A modular framework for early-phase seamless oncology trials

Philip S. Boonstra[*], Thomas M. Braun, and Elizabeth C. Chase

Department of Biostatistics, University of Michigan, USA

Abstract

Background: As our understanding of the etiology and mechanisms of cancer becomes more sophisticated and the number otf therapeutic options increases, phase I oncology trials today have multiple primary objectives. Many such designs are now 'seamless', meaning that the trial estimates both the maximum tolerated dose and the efficacy at this dose level. Sponsors often proceed with further study only with this additional efficacy evidence. However, with this increasing complexity in trial design, it becomes challenging to articulate fundamental operating characteristics of these trials, such as (i) what is the probability that the design will identify an acceptable, i.e. safe and efficacious, dose level? or (ii) how many patients will be assigned to an acceptable dose level on average? **Methods:** In this manuscript, we propose a new modular framework for designing and evaluating seamless oncology trials. Each module is comprised of either a dose assignment step or a dose-response evaluation, and multiple such modules can be implemented sequentially. We develop modules from existing phase I/II designs as well as a novel module for evaluating dose-response using a Bayesian isotonic regression scheme. **Results:** We also demonstrate a freely

---

[*]1415 Washington Heights, Ann Arbor, Michigan, USA, 48109-2029; Tel: +1 734 615 1580; philb@umich.edu

available R package called `seamlesssim` to numerically estimate, by means of simulation, the operating characteristics of these modular trials. **Conclusions:** Together, this design framework and its accompanying simulator allow the clinical trialist to compare multiple different candidate designs, more rigorously assess performance, better justify sample sizes, and ultimately select a higher quality design.

keywords: phase I; phase II; dose escalation; recommended phase II dose

1 Background

Current phase I oncology trial designs are often 'seamless' in nature [1], meaning both a maximum tolerated dose (MTD) and an initial efficacy estimate at this dose level are sought. Trial sponsors may be unwilling to proceed with further study without this additional efficacy evidence [2]. We recently designed such a trial for relapsed/refractory indolent lymphoma. The objective was to estimate the MTD and obtain preliminary estimates of efficacy as measured by complete response. Success for this trial reflects the seamless perspective: the primary objective is to identify a dose level that is both tolerable <u>and</u> efficacious. As in all such trials, a key statistical concern is the potential for uncertainty to compound across objectives. Future success in the pipeline depends upon the current trial correctly estimating both the MTD and the response rate at this estimated MTD.

Our understanding of both the etiology and mechanisms of many cancers has increased over the past decades, and the number of therapeutic options has correspondingly increased. However, one characteristic remaining relatively static is that most trial designs are still variants of the '3+3' algorithm [3, 4]. This is notable given that the canonical reference for this design recommends <u>against</u> its use [5]. The continued use of the 3+3 instead of demonstrably superior alternatives such as the continual reassessment method [CRM, 6] highlights a persistent gap between method and practice in this area, which is widened by the multifaceted objectives of seamless oncology trials.

We argue that this gap can be bridged with an accessible framework for seamless trial design and the software for simulating such designs. To that end, this paper has two primary contributions. First, we develop a modular perspective for seamless oncology trial design, which decomposes complex designs with multiple decision points into simpler components. This atomic approach individualizes the choices required for

seamless designs, thus providing a framework to evaluate the effect of each choice on overall performance. Second, we present a seamless trial design simulator to conduct this performance assessment, implemented in the R statistical environment [7–10] as a freely available package called `seamlesssim` [11]. Users can evaluate the impact of each design choice with respect to important operating characteristics.

Many model-based designs for recommending dose levels based upon safety and efficacy have been proposed, including the bivariate CRM [bCRM, 12], 'EffTox' [13], the trivariate CRM [TriCRM, 14], and others [15–26]. Some of these designs have even been implemented in trial practice, like the use of an EffTox design to assess the performance of sitravatinib and nivolumab in renal cell cancer [27] or a trial of FLAG-IDA and ponatinib in myeloid leukemia [28], or the use of the bivariate CRM in a trial on T cell co-stimulators for solid tumors [29]. However, these examples aside, the uptake of these advanced designs has been slow, and protocols often try to satisfy these dual objectives under the 3+3 framework. For example, a trial may estimate an MTD using the 3+3 and then, under the assumption that it has been correctly identified, enroll an additional 'expansion cohort(s)' of subjects at a fixed dose level(s) and/or in different subpopulations in order to evaluate efficacy [30–33]. An extreme example of this is the KEYNOTE-001 trial of pembrolizumab, which started as a 3+3 trial of 10 subjects across three dose levels and subsequently evolved into a multi-amendment trial enrolling a total of 1235 subjects across multiple expansion cohorts [34, 35]. As of 2016, there were at least four other phase I trials enrolling more than 1000 patients [36].

The favorable outcome of KEYNOTE-001 was unusual: a recent study found that only 3.4% of phase I oncology trials that opened between Jan 1, 2000 and Oct 31, 2015 led to eventual approval [37], which is consistent with reported success rates from an earlier study [38]. Stopping dose-escalation after a small number of subjects, as in the 3+3, risks ending at too low of a dose [31]; the subsequent expansion may stop for lack

4

of efficacy, when a higher dose level might have been found to be efficacious. Or escalation might have occurred too quickly, and lower dose levels, in addition to being safer, are equally efficacious, which would only be revealed through a more methodical exploration of efficacy across the full range of dose levels. This piecemeal approach to seamless trial design makes difficult a formal assessment of operating characteristics and highlights the challenges of quantifying the implications of multiple sequential decision trials.

Recognizing these weaknesses, the FDA published draft guidance on designing seamless oncology trials [1, 39]. However, because the FDA does not make design recommendations, this guidance does not resolve the tension between logistical simplicity on the one hand and a rigorous quantification of uncertainty on the other. To that end, this paper outlines a modular framework that captures the decision points made in designing a seamless oncology trial and gives several possible options for each decision. Section 2 describes each module and presents our specific design choices intended to adhere to the FDA guidance. We present a freely available trial simulator, implemented in an R package called `seamlesssim` and an accompanying vignette, both available via GitHub [11], allowing the user to evaluate and compare many design options at once. Section 3 demonstrates this simulator with a numerical study. We show that sensible choices for each module exist, and we provide the computational tools for clinical trialists and statisticians to make this assessment.
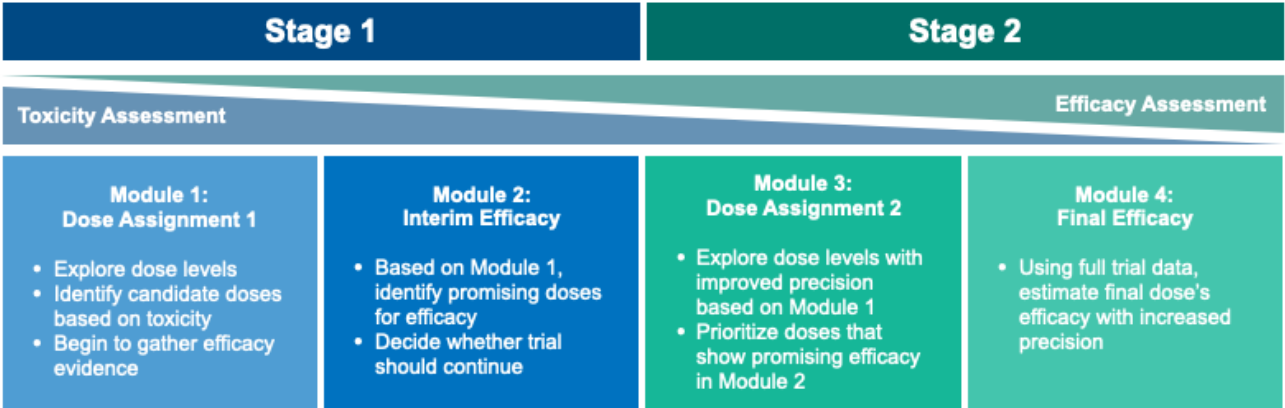
Figure 1: A high-level schematic of our two-stage, four-module framework for Phase I/II seamless designs. Note that toxicity and efficacy are assessed concurrently in both stages, rather than a stark divide between the two objectives, with toxicity the priority in Stage 1, while efficacy is the focus of Stage 2. Trials may choose to stop after Stage 1 if the Module 2 interim efficacy analysis is not promising.

2 Methods

2.1 Conceptual Framework

Our framework encompasses many styles of early phase, seamless-type trials. We allow for two stages, both enrolling cohorts of subjects to an assigned dose level(s) and following for toxicity or response endpoints; the second stage may be dropped if a certain response threshold is not achieved in the first stage. Each stage is subdivided into two 'modules', one for dosing subjects and one for assessing efficacy, for up to four modules total. Figure 1 provides a high-level schematic of our framework.

This paper defines a 'seamless design' to be any valid combination of the four modules. Since each module can be selected independently of the remaining modules, with some exceptions, there are many possible designs. Some modules can be skipped altogether, and thus our schema includes both classical phase I dose-escalation designs, such as the 3+3 or the CRM, and two-stage phase II designs [40]. It also yields novel hybridizations of the two, and we elucidate the benefits and pitfalls of these

hybridizations. The specific choices for each module that we have considered are summarized in Table 1. Throughout this document, we use a `typewriter font family` to refer to the simulator and its inputs.

Remark 1 Our "1st stage" and "2nd stage" terminology are not synonymous with the traditional phase 1 and phase 2 terminology. Rather, the seamless framework means that phase 1 and phase 2 occur throughout both stages. As we will show, the staging protects patients by allowing to stop for futility.

The data to be collected for each subject are dose level, occurrence of dose limiting toxicity (DLT), and indicator of response. The true MTD is the dose level with true probability of DLT closest to some pre-specified threshold, usually between 0.25 and 0.35. The objective of seamless oncology trials is typically to establish that the true response probability at the true MTD exceeds some minimal efficacy boundary. Well-designed seamless trials should guard against two distinct errors: the MTD may be incorrectly estimated and/or the estimated response rate may be inefficient or biased. Either error decreases the probability of correctly estimating the efficacy at the true MTD, and moreover, the consequences of either error may be different.

Remark 2 Because our goal is simplicity, the modules currently implemented assume that the probabilities of DLT and response for each dose are mutually independent. However, additional modules that relax this assumption could be implemented.

2.2   Module 1: Dose Assignments for 1st Stage

An initial cohort of n_1 subjects is enrolled and followed for DLT and/or response, which are assumed to be binary. Subjects may be adaptively assigned to one of several dose levels based upon the observed dose-toxicity relationship in the previous subjects, or all subjects in this module may be enrolled to a fixed dose level, and the sole primary

	Stage 1		Stage 2	
	Module 1	Module 2	Module 3	Module 4

	`crm`
	`$n`
	`$skeleton`
	`$starting_dose`
	`$beta_scale`
	`$dose_cohort_size`
	`$dose_cohort_size_first_only`
	`$earliest_stop`
		`bayes`
		`$prob_threshold`
		`$prob_means`
		`$prior_n_per`
			`crm`
			`$n`
			`$skeleton`[†]
			`$beta_scale`[†]
			`$dose_cohort_size`
			`$dose_cohort_size_first_only`
			`$earliest_stop`
				`bayes`
				`$prob_threshold`
				`$prob_means`
				`$prior_n_per`
				`$include_stage1_data`

	`3pl3`
	`$starting_dose`
		`bayes_isoreg`
		`$prob_threshold`
		`$alpha_scale`
			`3pl3`	`bayes_isoreg`
				`$prob_threshold`
				`$alpha_scale`
				`$include_stage1_data`

	`empiric`
	`$n`
	`$starting_dose`
	`$rule`
	`$first_patient_look`
	`$thresh_decrease`
		`inverted_score`
		`$ci_level_onesided`
			`empiric`
			`$n`
			`$rule`
			`$first_patient_look`
			`$thresh_decrease`
				`inverted_score`
				`$ci_level_onesided`
				`$include_stage1_data`

	`fixed`
	`$n`
	`$starting_dose`
		`min_num_resp`
		`$number`
			`fixed`
			`$n`
			`$include_stage1_data`
				`min_num_resp`
				`$number`

		`min_pct_resp`
		`$percent`
			`continume_crm`[‡]
			`$n`
				`min_pct_resp`
				`$percent`
				`$include_stage1_data`

		`none`	`none`	`none`

Table 1: Module options currently implemented in the seamless design trial simulator. The option names are in **bold**, and the names immediately below each option, which follow the '$' symbol, give the inputs required for that option.
† these arguments are optional if module 1 is `crm`. If this is the case and these arguments are not provided, the module 1 options will be used.
‡ this option can only be selected if module 1 is `crm` because this indicates a simple continuation of that crm

endpoint is response. Depending on the aims of the trial, a dose-modifying or fixed-dose module is possible.

We recommend the model-based CRM for module 1 [6, 10] (`crm`, Table 1). In this design, the first subject is assigned to a pre-specified starting dose. A statistical model is fit to estimate the DLT probabilities for each dose based upon the first subject's data; this model is usually fit using Bayesian methods to account for sparsity in the data. The MTD

is estimated after the first subject, and the next subject is assigned to the new estimated MTD. With each subsequent enrollment, the MTD is re-estimated using all previous subjects' data.

There are usually safety constraints accompanying the CRM to avoid unsafe escalation [41]. These include never escalating more than one dose level per subject, never escalating before at least one subject has completed a full DLT window at the current dose, and stopping the trial for toxicity if no dose falls below the toxicity threshold.

We have implemented other options for module 1 in our simulator, including the 3+3 algorithm (3p13); a de-escalation scheme that lowers the assigned dose whenever some percentage of subjects experience a DLT at the current dose level (empiric); and a fixed dose-assignment scheme (fixed). None of these alternatives explicitly estimate the MTD but instead simply assign dose levels according to a rule-based algorithm.

2.3   1st Stage Futility Analysis (Module 2)

Due to the longer timeframe typically required to observe response, it may be infeasible to re-evaluate the dose-response curve after each subject as is done with the dose-toxicity curve. Rather, it is common to assess efficacy at a few discrete time points during the trial. Module 2 demarcates the first such time point in our framework. Module 1 ends after dose assignments have been made for n_1 subjects, and module 2 occurs when response has been measured for them.

Module 2 is a gatekeeper: if the probability of finding evidence for efficacy at the end of the trial, given the data thus far, is low, then the trial should stop. If module 1 assigned all n_1 subjects to the same dose level, then module 2 can be as simple as requiring a minimum number of responders in order to proceed (min_num_resp). In the

absence of any subsequent modules, this would be the single-stage design of A'hern [42]. If dose assignments vary from subject to subject, such that the actual number of subjects at the estimated MTD is likely to be less than n_1, then it would make more sense to require a minimum proportion of responders in order to proceed (`min_pct_resp`). Even this may be insufficient, as simply requiring a proportion does not take into account uncertainty in observed response rates; this can be achieved by calculating a confidence interval for the probability of response at the estimated MTD, e.g. based on the inverted score test [`inverted_score` 43].

We also implement a Bayesian version (`bayes`, Table 1), which, assuming that the efficacy at the estimated MTD is a random variable, calculates the posterior probability that efficacy at the MTD exceeds some minimal response rate. The trial continues only if this posterior probability exceeds some prespecified value. It is common to assume that the distribution of the random variable is Beta, with hyperparameters $a_1 > 0$ and $a_2 > 0$, which can be linked to prior beliefs about the mean historical response probability, $a_1/(a_1 + a_2)$, and the historical effective sample size, $a_1 + a_2$. This is akin to the prior elicitation approach of Mayo and Gajewski [44].

One limitation facing all of the above approaches when the dose-assignment scheme in module 1 is adaptive is that the number of subjects at the estimated MTD can be highly variable. As an extreme example, suppose that, after the n_1th subject, the dose-assignment mechanism updates the estimated MTD to a never-before-tried dose level. If no dose-response model is in place, and assuming a Bayesian perspective, then the distribution of the MTD's efficacy will be entirely prior-driven. This is undesirable behavior if there is evidence that the response rate at the dose level just above the estimated MTD is itself quite low, the knowledge of which would actually suggest stopping for futility. A parametric dose-response model, e.g. logistic regression or a CRM-type model, could help overcome this limitation. Alternatively, one could impose a

weaker assumption that the true probability of response is simply non-decreasing between dose levels. To this end, our recommendation for module 2 is a novel Bayesian variant of isotonic regression that we have developed (`bayes_isoreg`, Table 1), which we briefly describe in Supplement 1. In standard isotonic regression, observing a large response rate at a lower dose level would be taken as evidence that the response rates at high dose levels are also large and, conversely, observing a small response rate at a higher dose level is taken as evidence that the response rate at low dose levels is also small [45]. Standard isotonic regression may be inadequate under sparse data scenarios, and our Bayesian variant addresses this limitation.

2.4   Dose Assignments for 2nd Stage (Module 3)

Seamless trials often enroll a second cohort of n_2 subjects, either using an adaptive dose-assignment scheme or at a fixed dose level, sometimes called a 'dose expansion cohort' [DEC, 30]. Module 3, which delineates this next set of dose assignments, can be any of the options used in module 1, but the starting dose is now the estimated MTD from module 1. An expansion cohort starting at the estimated MTD with or without a simple de-escalation rule would be simulated using `empiric` or `fixed`, respectively (Table 1).

Because the estimated MTD is subject to uncertainty, we recommend that module 3 continue adaptive dose assignments and re-estimation of the MTD [31, 32]. Under this recommendation, an entirely new CRM could be initiated, starting at the previously estimated MTD (`crm`), or the CRM from module 1 could resume at the exact state it ended at in module 1 (`continue_crm`). Alternatively, a new 3+3 starting at the estimated MTD could be selected (`3pl3`). Finally, module 3 can be skipped altogether if no second stage is required (`none`).

11

2.5 Dose Recommendation (Module 4)

A final efficacy analysis is conducted at the end of the trial, either using all enrolled subjects' data or data from just the second stage. The design choices for module 4 are the same as for module 2, and all of the same options exist (`bayes`, `bayes_isoreg`, `inverted_score`, `min_num_resp`, and `min_pct_resp`).

3   Results

3.1   Trial Simulator: Description

We present a numerical study of our seamless trial simulator to demonstrate how it can be used to compare the performance of heterogenous seamless trial designs. Code and instructions for recreating the simulation study, which makes use of our `seamlesssim` R package, are available in a separate GitHub repository [46], and the instructions are reprinted in Supplement 2. Ten data-generating scenarios were considered; the true dose-toxicity and dose-response curves are given with the results in Figures 2–4 and Figure S2 in Supplement 1. For all scenarios, the true MTD was the dose level with true probability of DLT closest to 0.25 but not exceeding 0.30, and an 'acceptable' dose level was defined as any dose level with true probability of DLT not exceeding 0.30 and true probability of response at least 0.20. Based on this definition, two scenarios (numbers 7 and 9) had no acceptable dose levels, and three scenarios (1, 3, and 10) had two acceptable dose levels. The remaining five had exactly one acceptable dose level.

We compared six designs comprising combinations of modules: two choices of dose-assignment mechanism each combined with three different efficacy analyses. All designs are described in Table 2. The first dose-assignment mechanism is a '3+3' algorithm (module1 = `3p13`), which will have a maximum sample size of 6 patients per

Design name	module 1	module 2	module 3	module 4	$\max n_1 + n_2$
3pl3:Bayes:DEC:Bayes	3pl3	bayes	empiric	bayes	65
3pl3:none:DEC:Bayes	3pl3	none	empiric	bayes	65
3pl3:none:DEC:BIsoReg	3pl3	none	empiric	bayes_isoreg	65
CRM:Bayes:CRM:Bayes	crm	bayes	crm	bayes	60
CRM:none:CRM:Bayes	crm	none	continue_crm	bayes	60
CRM:none:CRM:BIsoReg	crm	none	continue_crm	bayes_isoreg	60

Table 2: Choice of modules for six designs considered. The upper three designs will enroll up to $n_1 = 30$ patients (cohorts of six for each of five dose levels) in module 1 and $n_2 = 35$ in module 3; the lower three designs will enroll up to $n_1 + n_2 = 60$ subjects across both modules.

dose level, or $n_1 = 30$ patients, followed by a fixed DEC of size $n_2 = 35$ (module3 = empiric); the DEC enrolls subjects at the estimated MTD from the '3+3' algorithm and only de-escalates if the observed DLT rate after the first 10 patients in the cohort exceeds 33%. The second dose-assignment mechanism is a CRM (module1 = module3 = crm) with $n_1 = 25$ and $n_2 = 35$ subjects in the first and third modules, respectively.

We evaluated three sets of efficacy analyses. The first utilized a Bayesian futility analysis (module2 = bayes) and a final Bayesian analysis (module4 = bayes); the second conducted only a final Bayesian analysis (module4 = bayes); and the third implemented our Bayesian isotonic regression at the end (module4 = bayes_isoreg). To compare between all six designs, the module 2 and module 4 efficacy parameters were set so that the probability of recommending the MTD in scenario 9 was about 0.10. This corresponds to a type I error (false positive) rate of 0.10, since the MTD is not acceptable in this scenario. For each of the 60 combinations of six designs and ten scenarios, we ran 2000 simulated trials.

3.2 Trial Simulator: Results

Figures 2 and 3 display the simulation-based probability of recommending any dose and whether or not that recommendation was a good one. Each figure gives a set of three

designs, grouped by choice of dose-assignment mechanism. When at least one dose level is acceptable, as in scenarios 1–6, 8, and 10, there are three possible outcomes to the trial – no recommendation, recommendation of an unacceptable dose, or recommendation of an acceptable dose – and when no dose levels are acceptable, as in scenarios 7 and 9, there are two possible outcomes – no recommendation or recommendation of an unacceptable dose.

From Figure 2, which contains the 3+3-type designs, whether or not there is an interim futility analysis in module 2, i.e. comparing 3pl3:Bayes:DEC:Bayes to 3pl3:none:DEC:Bayes, has relatively little impact on the probability of making a correct dose recommendation. This is not surprising, given that the DEC in module 3 is unlikely to change the estimated MTD unless extreme toxicity is observed, making this comparison analogous to that of a two-stage phase II design against a one-stage phase II design. In contrast, the use of a Bayesian isotonic regression in module 4 in place of independent beta priors on each dose level, i.e. comparing 3pl3:none:DEC:Bayes to 3pl3:none:DEC:IsoReg, generally increases the overall probability of recommending a dose level. That is, the 'No Rec (correct)' and 'No Rec (wrong)' proportions are generally larger for 3pl3:none:DEC:Bayes, and the 'Rec (unaccept)' and 'Rec (accept)' proportions are generally larger for 3pl3:none:DEC:BIsoReg. These differences range from 1–4% across all ten scenarios. This is also an expected finding, since sharing information between dose levels allows the isotonic regression to more easily recommend dose levels.

Figure 3 gives the same sets of results for the CRM-type designs, all of which have the same maximum possible sample size of $n_1 + n_2 = 60$ subjects. As expected, the CRM-type designs select an acceptable dose with higher probability than the 3+3-type designs. Moreover, in contrast to Figure 2, the interim efficacy analysis has a greater effect on the final outcome of the trial, as seen from comparing CRM:Bayes:CRM:Bayes

14

to CRM:none:CRM:Bayes. This is likely because of the final 35 patients of the trial, who are only enrolled to CRM:Bayes:CRM:Bayes if the module 2 efficacy analysis is passed. Comparing CRM:none:CRM:Bayes to CRM:none:CRM:BIsoReg, the isotonic regression generally increases the recommendation probability.

Figure 4 gives the distribution of dose assignments at subjects 10, 30, and 50 for the three CRM-type-designs. Analogous results for the 3+3-type-designs are given in Figure S2 of Supplement 1. For the CRM designs, as the trial proceeds, subjects are assigned to the preferred dose level in increasing proportions. Although the 3+3 designs generally assign the 10th subject to the preferred dose with high probability, their performance degrades as the trial proceeds, meaning that the 30th and 50th subjects are often assigned to the preferred dose level with small probabilities. This is an expected result given the known deficiencies of the 3+3, but the inclusion of the interim efficacy analysis would make it difficult to quantify these probabilities in the absence of this simulator. Comparing Figure 4 to Figure S2, CRM:none:CRM:BIsoReg is uniformly preferred to 3pl3:Bayes:DEC:Bayes except for scenarios 7 and 9, in which no dose levels are acceptable in truth. Based on this, we would conclude that CRM:none:CRM:BIsoReg would be improved with a choice of module 2 = `bayes_isoreg`, if we were actually proposing the design for a clinical protocol.

4   Conclusion

This paper characterizes seamless oncology trial design in a modular framework, which confers several advantages. It formalizes each decision needed for a seamless-type trial, including a provision for interim futility analyses. It puts forward a concrete definition of a 'seamless trial', namely any logically coherent set of modules. Because this modular framework is implemented in `seamlesssim`[11], a freely available R package and an
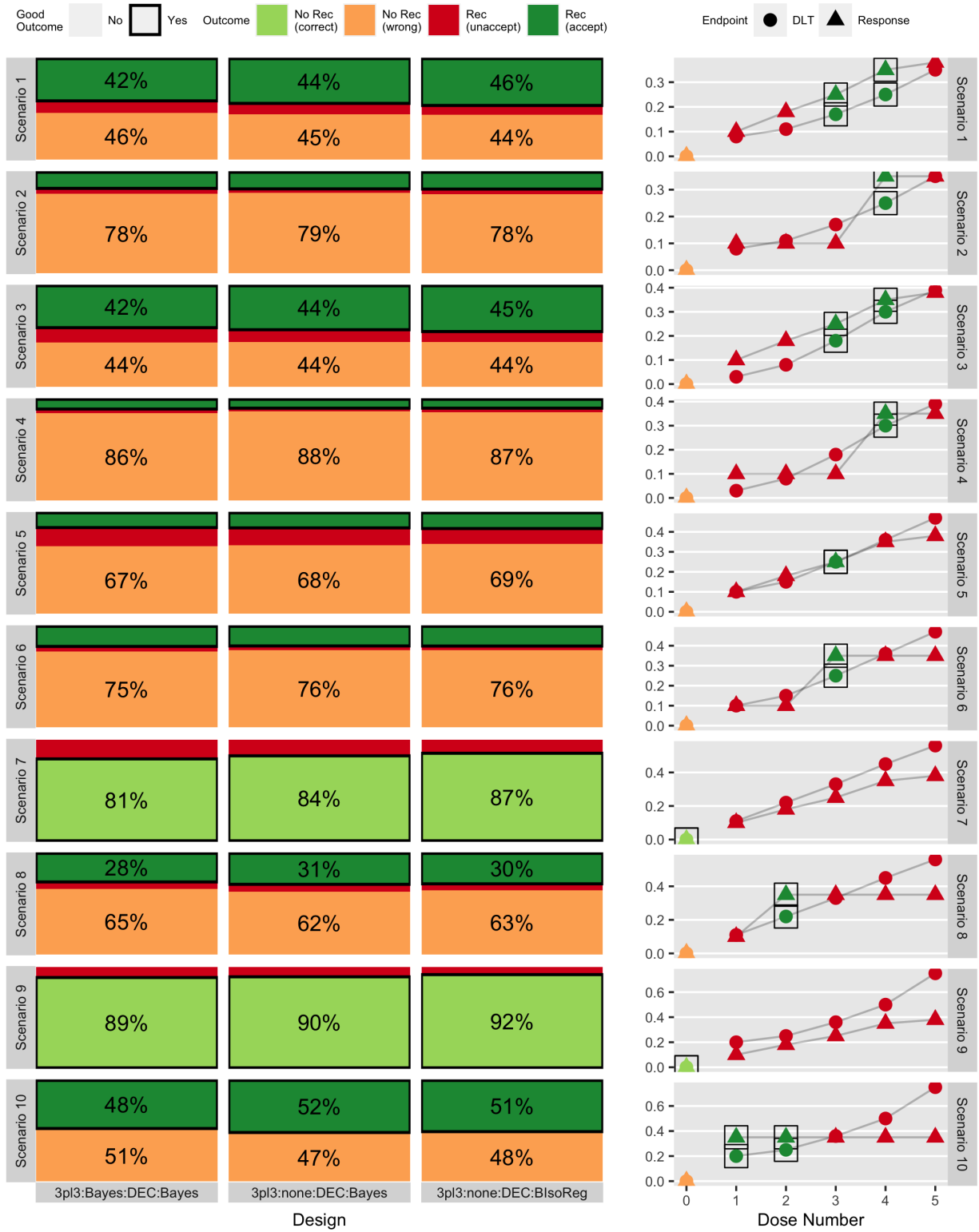
Figure 2: Simulation-based probability of outcomes (based on 2000 simulations) for three designs (out of six total) across ten scenarios, separated into whether or not any dose level was recommended and whether or not that outcome was acceptable or not. The right-hand column gives the true generating toxicity and efficacy curves. Each row corresponds to a different scenario, and consecutive pairs of scenarios (1&2, 3&4, etc.) are linked in that they share a common dose-toxicity curve but differ in the dose-response curve. Figure 3 gives the same results for three different designs.
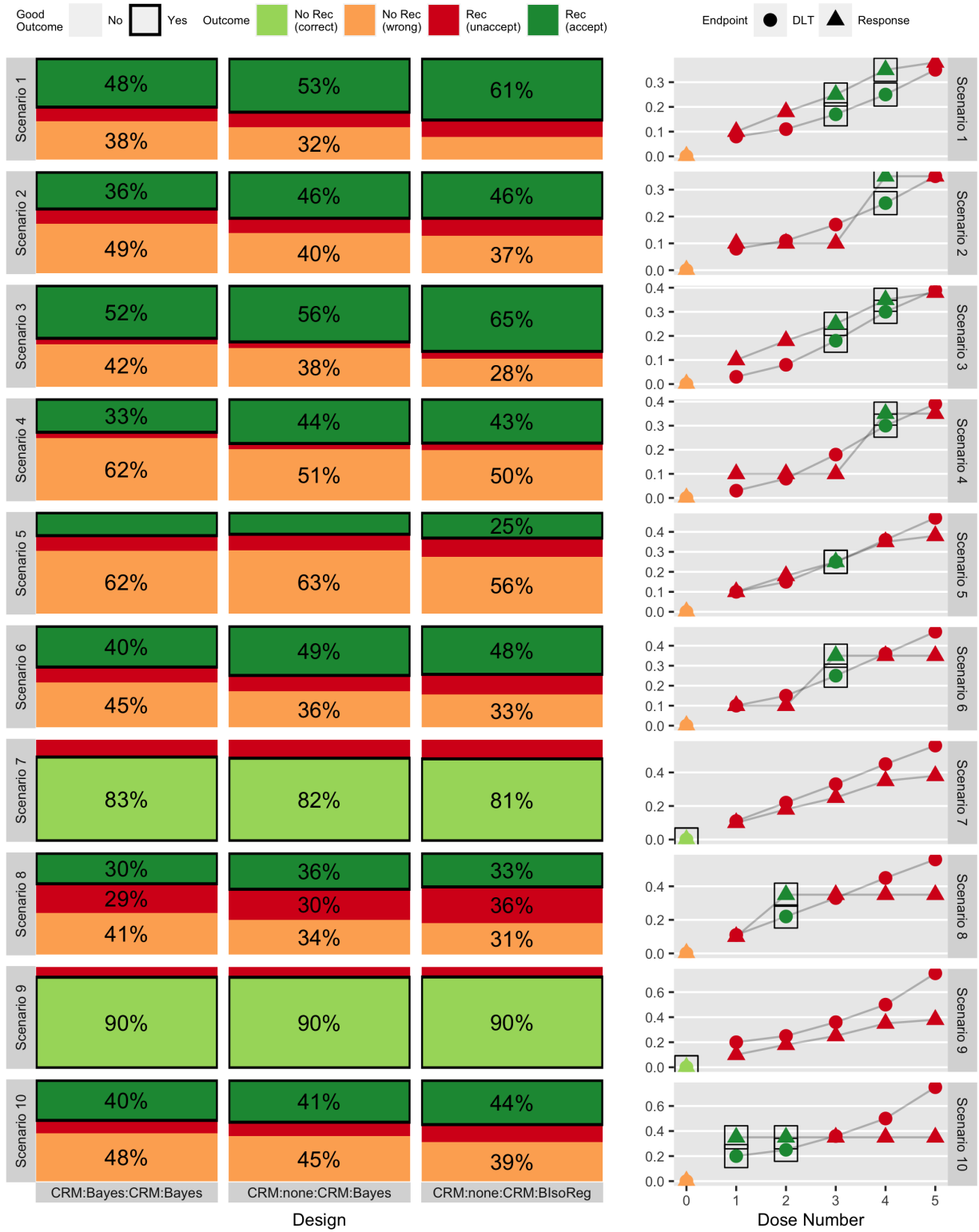
16

Figure 3: Simulation-based probability of outcomes (based on 2000 simulations) for three designs (out of six total) across ten scenarios, separated into whether or not any dose level was recommended and whether or not that outcome was acceptable or not. The right-hand column gives the true generating toxicity and efficacy curves. Each row corresponds to a different scenario, and consecutive pairs of scenarios (1&2, 3&4, etc.) are linked in that they share a common dose-toxicity curve but differ in the dose-response curve. Figure 2 gives the same results for three different designs.
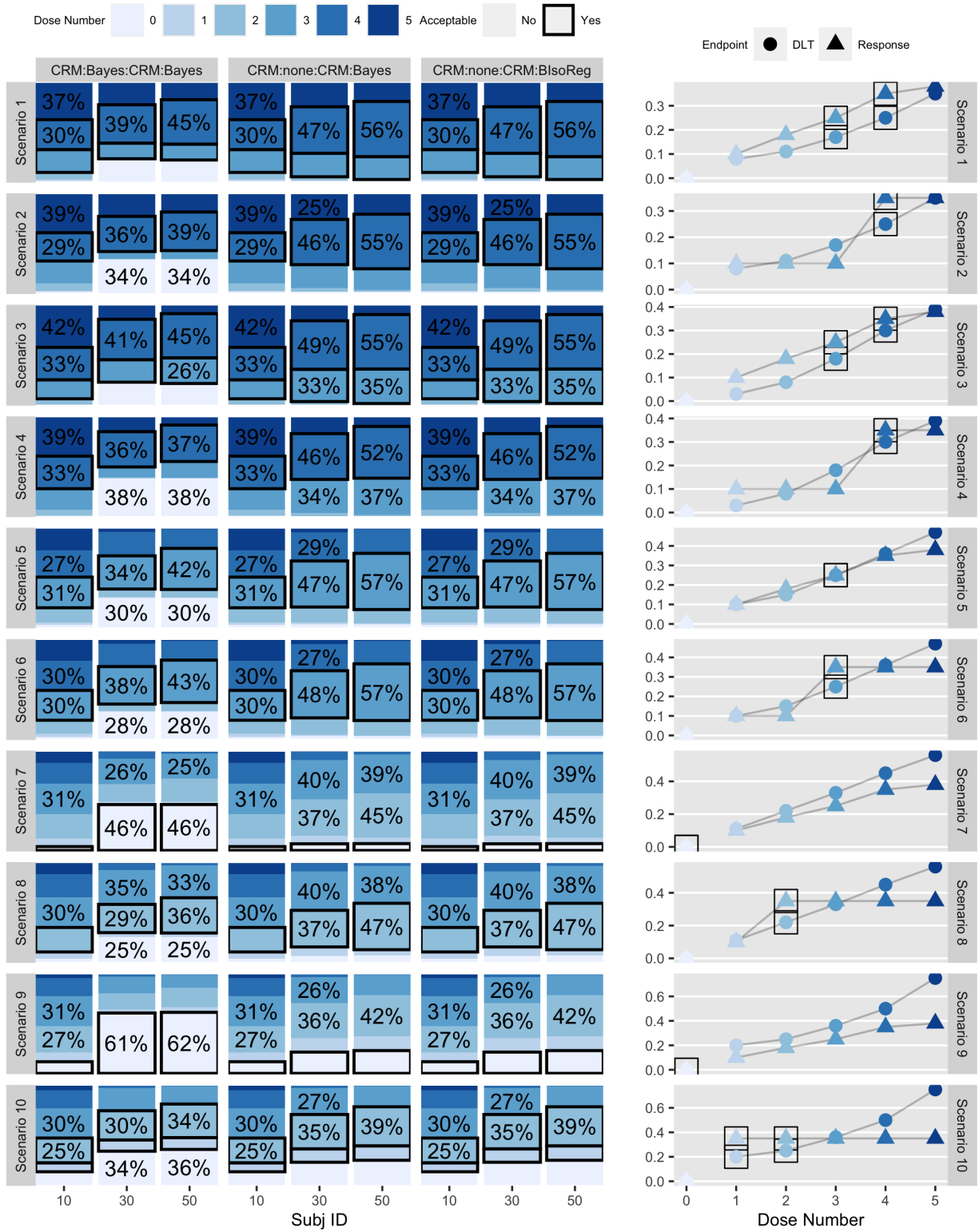
17

Figure 4: Distribution of dose assignments at subjects 10, 30, and 50 (based on 2000 simulations) for three CRM-type designs (out of six total) across ten scenarios. The right-hand column gives the true generating toxicity and efficacy curves. Each row corresponds to a different scenario, and consecutive pairs of scenarios (1&2, 3&4, etc.) are linked in that they share a common dose-toxicity curve but differ in the dose-response curve. The proportion(s) corresponding to the preferred dose level are bordered by a solid box. If a trial has stopped for futility or safety, the patient was treated as having been assigned to dose level '0'. Figure S2 in Supplement 1 gives the same results for the three 3+3-type designs.

accompanying vignette that together form the main contribution of this paper, it is possible to quantify design decisions' impact on operating characteristics. By providing the trial simulator, we have made it simpler for clinical trialists to quantify seamless design decisions' impact on operating characteristics. Even very heterogeneous designs can be compared using our simulator, which reports the same operating characteristics regardless of the module configuration. These advantages address three key challenges to the statistical design of seamless trials outlined in Hobbs, et al. [33]: lack of formal design for DECs, imprecise operating characteristics, and difficulties in implementing interim safety and futility analyses. Future extensions could use the time-to-event-CRM [47] in Modules 1 and 3 as a means of handling late-onset toxicities. Another important direction would be to extend Modules 2 and 4 to estimate an optimal biological dose or minimum efficacious dose.

Although our primary goal was not to recommend a particular configuration of modules, our findings reiterate the importance of keeping the dose assignment mechanism open for the full duration of the trial and efficiently sharing information across dose levels and stages. With regard to this second point, an isotonic regression model for estimating dose-response curves, such as `bayes_isoreg`, is sensible. This approach could also be adapted as an option for module 1 or module 3, to estimate the dose-toxicity curve.

Funding

Disclosure

The authors have declared no conflicts of interest

References

[1] Prowell TM, Theoret MR and Pazdur R. Seamless oncology-drug development. N Engl J Med 2016; 374(21): 2001–2003.

[2] Manji A, Brana I, Amir E et al. Evolution of clinical trial design in early drug development: Systematic review of expansion cohort use in single-agent phase I cancer trials. J Clin Oncol 2013; 31(33): 4260–4267.

[3] Rogatko A, Schoeneck D, Jonas W et al. Translation of innovative designs into phase I trials. J Clin Oncol 2007; 25(31): 4982–4986.

[4] Garrett-Mayer E and O'Connell N. The evolution of phase I trials, past, present, and future: A biostatistical perspective. In Kummar S and Takimoto C (eds.) Novel Designs of Early Phase Trials for Cancer Therapeutics, chapter 3. London, United Kingdom: Academic Press, 2018. pp. 17–30.

[5] Storer BE. Design and analysis of phase I clinical trials. Biometrics 1989; 45(3): 925–937.

[6] O'Quigley J, Pepe M and Fisher L. Continual reassessment method: A practical design for phase 1 clinical trials in cancer. Biometrics 1990; 46(1): 33–48.

[7] R Core Team. R: A Language and Environment for Statistical Computing. Vienna, Austria, 2015. URL https://www.R-project.org/.

[8] Wickham H, Averick M, Bryan J et al. Welcome to the tidyverse. Journal of Open Source Software 2019; 4(43): 1686.

[9] Stan Development Team. RStan: the R interface to Stan, 2018. URL http://mc-stan.org/. R package version 2.18.2.

[10] Cheung K. dfcrm: Dose-finding by the continual reassessment method, 2013. URL http://CRAN.R-project.org/package=dfcrm. R package version 0.2-2.

[11] Chase EC and Boonstra PS. seamlesssim: Simulates seamless phase I/II oncology trials. https://github.com/elizabethchase/seamlesssim, 2020. R package version 0.0.0.9000. Accessed: 17-Sep-2020.

[12] Braun TM. The bivariate continual reassessment method: Extending the CRM to phase I trials of two competing outcomes. Controlled Clin Trials 2002; 23(3): 240–256.

[13] Thall PF and Cook JD. Dose-finding based on efficacy–toxicity trade-offs. Biometrics 2004; 60(3): 684–693.

[14] Zhang W, Sargent DJ and Mandrekar S. An adaptive dose-finding design incorporating both toxicity and efficacy. Stat Med 2006; 25(14): 2365–2383.

[15] Dragalin V and Fedorov V. Adaptive designs for dose-finding based on efficacy–toxicity response. J Stat Plan Inference 2006; 136(6): 1800–1823.

[16] Yin G, Li Y and Ji Y. Bayesian dose-finding in phase I/II clinical trials using toxicity and efficacy odds ratios. Biometrics 2006; 62(3): 777–787.

[17] Wang M and Day R. Adaptive Bayesian design for phase I dose-finding trials using a joint model of response and toxicity. J Biopharm Stat 2009; 20(1): 125–144.

[18] Houede N, Thall PF, Nguyen H et al. Utility-based optimization of combination therapy using ordinal toxicity and efficacy in phase I/II trials. Biometrics 2010; 66(2): 532–540.

[19] Thall PF, Szabo A, Nguyen HQ et al. Optimizing the concentration and bolus of a drug delivered by continuous infusion. Biometrics 2011; 67(4): 1638–1646.

[20] Zhong W, Koopmeiners JS and Carlin BP. A trivariate continual reassessment method for phase I/II trials of toxicity, efficacy, and surrogate efficacy. Stat Med 2012; 31(29): 3885–3895.

[21] Thall PF and Nguyen HQ. Adaptive randomization to improve utility-based dose-finding with bivariate ordinal outcomes. J Biopharm Stat 2012; 22(4): 785–801.

[22] Tao Y, Liu J, Li Z et al. Dose-finding based on bivariate efficacy-toxicity outcome using archimedean copula. PloS One 2013; 8(11): e78805.

[23] Wages NA and Conaway MR. Phase I/II adaptive design for drug combination oncology trials. Stat Med 2014; 33(12): 1990–2003.

[24] Pan H, Xie F, Liu P et al. A phase I/II seamless dose escalation/expansion with adaptive randomization scheme (sears). Clin Trials 2014; 11(1): 49–59.

[25] Guo B and Li Y. Bayesian dose-finding designs for combination of molecularly targeted agents assuming partial stochastic ordering. Stat Med 2015; 34(5): 859–875.

[26] Dutton P, Love S, Billingham L et al. Analysis of phase II methodologies for single-arm clinical trials with multiple endpoints in rare cancers: An example in Ewing's sarcoma. Statistical methods in medical research 2018; 27(5): 1451–1463.

[27] Msaouel P, Thall P, Yuan Y et al. A phase I/II trial of sitravatinib (sitra) combined with nivolumab (nivo) in patients with advanced clear cell renal cell cancer (ACCRCC) that progressed on prior VEGF-targeted therapy. J Clin Oncol 2020; 38(6): 612.

[28] Copland M, Slade D, Byrne J et al. FLAG-IDA and ponatinib in patients with blast

phase chronic myeloid leukaemia: results from the phase I/II UK trials acceleration programme matchpoint trial. Blood 2019; 134: 497.

[29] Hansen A, Abdul-Karim R, Rizvi N et al. A phase I/II, open-label, two part study of GSK3359609 in combination with tremelimumab in participants with selected, advanced solid tumors. In AACR Annual Meeting.

[30] Iasonos A and O'Quigley J. Design considerations for dose-expansion cohorts in phase I trials. J Clin Oncol 2013; 31(31): 4014–4021.

[31] Boonstra PS, Shen J, Taylor JMG et al. A statistical evaluation of dose expansion cohorts in phase I clinical trials. J Natl Cancer Inst 2015; 107(3): dju429.

[32] Boonstra PS, Braun TM, Taylor JMG et al. Statistical controversies in clinical research: building the bridge to phase II–efficacy estimation in dose-expansion cohorts. Ann Oncol 2017; 28(7): 1427–1435.

[33] Hobbs BP, Barata PC, Kanjanapan Y et al. Seamless designs: Current practice and considerations for early-phase drug development in oncology. J Natl Cancer Inst 2019; 111(2): 118–128.

[34] Patnaik A, Kang SP, Rasco D et al. Phase pi. Clin Cancer Res ; : 4286–4293.

[35] Kang SP, Gergich K, Lubiniecki GM et al. Pembrolizumab KEYNOTE-001: an adaptive study leading to accelerated approval for two indications and a companion diagnostic. Ann Oncol 2017; 28(6): 1388–1398.

[36] Mullard A. Reining in the supersized Phase I cancer trial. Nat Rev Drug Discov 2016; 15(6): 371–373.

[37] Wong CH, Siah KW and Lo AW. Estimation of clinical trial success rates and related parameters. Biostatistics 2019; 20(2): 273–286.

[38] Hay M, Thomas DW, Craighead JL et al. Clinical development success rates for investigational drugs. Nat Biotechnol 2014; 32(1): 40.

[39] Food and Drug Administration. Expansion cohorts: Use in first-in-human clinical trials to expedite development of oncology drugs and biologics guidance for industry draft guidance, 2018. https://www.fda.gov/media/115172/download; accessed 31-May-2019.

[40] Simon R. Optimal two-stage designs for phase II clinical trials. Controlled Clin Trials 1989; 10(1): 1–10.

[41] Garrett-Mayer E. The continual reassessment method for dose-finding studies: a tutorial. Clin Trials 2006; 3(1): 57–71.

[42] A'Hern RP. Sample size tables for exact single-stage phase ii designs. Statistics in Medicine 2001; 20(6): 859–866.

[43] Wilson EB. Probable inference, the law of succession, and statistical inference. J Am Stat Assoc 1927; 22(158): 209–212.

[44] Mayo MS and Gajewski BJ. Bayesian sample size calculations in phase ii clinical trials using informative conjugate priors. Controlled Clinical Trials 2004; 25(2): 157 – 167.

[45] Barlow RE. Statistical inference under order restrictions; the theory and application of isotonic regression. Technical report, 1972.

[46] Boonstra PS and Chase EC. Simulation study instructions. https://github.com/psboonstra/simulations_seamlesssim, 2020. Accessed: 17-Sep-2020.

[47] Cheung YK and Chappell R. Sequential designs for phase I clinical trials with late-onset toxicities. Biometrics 2000; 56: 1177–1182.