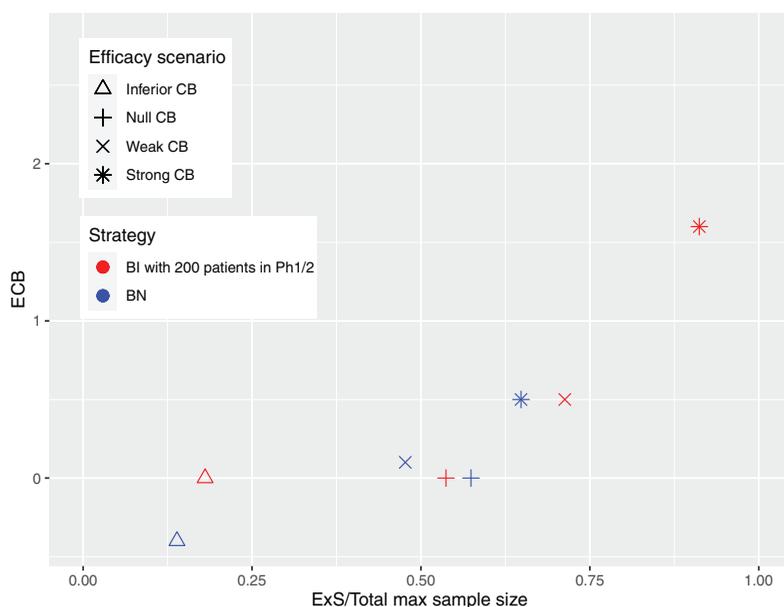
The maximum sample size used in the confirmatory study here is 500 participants and 40% subjects who would be enrolled in the confirmatory study was reserved for the BI strategy expansion phase. Figures S.9-S.23 in the Supplementary Material show the simulation results for different sizes of the BI expansion phase and different confirmatory study sample sizes.

Figures 3, 4, 5 show the expected proportion of participants enrolled (ExS) versus the ECB of each development strategy (vertical axes). Table S.9 in the Supplementary Material reports the corresponding proportions of simulations that did not progress to phase III due to failure in estimating logistic regression efficacy parameters in the phase I study, or due to lack of fit of the logistic regression model to the simulated efficacy data in phase I, or to no dose showing an estimated response rate greater than the 33% comparator threshold.

Figure 3 illustrates the simulation results when the prevalence of b+ patients is relatively low (20%). In this setting, the BN strategy consistently shows negative or null ECB in all scenarios. In contrast, the BI strategy shows a positive ECB in the weak and strong efficacy scenarios (0.25–0.5 months on the dRMST scale over 36 months follow-up). In addition, Table S.9 shows that under low biomarker prevalence both strategies suffer from numerical issues, due to few b+ data points being available in phase I for reliably estimating the dose-response curve. Table S.9 also shows that the proportions of simulations failing to progress to phase III for futility are always greater for the BN strategy compared to the BI strategy in all scenarios except for the null. Under the null scenario, the BI strategy is likely to



**Figure 3.** Expected clinical benefit versus expected cost (expressed as ExS/total max sample size in the whole trial) of clinical development strategies for each scenario when  $P(b+) = 20\%$ , the maximum sample size for the whole trial consists of 540 patients and 200 patients are enrolled in the expansion phase in the BI strategy.



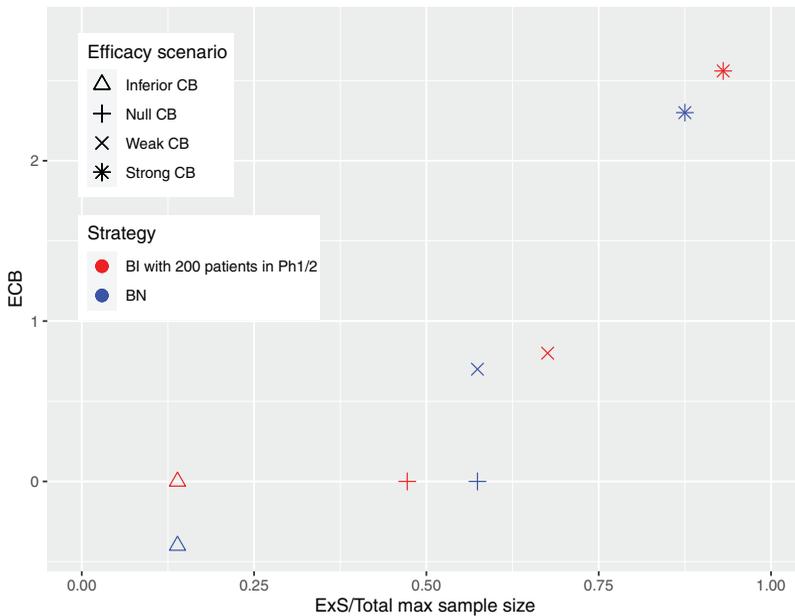
**Figure 4.** Expected clinical benefit versus expected cost (expressed as ExS/total max sample size in the whole trial) of clinical development strategies for each scenario when  $P(b+) = 50\%$ , the maximum sample size for the whole trial consists of 540 patients and 200 patients are enrolled in the expansion phase in the BI strategy.

stop earlier compared to the BN strategy because any dose is found to be effective and safe after Phase I–II. This difference could be the result of incorrect failing to stop under the BN strategy due to a biased estimate of the response rate in the whole population. Overall, these simulations show that under low  $b+$  prevalence the BI strategy has greater sensitivity – reflected in a larger ECB – than BN in detecting the presence of true biomarker-mediated incremental efficacy, but the BN strategy stops development earlier in absence of efficacy.

Figure 4 shows that the ECBs of both strategies are negative or null in absence of biomarker-driven incremental efficacy when half of the population is assumed to be biomarker positive. The ECB of the BI strategy is strictly greater than that of the BN strategy in all biomarker-driven efficacy scenarios (0.5–1.3 months on the dRMST scale), this difference being significantly greater than under the low biomarker prevalence scenario. Consistently with this difference in sensitivity between strategies, Table S.9 shows that the BN strategy incorrectly stops development in presence of biomarker-mediated efficacy more often than BI. In addition, Table S.9 shows that stops due to numerical issues occur in both strategies to a lesser extent than in the low prevalence  $b+$  simulation scenario.

Figure 5 depicts the simulation results when 80% of the population is  $b+$ . The ECB of the BI strategy is again greater than that of the BN strategy under the biomarker-mediated efficacy scenarios, with the ECB difference being less than 0.5 months on the dRMST scale. As for the 20% and 50% biomarker prevalence scenario, the expected number of participants enrolled under the BI strategy is consistently larger than that of the BN strategy in presence of biomarker-mediated efficacy. Finally, the proportions of stops in Table S.9 under high  $b+$  prevalence scenario are similar to those reported for the low prevalence scenario, showing again that the BN strategy stops development more often than the BI strategy in presence of efficacy in the  $b+$  group, with both strategies suffering from numerical issues.

Overall, the ECB and the expected number of subjects enrolled in clinical trials were shown to be consistently greater for the BI strategy compared to BN under all considered biomarker-driven efficacy scenarios. When efficacy of the investigational compound was either the same or inferior to that of the



**Figure 5.** Expected clinical benefit versus expected cost (expressed as ExS/total max sample size in the whole trial) of clinical development strategies for each scenario when  $P(b+) = 80\%$ , the maximum sample size for the whole trial consists of 540 patients and 200 patients are enrolled in the expansion phase in the BI strategy.

standard of care, the ECB of both strategies was less or equal to zero. These results demonstrate that the proposed ECB metric correctly ranked the two clinical development strategies in all analysed scenarios.

### 4.3. Sensitivity analyses

Additional analyses were conducted to assess the robustness of the ECB estimates under alternative simulation scenarios of practical relevance. Tables S.4, S.5 and S.6 in the Supplementary Material show the results when any dose can be selected in the first phase and when no interim analysis is planned in the confirmatory study. The results show that the ECB correctly ranks the BI and BN strategies when a 50% target response rate is used to trigger progression beyond dose escalation and in presence of lower true toxicity rates. The ExS here depends on whether the strategies are selecting a dose that is safe and efficacious to proceed to Phase III. In this case, the BI strategy is usually more likely to progress to phase III compared to the BN strategy.

Further simulations were also run using fixed safe doses for the expansion and confirmatory phases, thus bypassing dose escalation, and when a pre-planned interim analysis is used to assess efficacy and enable early stop of the phase III trial. The objective of these simulations was to confirm sensitivity and specificity of the ECB not only prior to starting a development program, but also after having observed the results of part of the planned studies. Specific details of these simulations are described in Appendix A.3. Table 1 shows the results of these further simulations when the maximum safe and efficacious dose (8 mg/kg) is used for the BI dose expansion and for the phase III confirmatory trial and when an interim analysis is performed at 50% information fraction in phase III. The ECB again correctly ranks the BN and BI strategies, but here the BI strategy requires a smaller expected sample size compared to the BN strategy in the biomarker-driven efficacy scenarios. In this setting, the ExS does only depend on whether the interim analysis provides sufficient information to declare that the

**Table 1.** Simulation results when the true toxicity rates for each dose level are as represented in Figure 2 and the RR of the comparator is set to 33%. A total of 540 patients are enrolled for each strategy and an interim analysis at half of the population is planned in the confirmatory study. Dose 8 mg/kg is recommended after phase I. For the BI strategy, a total of 200 patients are enrolled in the expansion phase. ECB: expected clinical benefit. ExS: expected sample size.

<b>P(b+)</b>	<b>Strategy</b>	<b>ECB</b>	<b>ExS</b>	<b>Scenario</b>
20%	BN	0.0	536.0	Strong
	BI	0.9	290.7	
	BN	-0.1	531.3	Weak
	BI	0.8	296.4	
	BN	0.0	537.8	Null
	BI	0.0	537.5	
	BN	-5.5	293.2	Inferior
	BI	0.0	363.5	
50%	BN	1.6	407.5	Strong
	BI	2.3	290.6	
	BN	0.9	469.1	Weak
	BI	1.9	296.4	
	BN	0.0	537.8	Null
	BI	0.0	537.4	
	BN	-5.5	293.3	Inferior
	BI	-0.1	363.5	
80%	BN	3.4	290.8	Strong
	BI	3.5	290.6	
	BN	2.8	298.4	Weak
	BI	3.0	296.6	
	BN	0.0	537.5	Null
	BI	0.0	537.8	
	BN	-5.5	293.5	Inferior
	BI	-0.1	362.5	

selected dose is efficacious compared to the standard of care. In this case, the BI strategy results more efficient – in terms of the expected number of patients that are enrolled – than the BN because it is likely to declare efficacy earlier when this is truly present.

Tables S.7 and S.8 in the Supplementary Material, respectively, show that the ECB correctly ranks the two strategies when the minimum safe and active dose in the weak efficacy scenario (3 mg/kg) and 8 mg/kg are assumed safe and efficacious, and no interim analysis is performed in the phase III study.

## 5. Discussion

An ECB metric was proposed to measure clearly and comprehensively the expected benefit of clinical development and to rank alternative development strategies in an oncology setting. However, alternative approaches can be considered for measuring ECB. For example, the effect size in Eqn. (1) could be replaced with the probability that the effect size is greater than a minimum clinical effect of interest, or more specifically replacing the expected dRMST with the probability  $P(dRMST > \text{minimum clinical significant survival gain})$ . This would then lead to the contributions of the three factors in determining estimates of the ECB being balanced as all three multiplied terms will be on a probability scale and the resulting measure would range between 0 and 1. Highlighting this property of the alternative, we also note that such a definition might require further assumptions and can lead to some communication challenges, compared to the proposed measure. Including this probability term will require specification of the minimal clinically meaningful difference. In the survival setting, this can be challenging as prolonging survival by 1 day can be

argued to be meaningful and hence agreement on this measure can be challenging to achieve. This approach may also be less clear to patients compared to the proposed metric where the magnitude of the dRMST is the most relevant and meaningful factor for patients.

An alternative form of the metric could be obtained via standardisation of the treatment effect in the proposed metric (as it is commonly done in the quantitative HTA methods, for example, by the multi-criteria decision analysis (MCDA), Mussen et al. 2007, Saint-Hilary et al. 2017). Via the standardisation, the treatment effect will be forced to be between 0 and 1, i.e. on the same scale as the rest of the terms. At the same time, we note that our proposed measure can be seen as more favourable choice. We argue that the magnitude of the effect size itself is useful because dRMST is the only factor measuring utility units directly relevant to the patient. Instead, the other two terms – probability of licensure and the biomarker population incidence – are indirectly relevant and less interpretable to patients and they may be thought as weights of the treatment effect.

In the simulation study, the ExS was also measured in addition to the ECB metric. While the ECB metric is relevant from the perspective of a future patient, the ExS can be seen as a measure evaluating the feasibility of the study. Thus, the ExS is needed in practice, but one can argue that a positive/negative development decision should depend on ECB only. Presenting both these metrics is thought to provide full information to sponsors on both the expected clinical benefit and the feasibility of the study in order to offer them the possibility to evaluate which strategy is the most appropriate and feasible and should be prioritised.

Overall, the proposed ECB metric was shown to be specific and sensitive in data simulations of practical relevance. No expected benefit was detected under the inferior and null efficacy scenarios and the biomarker-informed strategy is correctly preferred to the biomarker naïve strategy when efficacy is truly biomarker-driven. Sensitivity analyses showed analogous results when varying the number of patients enrolled in the BI strategy expansion phase and in the phase III trial.

The parameters used in the simulations were informed by historical immune-oncology studies. However, one could argue that historical estimates could be used directly as a tool for decision-making. What is more challenging is to then assess the robustness of these conclusions (i.e. undertake sensitivity analysis) using the same framework. Using simulations means that one can use the same principled approach to make decisions and robustness assessment.

In all scenarios, the maximum likelihood estimates of the logistic regression model coefficients could not be calculated for all simulated phase I data because of numerical convergence issues due to data separation (Albert and Anderson 1984). Separation takes place when a linear function of the covariates generates perfect predictions. In the scenarios used here, separation occurred when dose and biomarker status were strongly collinear. Regularization of the maximum likelihood estimates of the logistic regression coefficients (Firth 1993; Heinze and Schemper 2002) can address this weakness. For instance, figures S.25, S.26, S.27 in the Supplementary Material show that use of Firth's penalized regression as implemented in the R package *logistf* (Heinze et al. 2020) did not show the numerical issues noted under maximum likelihood and the ranking between the BN and BI strategies was the same as when using the maximum likelihood estimates.

Notwithstanding the breadth of the simulation scenarios explored here, there is clear scope for further work. Specifically, no difference in safety was considered here in relation to biomarker status. Also, simulations were run using a 30% censoring rate in the confirmatory study. However, the log-rank test statistic is sensitive to the number of patients that are censored. Additional work is also needed to derive probability intervals of the ECB, as only point estimates were provided here. Finally, development strategies compared via the ECB may be extended well beyond what was considered here, as in practice a greater number of study designs, multiple endpoints and more complexity statistical methods are routinely used in oncology as well as in other therapeutic areas. Despite its current limitations, the simple framework proposed here can be further extended to encompass additional

factors relevant clinical development planning in other therapeutic areas and thus to guide the selection of development strategies in practice by combining transparently their clinical, epidemiological and statistical dimensions.

## Note

1. <https://clinicaltrials.gov/ct2/show/NCT01975519>

## Acknowledgments

The Authors wish to thank the Editor and Referees for their insightful comments on earlier versions of this paper. FR wishes to thank Sudhakar Rao for his support during the development of this work.

This report is independent research supported by the National Institute for Health Research (NIHR Advanced Fellowship, Dr Pavel Mozgunov, NIHR300576) and by the NIHR Cambridge Biomedical Research Centre (BRC-1215-20014). The views expressed in this publication are those of the authors and not necessarily those of the NHS, the National Institute for Health Research or the Department of Health and Social Care (DHSC). T Jaki and P Mozgunov also received funding from UK Medical Research Council (MC\_UU\_00002/14). FR was employee of Janssen during the development of this work.

## Disclosure statement

No potential conflict of interest was reported by the authors.

## Funding

This work was supported by the National Institute for Health Research [NIHR300576]; NIHR Cambridge Biomedical Research Centre [BRC-1215-20014]; UK Medical Research Council [MC\_UU\_00002/14].

## ORCID

Alessandra Serra  <http://orcid.org/0000-0001-8431-5154>

Pavel Mozgunov  <http://orcid.org/0000-0001-6810-0284>

Thomas Jaki  <http://orcid.org/0000-0002-1096-188X>

## References

- Albert, A., and J. Anderson. 1984. *Biometrika. On the Existence of Maximum Likelihood Estimates in Logistic Regression Models* 71:1–10.
- Beckman, R., J. Clark, and C. Chen. 2011. *Nature reviews drug discovery. Integrating Predictive Biomarkers and Classifiers into Oncology Clinical Development Programmes* 10:735–748.
- Beckman, R., and C. Chen. 2013. *Chinese journal of cancer. New evidence-based adaptive clinical trial methods for optimally integrating predictive biomarkers into oncology clinical. Development Programs* 32 (5):233–241.
- Biankin, A. V., S. Piantadosi, and S. J. Hollingsworth. 2015. *Nature. Patient-centric Trials for Therapeutic Development in Precision Oncology* 526 (7573):361–370.
- Chen, C., and R. A. Beckman. 2009. *Journal of Biopharmaceutical Statistics. Optimal Cost-effective Designs of Phase II Proof of Concept Trials and Associated Go-no Go Decisions* 19:424–436.
- European Medicine Agency. 2017. *Concept paper on predictive biomarker-based assay development in the context of drug development and lifecycle.*
- Firth, D. B. 1993. *Biometrika. Reduction of Maximum Likelihood Estimates* 80:27–38.
- Fontes Jardim, D. L., M. Schwaederle, C. Wei, J. J. Lee, D. S. Hong, A. M. Eggermont, R. L. Schilsky, J. Mendelsohn, V. Lazar, and R. Kurzrock. 2015. *JNCI: Journal of the National Cancer Institute. Impact of A Biomarker-Based Strategy on Oncology Drug Development: A Meta-Analysis of Clinical Trials Leading to FDA Approval* 107 (11):djv253.
- Freidlin, B., and R. Simon. 2005. *Clinical Cancer Research. Adaptive Signature Design: An Adaptive Clinical Trial Design for Generating and Prospectively Testing A Gene Expression Signature for Sensitive Patients* 11 (21):7872–7878.
- Freidlin, B., W. Jiang, and R. Simon. 2010. *The Cross-Validated Adaptive Signature Design. Clinical Cancer Research* 16 (2):691–698.

- Graf, A. C., M. Posch, and F. Koenig. 2015. *Biometrical Journal. Adaptive Designs for Subpopulation Analysis Optimizing Utility Functions* 57 (1):76–89.
- Heinze, G., and M. Schemper. 2002. *Statistics in Medicine. A Solution to the Problem of Separation in Logistic Regression* 21:2409–2419.
- Heinze, G., M. Ploner, D. Dunkler, H. Southworth, and L. Jiricka. 2020. *logistf: Firth's Bias-Reduced Logistic Regression*, R package version 1.24.
- Hu, C., and J. J. Dignam. 2019. *JCO Precis Oncology. Biomarker-Driven Oncology Clinical Trials: Key Design Elements, Types, Features, and Practical Considerations* 1:1-12.
- Huang, M., and B. P. Hobbs. 2019. *Statistical Methods in Medical Research. Estimating Mean Local Posterior Predictive Benefit for Biomarker-guided Treatment Strategies* 28 (9):2820–2833.
- Janiauda, P., S. Serghioua, and J. P. A. Ioannidis. 2019. *Cancer treatment reviews. New Clinical Trial Designs in the Era of Precision Medicine: An Overview of Definitions, Strengths, Weaknesses, and Current Use in Oncology* 73:23–30.
- Joshi, N., C. Nguyen, and A. Ivanova. 2020. *Journal of Biopharmaceutical Statistics. Multi-stage Adaptive Enrichment Trial Design with Subgroup Estimation* 30 (6):1038–1049.
- Kelloff, G. J., and C. C. Sigman. 2012. *Nature reviews drug discovery. Cancer Biomarkers: Selecting the Right Drug for the Right Patient* 11:201–214.
- Kleihues, P., and H. S. Leslie. 2000. *Cancer. World Health Organization Classification of Tumors* 88:12.
- Liu, J., T. Lichtenberg, K. A. Hoadley, L. M. Poisson, A. J. Lazar, A. D. Cherniack, A. J. Kovatich, C. C. Benz, D. A. Levine, and A. V. Lee, et al. 2018. *The Cancer Genome Atlas Research Network. An Integrated TCGA Pan-Cancer Clinical Data Resource to Drive High-Quality Survival Outcome Analytics* 1732:400–416
- Mehta, C., L. Liu, and C. Theuer. 2019. *Ann Oncol. An Adaptive Population Enrichment Phase III Trial of TRC105 and Pazopanib versus Pazopanib Alone in Patients with Advanced Angiosarcoma (TAPPAS Trial)* 30 (1):103–108.
- Mussen, F., S. Salek, and S. Walker. 2007. *Pharmacoepidemiology and drug safety. A Quantitative Approach to Benefit-risk Assessment of Medicines – Part 1: The Development of A New Model Using Multi-criteria Decision Analysis* 16:S2–S15.
- O'Brien, P. C., and T. R. Fleming. 1979. *Biometrics. A Multiple Testing Procedure for Clinical Trials* 35:549.
- Pak, K., U. BPharm, H. Kim, D. H. L. Tian, R. C. Kane, M. Takeuchi, H. Fu, B. Claggett, and L. Wei. 2017. *JAMA Oncology. Interpretability of Cancer Clinical Trial Results Using Restricted Mean Survival Time as an Alternative to the Hazard Ratio* 3 (12):1692–1696.
- Perez-Gracia, J. L., M. F. Sanmamed, A. Bosch, A. Patiño-García, K. A. Schalper, V. Segura, J. Bellmunt, J. Tabernero, C. J. Sweeney, T. K. Choueiri, et al. 2017. *Cancer treatment reviews. Strategies to Design Clinical Studies to Identify Predictive Biomarkers in Cancer Research* 53:79–97.
- Polley, M. C., E. L. Korn, and B. Freidlin. 2019. *JCO Precis Oncology. Phase III Precision Medicine Clinical Trial Designs that Integrate Treatment and Biomarker Evaluation* 3:1-9.
- Saint-Hilary, G., S. Cadour, V. Robert, and M. Gasparini. 2017. *Biometrical Journal. A Simple Way to Unify Multicriteria Decision Analysis (MCDA) and Stochastic Multicriteria Acceptability Analysis (SMAA) Using A Dirichlet Distribution in Benefit-risk Assessment* 59 (3):567–578.
- Saint-Hilary, G., V. G. Robert, T. Jaki, and P. Mozgunov. 2019. *Statistical Methods in Medical Research. A Novel Measure of Drug Benefit-risk Assessment Based on Scale Loss Score* 28:2738–2753.
- Sweeting, M., A. Mander, and T. Sabin. 2013. *Journal of statistical software. Bcrm: Bayesian Continual Reassessment Method Designs for Phase I Dose-Finding Trials* 54:1-26.
- US Food and Drug Administration. 2014. Expedited programs for serious conditions-drugs and biologics. <https://www.fda.gov/regulatory-information/searchfda-guidance-documents/expedited-programs-serious-conditionsdrugs-and-biologics>
- US Food and Drug Administration. 2018a. Biomarker Qualification: Evidentiary Framework. <https://www.fda.gov/regulatory-information/search-fda-guidancedocuments/biomarker-qualification-evidentiary-framework>
- US Food and Drug Administration. 2018b. *Clinical Trial Endpoints for the Approval of Cancer Drugs and Biologics*. <https://www.fda.gov/regulatory-information/search-fdaguidance-documents/clinical-trial-endpoints-approval-cancer-drugsand-biologics>
- Van der, V. 1998. *Asymptotic Statistics*. (Cambridge Series in Statistical and Probabilistic Mathematics). Cambridge: Cambridge University Press. doi:10.1017/CBO9780511802256
- Wilson, J. L., and R. B. Altman. 2017. *Experimental Biology and Medicine. Biomarkers: Delivering on the Expectation of Molecularly Driven, Quantitative Health* 243 (3):313–322.
- Zhao, L., B. Claggett, L. Tian, H. Uno, M. A. Pfeffer, S. D. Solomon, L. Trippa, and L. J. Wei. 2016. *Biometrics. On the Restricted Mean Survival Time Curve in Survival Analysis* 72:215–221.

## Appendices

### A.1 BN strategy

In phase I, the response data are simulated considering the biomarker status. Logistic regression is used to model the dose-response curve ignoring the baseline biomarker status. If the model fits the data adequately, then the asymptotic posterior distribution of regression coefficients is found as per the Bernstein-von Mises theorem (Van der 1998). If at least one dose is identified which response rate is greater than the fixed control rate (set as 33%), then the study proceeds to phase III with the maximum dose that is safe and effective. In the confirmatory study, the PFS model is simulated and the power of the log-rank test is computed using several sample sizes. At the end of the sequence of studies, the expected clinical benefit is computed as

$$ECB_{BN} = E(dRMST \times \text{power of logrank test} \times 1_{\{\text{dose is safe}\}}) \quad (4),$$

where the dRMST estimate is based on the Kaplan–Meier curves and  $1_{\{\text{dose is safe}\}}$  takes value one at doses that are not truly toxic continue to phase III and zero otherwise.

### A.2 BI strategy

In phase I, the response data are simulated using dose and biomarker status. A logistic regression is used to model the dose-response curve considering the baseline biomarker status, the doses and the interaction between them as independent variables. If the model fits the data adequately, then the asymptotic posterior distribution of the regression coefficients is found as per the Bernstein-von Mises theorem (Van der 1998). If the point estimate of the response rate in the b+ population for at least one dose is greater than that of the b- population and greater than the response rate of the control (set as 33%), then the study proceeds to phase III with the maximum safe and effective dose. In the confirmatory study, the power of the biomarker cohort analysis is computed using the chi-square test. Moreover, the PFS model is simulated and the power of the log-rank test is computed using several sample sizes. The expected clinical benefit is computed as

$$ECB_{BI} = E(dRMST \times \text{power of logrank test} \times \text{power of chisquare test} \times 1_{\{\text{dose is safe}\}}|b+) \times P(b+) \quad (5)$$

### A.3 Estimation of the ECB when a specific dose is recommended and an interim analysis is planned in the confirmatory studies.

A dose is selected to proceed to the expansion phase in the BI strategy and to the confirmatory studies for both strategies. A maximum number of 500 patients are enrolled in the confirmatory studies and an interim analysis is performed after acquiring 50% of the information fraction. The PFS model is simulated and if the power of the log-rank test at the interim analysis is below the O'Brien and Fleming (1979) p-value threshold, then the trial stops. Otherwise, the study can proceed to the final analysis. In the BI strategy, the power of the biomarker cohort analysis is computed using the chi-square test. The ECB is computed as described in (4) and (5) and the expected sample size (ExS) is computed in each simulation as

$$ExS = N.\text{subjects prior to Phase III} + (N.\text{subjects at Phase III interim}) \times P(\text{stop at interim}) \\ + (N.\text{subjects at Phase III final}) \times (1 - P(\text{stop at interim})).$$