Taylor & Francis
Taylor & Francis Group

Check for updates

A simulation study of approaches for handling disease progression in dose-finding clinical trials

Lucie Biard[a,b], Bin Cheng[a], Gulam A. Manji[c], and Shing M Lee[a]

[a]Department of Biostatistics, Mailman School of Public Health, Columbia University, New York, NY, USA; [b]INSERM U1153 team ECSTRRA, Université de Paris, Paris, France; [c]Division of Medical Oncology, Columbia University Irving Medical Center, and New York Presbyterian Hospital, Herbert Irving Pavilion, New York, NY, USA

ABSTRACT

In traditional dose-finding studies, dose-limiting toxicity (DLT) is determined within a fixed time observation window where DLT is often defined as a binary outcome. In the setting of oncology dose-finding trials, often patients in advanced stage of diseases are enrolled. Therefore, disease progression may occur within the DLT observation window leading to treatment discontinuation and rendering the patient unevaluable for DLT assessment. As a result, additional patients have to be enrolled, increasing the sample size. We propose and compare several practical approaches for handling disease progression which occurs within the DLT observation window, while in the framework of the time-to-event continual reassessment method (TITE-CRM) which allows using partial observations. The approaches differ on the way they define an evaluable patient and in the way incomplete observations are included. The practical approaches, which we call strategies A, B and C, are illustrated and contrasted in the context of a single simulated trial, and compared via simulations under various scenarios of dose-progression relationship, in the setting of advanced soft-tissue sarcoma.

1. Introduction

Phase I dose-finding trials aim to determine a safe dose of a new drug to be used and further assessed for activity and efficacy in subsequent trials (Chevret 2006). Depending on the targeted condition or disease, and the type of expected side effects of the drug, phase I trials are performed in either healthy volunteers or patients. Based on ethical considerations, cancer dose-finding trials have usually been conducted in patients with advanced-stage disease that have grown refractory to multiple lines of treatment, due to the serious side effects of cytotoxic chemotherapies.

In a traditional dose-finding design in oncology with cytotoxic agents, several assumptions are made. First, it assumes monotone increasing dose-efficacy and dose-toxicity relationships. Second, it assumes acute severe toxicities, usually within one or two treatment cycles (4–8 weeks after treatment initiation). Third, it assumes delayed assessment of response compared to the toxicity observation window and interval drug exposure (the cytotoxic treatment is discontinued once response is definitely assessed, as opposed to chronic exposure with cytostatic agents) (Chevret 2006; Mathijssen et al. 2014).

This paradigm has been challenged in recent years with the development of new classes of anti-cancer agents such as molecularly targeted agents and immunotherapy, or refinement of pre-existing treatments (*e.g.*, radiation therapy) (Mathijssen et al. 2014). These new treatments, some of them cytostatic, result in new frameworks for safety evaluation: plateau dose-efficacy relationship, late-onset

or chronic toxicities, prolonged drug exposure, and the convergence between the safety and the efficacy time observation windows. These suggest the need for longer toxicity observation windows beyond one or two treatment cycles to capture the toxicities resulting from such prolonged treatments (Lee et al. 2016; Postel-Vinay et al. 2014, 2011).

However, given the advanced-stage disease of patients enrolled in dose-finding cancer trials, disease progression may occur within the toxicity observation window. For ethical reasons, dose-finding study protocols usually allow handling progression by discontinuing the experimental treatment and proposing alternative therapies if available. For dose-finding studies with dichotomous or polytomous toxicity outcome observed within a pre-specified time window (i.e. 3 + 3 design, continual reassessment method (CRM), escalation with overdose control (EWOC), modified toxicity probability interval design (mTPI)), the occurrence of progression precludes complete safety assessment as required by the design, rendering the patient unevaluable and resulting in mandatory patient replacement, which in turn, leads to increased study duration and cost. Replacing patients due to progression in studies with late-onset toxicity or short median progression-free survival can be problematic and inefficient, especially in the context of rare diseases.

In this paper, we propose and evaluate by simulation practical approaches for handling disease progression in the context of the time-to-event CRM, which allows use of partial observations. The proposed approaches allow various proportions of patients who did not complete follow-up due to disease progression to be included in the TITE-CRM for the estimation of the MTD.

2. Motivating examples

The present work was motivated by two clinical examples. The first is a dose-finding trial in patients with unresectable sarcoma or malignant peripheral nerve sheath tumors that was recently completed Columbia University Medical Center (NCT02584647). The goal of the trial was to identify the maximum tolerated dose (MTD) out of five doses of interest. The trial was designed using the time-to-event CRM (TITE-CRM) (Cheung and Chappell 2000) with a sample size of 24 patients initially, revised to 18 patients during the trial, and a pre-specified target dose-limiting toxicity (DLT) probability of 0.25. DLT was defined as any grade 3 or higher non-hematologic toxicity or grade 4 or higher hematologic toxicity in the first two cycles of treatment (i.e. 8 weeks). The TITE-CRM was selected to allow for enrollment before fully observed outcomes at the end of 8 weeks. However, given a median progression-free survival (PFS) of 6 weeks under standard of care (Maki et al. 2009), we expected almost 50% of discontinuation due to progression by the end of the 8-week toxicity observation window.

The second motivating example arises from a published re-analysis of dose-finding trials of bortezomib in hematologic malignancies (Lee et al. 2016). The results pointed toward an extension of the DLT observation window (to approximately five treatment cycles) to be able to capture and evaluate safety including non negligible late-onset toxicities. At the approved dose of 1.3 mg/m2, the cumulative incidence of DLT over the whole course of treatment largely exceeded the conventional target of 20–33%. However, the median PFS was less than 5 cycles.

In these examples, the simplest approach for handling patients who progress within a pre-specified toxicity observation window without a DLT is to render them not evaluable and enroll additional patients to replace them. In the first motivating example, with a DLT observation window of 8 weeks and assuming a median PFS ranging from 6 weeks under standard of care to 12 weeks at the highest dose level, the increase in the number of patients due to replacements would be of at least 37%, and potentially as high as 60% (i.e. more than every other patient needing replacement). The operating characteristics of the design remain the same given the maintained sample size (with replacements) used for the estimation of the MTD. However, it can lead to a large number of replacements increasing trial duration and cost, and the potential bias of selection of healthier patients (Winther et al. 2016).

In the context of the TITE-CRM, which allows for inclusion of partial information on patient follow-up during the trial, there is flexibility in the handling of early progressions occurring within the

toxicity observation window, in the absence of DLT. We propose and compare three approaches for handling treatment discontinuation, notably due to disease progression (termed "progression" in the remaining of the paper for simplicity), in the context of the TITE-CRM and evaluate the impact on the operating characteristics of the TITE-CRM.

The Methods section introduces notations, describes three strategies to handle progression within the TITE-CRM design and presents our simulation settings. The Results section reports the results of the simulation study and concluding remarks are provided in the Discussion section.

3. Methods

3.1. Notations

Consider a dose-finding trial with K dose-levels of interest, $d_1, \ldots, d_K$, and a sample size of N. Dose-finding designs usually require DLT to be defined relative to a toxicity observation window, of a known duration T, to identify the MTD corresponding to a pre-specified target DLT probability. Thus, DLT is often modeled as a binary variable and may occur any time within the toxicity observation window (Cheung and Chappell 2000; Chevret 2006).

Let U be the time from inclusion to DLT for a patient. DLT occurs within the toxicity window if $U \leq T$. In case of early progression at time P for a patient, that is, $P \leq \min\{U, T\}$, a follow-up of duration P without toxicity is available for this patient, and the toxicity outcome is censored earlier than the scheduled time T by $T - P$ time-units.

For simplicity, and given the framework of oncology dose-finding trials, we assume that there is no risk of loss to follow-up in this setting.

3.2. Strategies to handle disease progression

Partial information due to progression events within the toxicity observation window can be handled using the TITE-CRM. This model-based approach incorporates partial observations through a weighted likelihood. Thus, incomplete observations due to progression can be treated as censored and included in the model likelihood, as observations without toxicity, weighted according to the available follow-up.

Moreover, we can define a minimum follow-up threshold to render patients who progressed evaluable for toxicity assessment. Formally, let ϕ be a pre-specified fraction of the toxicity window T, where $0 \leq \phi \leq 1$; a patient who progresses is deemed evaluable at the end of follow-up if $P \geq \phi T$. If $\phi = 0$, all incomplete observations censored by progression are deemed evaluable. If $\phi = 1$, all incomplete observations censored by progression are deemed unevaluable and need to be replaced. The sample size for the TITE-CRM is the number of evaluable patients. Thus, additional patients may need to be included to compensate for unevaluable patients.

Different strategies can be implemented depending on: the threshold ϕ and the use of partial observations collected data for unevaluable patients with $P < \phi T$. Even though unevaluable patients will not count in the effective sample size, their partial information can be used or not in the estimation of the MTD for dose assignments. For instance, a progression, with $P < \phi T$, may occur before the enrollment of any new patient; in that case, one could argue that this unevaluable observation could be excluded.

Three strategies were therefore considered within the TITE-CRM design. In all of them, progression without DLT was considered as censoring resulting in incomplete observation of the toxicity window. While the strategies all maintained the same number of evaluable patients, they differed in (1) the choice of ϕ, defining an evaluable patient, (2) the inclusion of unevaluable patients for the estimation of the MTD. The strategies can be summarized as follows:

- Strategy A

Let $\phi = 0$. All patients are evaluable regardless of time of progression.

- Strategy B

(1) Choose a $\phi > 0$.
(2) Include all available partial follow-up time P from unevaluable patients with $P < \phi T$.

- Strategy C
(1) Choose a $\phi > 0$.
(2) Include partial information from unevaluable patients with $P < \phi T$ only if the information has been used for previous dose assignments.

Of note, the number of evaluable patients is the same in all three strategies. In strategy A, all patients are evaluable since $\phi = 0$, so the number of evaluable patients corresponds to the included patients. In strategies, B and C, $\phi > 0$ implies that the number of included patients may exceed the number of evaluable patients, to compensate for unevaluable patients.

The strategies are further exemplified in a sample trial detailed in section 3.4 and illustrated in Figure 1.

3.3. Simulation settings

We examined the different strategies to handle progression within the TITE-CRM design, in a simulation study in the context of the sarcoma motivating example. The design parameters were $K = 5$, $N = 24$, $T = 8$ (corresponding to two cycles of treatment, with toxicity and progression assessment every week) and a target DLT probability of 0.25. The median PFS under standard of care was assumed to be 6 weeks, resulting in a progression rate of approximately 60% by T. For the estimation of the MTD, the Bayesian framework of the TITE-CRM with the empirical model for the probability of DLT by T was used. The cohort size was 1 and no dose-skipping was allowed during dose escalation. The skeleton was selected using the indifference interval calibration approach. The optimal length of the indifference interval was $\delta = 0.10$ based on 2000 simulations (Lee and Cheung 2009). To evaluate the impact of sample size, we also performed simulations with a sample size $N = 18$ and its corresponding indifference interval of $\delta = 0.09$.

In the design of the trial, five different main scenarios of dose-toxicity relationship were evaluated, depending on the MTD true location. The dose-toxicity relationship was assumed to be monotone increasing, in line with a previous publication (see Table 1) (Wages and Tait 2015). We evaluated 16 additional dose-toxicity scenarios for sensitivity analysis, in which the dose-toxicity curve was less favourable for dose finding: in scenarios s1 to s4, a true MTD with a DLT probability slightly above (0.28) or, in scenarios s5 to s8, below (0.22) the targeted value 0.25, then less steep dose-toxicity curve around the MTD, with odds ratio for DLT of 1.5, below (s9 to s12) and above (s13 to s16) the MTD.

For each of the toxicity scenarios, we evaluated 11 scenarios of dose-progression relationships. More specifically, five of them assumed constant progression rates of 0%, 20%, 40%, 60%, or 80% by T across all dose levels, corresponding to a reference situation or absence of dose effect on progression. The other six progression scenarios assumed dose-dependent progression rates: monotone decreasing (60%, 50%, 40%, 30%, 20%) corresponding to the setting of conventional cytotoxic chemotherapy, plateau with step-decrease from 60% to 40% at dose-level 2, 3, 4 or 5, or U-shaped (60%, 50%, 40%, 50%, 60%), to evaluate non strictly monotone relationships as found with new anticancer agents such as immunotherapies (Riviere et al. 2018; Tait and Wages 2015). Of note, the progression rates used in
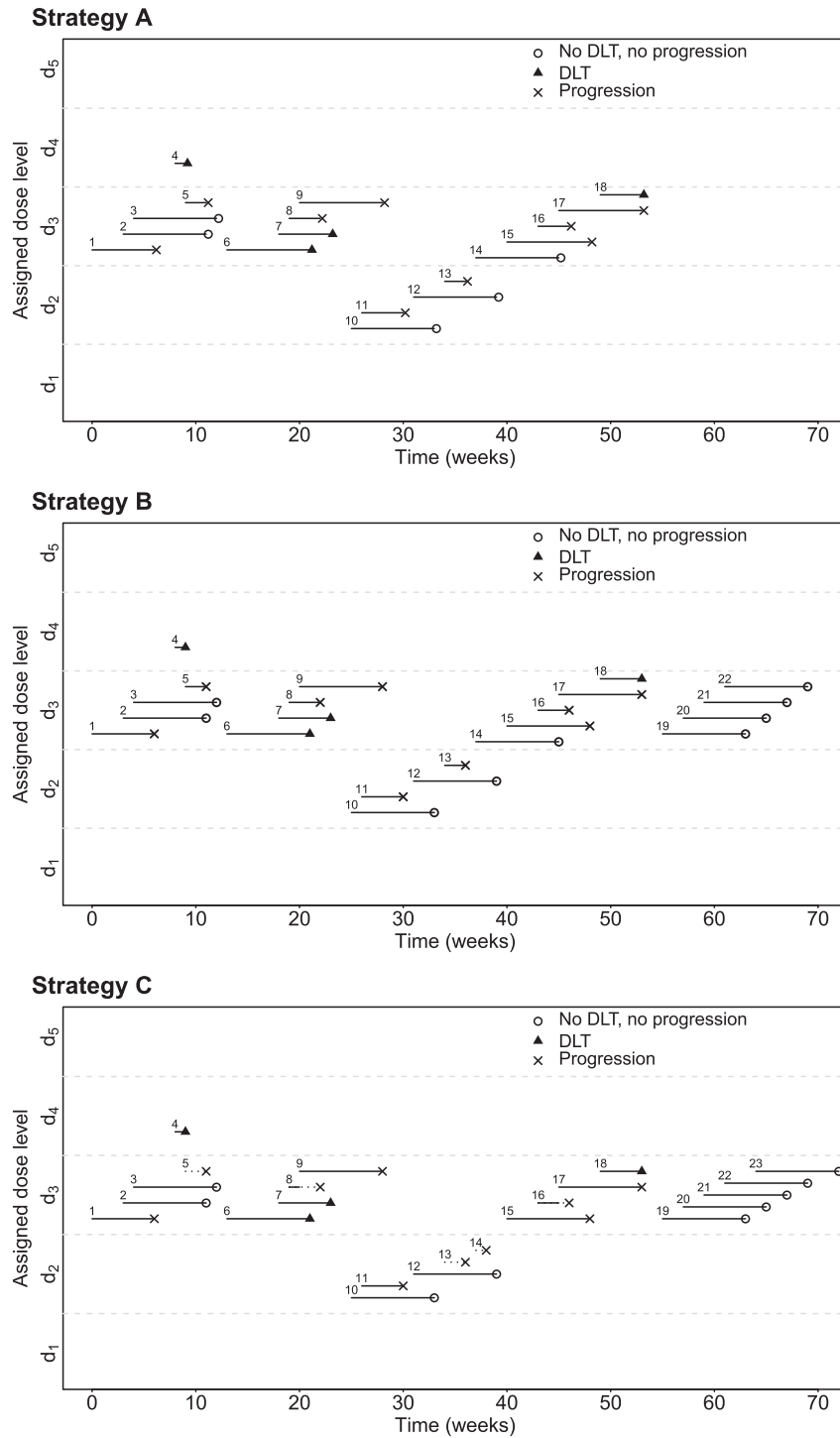
Figure 1. Simulated sample trial. Patients flow according to the three strategies, in a trial scenario with targeted sample size $N = 18$ evaluable patients, five candidate dose levels, $T = 8$ weeks, $\phi = 0.5$. Enumerated segments represent patients' observations, with solid lines for follow-up information included in the MTD TITE-CRM estimation, and dotted lines for excluded follow-up. Patients within dashed horizontal lines received the same dose level, denoted on the y-axis. (DLT: Dose-Limiting Toxicity observed).

Table 1. Dose-toxicity relationship in simulations scenarios: true dose limiting toxicity (DLT) rate by toxicity window T for each dose level, 1 to 5.

Pr(DLT) by T for dose levels 1 to 5 (%)
(00, 01, 05, 10, **25**)
(01, 05, 10, **25**, 40)
(05, 10, **25**, 40, 55)
(10, **25**, 40, 55, 65)
(**25**, 40, 55, 65,70)

the scenarios (20%, 40%, 50%, 60%, 80%) corresponded to median progression-free survival of roughly 24, 11, 8, 6, 3.5 weeks, respectively, assuming constant hazard.

Time to event outcomes were generated using conditional uniform models, similar to the ones used in the original TITE-CRM publication (Cheung and Chappell 2000). For each observation, the presence of each event within the observation window T was first determined using a Bernoulli distribution with parameter corresponding to the latent cumulative incidences at T described above, and if so, a time to the event was drawn from a uniform distribution on $[0, T]$. Time to toxicity and time to progression were generated independently; we then defined the observation as the time to the first event or the end of the window T (in case of no event). Observations with no event or with a DLT are complete, as per the TITE-CRM design, whereas observations with a progression ($\min(U, P, T) = P$) are incomplete.

Simulations were performed with a fixed accrual of 2 patients per toxicity observation window (1 per 4-week treatment cycle). Moreover, to evaluate the impact of ϕ, simulations were performed for ϕ values of 0.25, 0.50, and 0.75.

The progression-handling strategies were compared based on probability of correct dose selection (PCS), probability of overdose selection (POS), the number of additional patients enrolled in replacement of unevaluable patients and the duration of the trial.

For each scenario, 10000 replicates were performed, so that the observed PCS would be expected to lie within a 2%-wide interval in 95% of cases, if the true PCS of the design was 0.60. R statistical platform, version 3.4.2, and the dfcrm library, version 0.2–2, were used for the computations.

3.4. Sample trial

To illustrate and contrast the three strategies, we applied them to a simulated sample trial, presented in Figure 1. A trial with sample size $N = 18$ was generated under the simulation scenario with a plateau dose-progression relationship with marginal progression incidences at T: (60%, 60%, 40%, 40%, 40%) as described above. The observation window T was 8 weeks, and ϕ=0.5 defining as evaluable patients with progression in strategies B and C when $P \geq 4$ weeks.

Figure 1 illustrates the differences among the strategies. Solid segments represent follow-up which is included in the estimation of the MTD (through weights in the TITE-CRM likelihood); dotted segments correspond to follow-up which is not included. Both strategies A and B include all available information, even very short partial observations due to early progression (patients #5, 8, 13, 16). However, these patients are considered unevaluable under strategy B and replaced, resulting in four additional recruitments, for a total sample size of 22. Strategy C includes partial follow up from unevaluable patients to the extent that it has been used for dose assignment computations during the trial. In other words, if a patient is deemed unevaluable, $P < \phi T$, before the next patient is included, this incomplete observation is excluded in strategy C, while included in strategies A and B (patients #5, 13, and 14). Moreover, if the patient is deemed unevaluable ($P < \phi T$) after subsequent patient(s) has(have) been enrolled, only the information that has previously been used for dose assignment computations is included in the remaining of the trial with strategy C (e.g. patients #8 and 16). Overall, this may result in slightly different dose assignments and total number of included patients, as

illustrated in Figure 1. Notably, patient #14 is assigned to dose level 3 with strategies A and B, but, more conservatively, to dose level 2 in strategy C. Since patient #14 progressed early when assigned to dose level 2, and before patient #15 is included, strategy C requires one more patient compared to strategy B, in this example.

4. Results

For simplicity and brevity, results from five progression scenarios along with a reference scenario with no progression, with true MTD dose level ranging from 1 to 5, and $\phi = 0.5$, are reported hereafter. Full results, including all combinations of toxicity and progression scenarios and ϕ parameter, are available in Supplemental material. Tables 2 and 3 present results for planned sample sizes of 24 and 18 patients, respectively.

In the absence of progression, the probability of correct dose selection (PCS) using the TITE-CRM design ranged from 62.7% to 69.1% for $N = 24$ and 57.6% to 67.8% for $N = 18$ in the reported scenarios, depending on the true correct MTD level. The probability of overdose selection (POS) ranged from 13.7% to 31.5% for $N = 24$ and from 17.8% to 34.0% for $N = 18$.

Overall, increased progression rates resulted in decreased PCS and increased POS, regardless of the strategy used and similarly in trials with $N = 18$ or 24. However, strategy A showed greater decrease in PCS and greater increase in POS compared to strategies B and C, in all scenarios, regardless of the planned sample size. For instance, with $N = 24$ and $\phi = 0.5$, PCS decreased by 9.3% in the scenario with true MTD being dose level 2 and 60% constant progression rate (from 62.7% in the absence of progression to 53.4%), compared to 5.4% decrease for strategy B (57.3%) and 3.9% decrease for

Table 2. Percent of correct selection (PCS), overdose selection (POS), and mean increase in sample size ($ + N$ (%)) for strategies A, B, and C, according to DLT and progression latent rates at T (%), in scenarios with five candidate dose levels, from 10000 simulations with $N = 24$.

Pr(DLT) (%)	Pr(Progression)	Strategy A		Strategy B			Strategy C		
		PCS	POS	PCS	POS	N(%)	PCS	POS	N(%)
(**25**,40,55,65,70)	(00,00,00,00,00)	**68.5**	**31.5**	**68.5**	**31.5**	**0 (0)**	**68.5**	**31.5**	**0 (0)**
	(20,20,20,20,20)	64.7	35.3	66.0	34.0	1.9 (8)	67.5	32.5	1.9 (8)
	(60,60,60,60,60)	57.2	42.8	61.1	38.9	6.2 (26)	67.3	32.6	6.5 (27)
	(60,50,40,30,20)	60.5	39.5	62.5	37.5	5.5 (23)	67.7	32.3	5.5 (23)
	(60,40,40,40,40)	61.6	38.4	63.5	36.5	6.2 (21)	68.3	31.7	5.3 (22)
	(60,60,60,60,40)	57.1	42.9	61	39	6.2 (26)	67.4	32.6	6.2 (26)
(10,**25**,40,55,65)	(00,00,00,00,00)	**62.7**	**25.0**	**62.7**	**25.0**	**0 (0)**	**62.7**	**25.0**	**0 (0)**
	(20,20,20,20,20)	59.9	27.5	61.6	26.6	1.9 (8)	62.6	24.8	1.9 (8)
	(60,60,60,60,60)	53.4	33.9	57.3	31.6	6.5 (27)	58.8	26.6	6.5 (27)
	(60,50,40,30,20)	56.0	30.9	60.1	28.8	4.8 (20)	60.8	25.5	4.8 (20)
	(60,40,40,40,40)	56.4	30.1	60.9	28.2	4.3 (18)	60.4	25.7	4.3 (18)
	(60,60,60,60,40)	53.5	33.9	57.4	31.6	6.5 (27)	59.1	26.5	6.5 (27)
(05,10,**25**,40,55)	(00,00,00,00,00)	**64.6**	**18.9**	**64.6**	**18.9**	**0 (0)**	**64.6**	**18.9**	**0 (0)**
	(20,20,20,20,20)	61.8	21.1	63	20.7	1.9 (8)	62.6	19.8	1.9 (8)
	(60,60,60,60,60)	55.8	26.1	60.7	23.9	6.5 (27)	60.6	19.7	6.5 (27)
	(60,50,40,30,20)	59.8	22.4	62.8	21.7	4.1 (17)	61.7	19.6	4.1 (17)
	(60,40,40,40,40)	58.8	23.6	62.1	22.2	4.1 (17)	61.6	20.1	4.1 (17)
	(60,60,60,60,40)	56.5	25.9	60.7	24.3	6.5 (27)	60.7	20.1	6.5 (27)
(01,05,10,**25**,40)	(00,00,00,00,00)	**65.5**	**13.7**	**65.5**	**13.7**	**0 (0)**	**65.5**	**13.7**	**0 (0)**
	(20,20,20,20,20)	62.1	16.4	64.4	15.4	1.9 (8)	64.2	14.3	1.9 (8)
	(60,60,60,60,60)	56.1	20.6	61.3	18.8	6.5 (27)	60.5	15.4	6.5 (27)
	(60,50,40,30,20)	61.2	16.4	63.9	15.8	3.1 (13)	62.8	14.6	3.1 (13)
	(60,40,40,40,40)	60.7	17.4	63.0	16.8	4.1 (17)	62.3	14.9	4.1 (17)
	(60,60,60,60,40)	59.3	17.8	62.9	17.5	6.0 (25)	61.5	14.3	6.2 (26)
(00,01,05,10,**25**)	(00,00,00,00,00)	**69.1**	n/a	**69.1**	n/a	**0 (0)**	**69.1**	n/a	**0 (0)**
	(20,20,20,20,20)	69.2	n/a	70.6	n/a	1.9 (8)	69.5	n/a	1.9 (8)
	(60,60,60,60,60)	67.7	n/a	73	n/a	6.7 (28)	67.3	n/a	6.7 (28)
	(60,50,40,30,20)	68.3	n/a	71.1	n/a	2.6 (11)	69.2	n/a	2.6 (11)
	(60,40,40,40,40)	67.8	n/a	71.6	n/a	4.1 (17)	68.2	n/a	4.1 (17)
	(60,60,60,60,40)	66.4	n/a	72.2	n/a	5.3 (22)	68.0	n/a	5.5 (23)

Table 3. Percent of correct selection (PCS), overdose selection (POS), and increase in sample size (+N (%)) for strategies A, B, and C, according to DLT and progression latent rates at T (%), in scenarios with five candidate dose levels, from 10000 simulations with $N = 18$.

		Strategy A		Strategy B			Strategy C		
Pr(DLT)	Pr(Progression)	PCS	POS	PCS	POS	N(%)	PCS	POS	N(%)
(**25**,40,55,65,70)	(00,00,00,00,00)	**66.0**	**34.0**	**66.0**	**34.0**	**0 (0)**	**66.0**	**34.0**	**0 (0)**
	(20,20,20,20,20)	63.4	36.6	64.7	35.3	1.4 (8)	65.9	34.1	1.4 (8)
	(60,60,60,60,60)	56.9	43.1	60.0	40.0	4.7 (26)	64.3	35.7	4.7 (26)
	(60,50,40,30,20)	59.7	40.3	61.4	38.6	4.0 (22)	65.5	34.5	4.0 (22)
	(60,40,40,40,40)	60.2	39.8	62.5	37.5	3.8 (21)	65.9	34.1	3.8 (21)
	(60,60,60,60,40)	56.5	43.5	59.5	40.5	4.9 (26)	64.2	35.8	4.9 (26)
(10,**25**,40,55,65)	(00,00,00,00,00)	**57.6**	**27.0**	**57.6**	**27.0**	**0 (0)**	**57.6**	**27.0**	**0 (0)**
	(20,20,20,20,20)	54.5	29.9	55.5	29.1	1.4 (8)	56.1	27.6	1.4 (8)
	(60,60,60,60,60)	48.1	35.7	52.6	32.9	4.9 (27)	53.7	28.5	4.9 (27)
	(60,50,40,30,20)	50.7	32.9	52.8	31.8	3.6 (20)	54.2	28.4	3.6 (20)
	(60,40,40,40,40)	50.5	32.8	53.8	30.8	3.2 (18)	54.0	28	3.4 (19)
	(60,60,60,60,40)	47.9	36.0	52.3	33.1	4.9 (27)	53.4	28.8	4.9 (27)
(05,10,**25**,40,55)	(00,00,00,00,00)	**57.7**	**21.5**	**57.7**	**21.5**	**0 (0)**	**57.7**	**21.5**	**0 (0)**
	(20,20,20,20,20)	54.5	24.4	55.9	23.9	1.4 (8)	55.4	22.8	1.4 (8)
	(60,60,60,60,60)	49.5	28.6	53.8	27.0	4.9 (27)	53.7	23.0	4.9 (27)
	(60,50,40,30,20)	53.7	24.5	56.6	24.2	3.1 (17)	55.0	22.5	3.1 (17)
	(60,40,40,40,40)	53.0	26.0	55.1	25.0	3.1 (17)	54.7	22.4	3.1 (17)
	(60,60,60,60,40)	50.5	27.6	54.9	26.1	4.9 (27)	54.2	22.8	4.9 (27)
(01,05,10,**25**,40)	(00,00,00,00,00)	**58.0**	**17.8**	**58.0**	**17.8**	**0 (0)**	**58.0**	**17.8**	**0 (0)**
	(20,20,20,20,20)	55.8	19.3	57.2	19.5	1.4 (8)	56.7	18.4	1.4 (8)
	(60,60,60,60,60)	50.2	23.4	55.5	21.3	5.0 (28)	54.2	18.2	5.0 (28)
	(60,50,40,30,20)	54.3	19.9	57.4	19.4	2.3 (13)	56.5	18.0	2.5 (14)
	(60,40,40,40,40)	54.0	20.3	57.1	19.7	3.1 (17)	56.0	18.0	3.1 (17)
	(60,60,60,60,40)	51.9	21.0	56.3	20.2	4.5 (25)	54.9	17.5	4.7 (26)
(00,01,05,10,**25**)	(00,00,00,00,00)	**67.8**	n/a	**67.8**	n/a	**0 (0)**	**67.8**	n/a	**0 (0)**
	(20,20,20,20,20)	65.5	n/a	67.6	n/a	1.4 (8)	66.1	n/a	1.4 (8)
	(60,60,60,60,60)	64.8	n/a	69.1	n/a	5.0 (28)	64.1	n/a	5.0 (28)
	(60,50,40,30,20)	64.8	n/a	67.8	n/a	2.0 (11)	65.6	n/a	2.0 (11)
	(60,40,40,40,40)	65.0	n/a	68.2	n/a	3.1 (17)	64.4	n/a	3.1 (17)
	(60,60,60,60,40)	64.0	n/a	67.8	n/a	4.1 (23)	64.7	n/a	4.1 (23)

strategy C (58.8%) (Table 2). When comparing between strategy B and C, strategy C had better PCS, in scenarios where the true MTD was dose level 1 and 2, and slightly lower PCS when the true MTD was dose level 3 to 5 (less than 1% difference when the true MTD was dose level 3 to 8.1% difference in extreme settings when the true MTD was dose level 5, 80% constant progression rate). However, at the same time, strategy C's POS was consistently lower than that of strategy B, across all scenarios and settings, even reducing POS compared to reference situations without progression when $\phi = 0.75$ (Table S4, Supplemental material).

Regarding the impact of the dose-progression relationship, PCS was most impacted in cases of constant progression across dose levels, in comparison with decreasing, plateau or U-shaped progression rates at T (Tables 2 and 3, and Supplemental material).

While strategy A maintained the initial planned sample size by definition, the replacements compensating unevaluable patients with progression were very similar between strategies B and C. With $\phi = 0.5$, it ranged from 8% to 28% of the initially planned sample size N, in the reported scenarios (Tables 2 and 3), which corresponds to 2 to 7 supplemental patients when $N = 24$.

Tables S3 and S4 (Supplemental material) present simulations results with N = 24, for strategies B and C when $\phi = 0.25$ or 0.75, respectively. Although similar trends to those mentioned above were observed in these settings, the gain in PCS and POS of strategies B and C relative to strategy A were diminished when $\phi = 0.25$, with values of PCS and POS converging to strategy A's values. On the contrary, when $\phi = 0.75$, both PCS and POS improved, outperforming the reference scenario with no progression, at the cost of a greater number of unevaluable patients and increase in the number of included patients, compared to designs with $\phi = 0.5$.

Tables S5 to S8 (Supplemental material) present simulations results with N = 24, and sensitivity toxicity scenarios (reduced steepness of the dose-toxicity curve). Despite poorer performances of the TITE-CRM design overall, compared to the main toxicity scenarios, the proposed approaches consistently resulted in preventing the decrease in PCS and increase in POS and trial duration, in the presence of disease progression, similarly to the main results.

5. Discussion

When designing dose-finding clinical trials, it is important to consider the median progression-free survival of the population that will be included in the study if treatment will be discontinued after disease progression leading to unevaluable patients. In the setting where a substantial number of patients are expected to progress within the toxicity observation window, consideration should be given to the use of the TITE-CRM given that it allows the flexibility of using partial information. Indeed, other methods with strictly dichotomous endpoints leave no other choice than replacing every patient who progressed. Instead, we show that the use of partial information within the TITE-CRM can reduce the final number of included patients, the duration of the study and the costs associated with both.

This paper proposes three practical strategies to handle disease progression events within the TITE CRM, as alternatives to systematic replacement, and compared their operating characteristics via extensive simulations. Based on our simulation study, progression that occurs within the observation window can have significant impact in the operating characteristics of the TITE-CRM, as progression led to reduction in the probability of correct selection (PCS) of the true MTD, and increase in the probability of overdose selection (POS) at the end of trial in most cases, regardless of how progression cases are handled.

As expected, we show that operating characteristics of the design can be maintained only at the cost of supplemental patients and prolonged duration. However, we show that the definition of evaluable patients with the threshold ϕT and partial replenishment of unevaluable patients only can yield acceptable performances. More precisely, the choice of strategy should be trade-off between the loss in design performances (PCS, POS, duration) and the feasibility of replacing unevaluable patients. Both strategy B and C, which allowed for the replacement of pre-defined unevaluable patients, outperformed strategy A with better PCS and lower POS. Furthermore, strategy B was more aggressive than strategy C in including all partial observations, even very short ones, and recommending doses above the MTD more often. Excluding unevaluable observations due to progression, when possible, reduces aggressiveness in dose-escalation and POS. The choice of $\phi = 0.5$, the fraction parameter defining evaluable patients for toxicity assessment, relatively to the observation window, seemed to be optimal in the sense that it balanced favorably the improvement in PCS and POS with a moderate increase in sample size (and therefore trial duration) compared to strategy A (no increase in sample size). Increasing ϕ, and up to $\phi = 1$, consists in systematically compensating a progression with an additional patient, whenever it happens within the toxicity window. As expected with more complete observations, it improves PCS and POS, but at the cost of a prolonged trial duration and increased number of inclusions (Tables S3 and S4, $\phi = 0.75$). Of note, for ethical reasons in the TITE-CRM, we consider not possible to fully exclude the incomplete observations of patients with progression from further MTD estimations if they have already been used for dose assignments. This is the rationale behind Strategy C, which compromises between exclusion and use of incomplete observations, in accordance with the TITE-CRM design requirements.

In sensitivity analyses, we confirmed the consistency of the approaches' performance in more difficult scenarios, where designs performances are expected to be poorer due to the reduced steepness of the dose-toxicity curve, even in the absence of progression. If the dose-toxicity curve is expected to be somewhat flat, with small odds ratios between the probabilities of DLT above and below the MTD, it is expected that the PCS will be lower, whichever the progression risk. It is then recommended to

increase the sample size, in order to maintain the design's performances, independently of disease progression (Cheung 2013).

In terms of practical guidelines, our simulation study suggests that low progression (no more than 20%) had limited impact on PCS and POS (less than 5% decrease in PCS and increase in POS). We would therefore recommend strategy A when the expected rate of progression is low within the DLT observation window. Thus, rather than secondarily excluding and replacing patients with progression, investigators may consider including all partial observations, as evaluable, especially in the context of rare diseases, even though some observations will be incomplete due to progression. When the expected rate of progression is moderate (from 20% to 60%) within the desired toxicity observation window, we recommend implementing strategy C with $\phi = 0.5$, to maximize PCS while controlling POS, as a trade-off for avoiding replacement of every patient who progresses. If the expected rate of progression is high (above 60%), the definition of the toxicity observation window and/or the eligibility criteria for the trial should be reconsidered or alternative designs, such as phase I/II designs which account for efficacy information, should be evaluated (Tait and Wages 2015; Yuan and Yin 2009; Zang et al. 2014; Zhang et al. 2006).

In addition to the expected rate of progression, it is important to consider other logistical issues when designing a dose-finding trial; for example, the toxicity and efficacy assessment schedule, the criteria for treatment discontinuation and the rate of accrual and acceptability of prolonged trial duration. In our motivating example, we assumed that toxicity and efficacy could be assessed within the same time frame and that the accrual rate was relatively fast (i.e. multiple patients could be accrued within the toxicity observation window). Thus, for reproducibility and both practical and ethical reasons, we could not systematically exclude patients with progression: part of their follow-up information might have been already used in the TITE-CRM dose-assignment computations for subsequent patient entries. However, if accrual is sufficiently slow to guarantee no more than one entry per toxicity window, patients with progression could be systematically excluded and replaced; in that case a conventional CRM design might be used instead of the TITE-CRM. On the other hand, in the presence of moderate risk of progression, as long as the accrual rate is not too slow, we argue that a TITE-CRM design in conjunction with strategy C and $\phi = 0.5$ should be preferable to a CRM design (or any other design relying on complete observation of dichotomous toxicity outcome) because the former results in improved efficiency, in avoiding to replace all patients who progressed and limiting the increase in trial duration and costs.

Finally, with anti-cancer molecularly targeted agents (MTAs) and immunotherapy agents changing the landscape of dose-finding in oncology (Kummar et al. 2006; Lee et al. 2016; Le Tourneau et al. 2010; Postel-Vinay et al. 2011, 2014), with prolonged toxicity windows, researchers have been advocating designs including efficacy in the dose-finding process for MTAs, often relying on an optimal biological dose (OBD) instead of the MTD (Wages and Tait 2015; Yuan and Yin 2009; Zang et al. 2014; Koopmeiners and Modiano 2014). Many of these designs require both toxicity and efficacy outcomes to be observed. However, in oncology settings, it is common that, due to ethical concerns, patients are taken off study at progression and the experimental drug is discontinued. This prevents further assessment of DLT. At the same time, if patients experience a DLT, the drug is generally discontinued as well, preventing further assessment for efficacy. Dual assessment for toxicity and efficacy is only possible if the respective endpoint events do not prevent the complete observation one another, contrary to DLT and disease progression in our setting.

6. Conclusion

Disease progression can occur during patient follow-up in dose-finding cancer clinical trials. However, current methods generally consider patients who progress not evaluable for toxicity, increasing both trial duration and cost. This paper proposes approaches for handling disease progression in the context of the TITE-CRM, which can accommodate partial observations. Practical strategies, relying on the inclusion of incomplete observations in the sequential MTD estimation and on patient's

replacement, can limit the impact of treatment discontinuation due to disease progression. These strategies are particularly useful to handle moderate degree of progression. Thus, it is important to assess time to progression in the design of dose-finding cancer trials. Our conclusions can be extended to any cause of discontinuation, assumed independent of the toxicity endpoint. If both progression and toxicity are of interest in the dose-finding process, designs relying on both toxicity and efficacy should be considered.

Abbreviations

CRM: Continual reassessment method; DLT: Dose-Limiting Toxicity; MTAs: Molecularly targeted agents; MTD: Maximum tolerated dose; OBD: Optimal biological dose; PCS: Probability of correct dose selection; POS: Probability of overdose selection; TITE-CRM: Time-to-event continual reassessment method.

Funding

Declaration of interests

The authors declare that they have no competing interests.

References

Cheung, Y. K. 2013. Sample size formulae for the Bayesian continual reassessment method. *Clinical Trials* 10 (6):852–861. doi:10.1177/1740774513497294.

Cheung, Y. K., and R. Chappell. 2000. Sequential designs for phase I clinical trials with late-onset toxicities. *Biometrics* 56 (4):1177–1182. doi:10.1111/j.0006-341X.2000.01177.x.

Chevret, S., ed. 2006. *Statistical methods for dose-finding experiments*. Chichester, England: John Wiley & Sons, Ltd.

Koopmeiners JS and Modiano J. A Bayesian adaptive phase III clinical trial for evaluating efficacy and toxicity with delayed outcomes. Clin Trials 2014; 11: 38–48

Kummar, S., M. Gutierrez, J. H. Doroshow, and A. J. Murgo. 2006. Drug development in oncology: Classical cytotoxics and molecularly targeted agents. *British Journal of Clinical Pharmacology* 62 (1):15–26. doi:10.1111/j.1365-2125.2006.02713.x.

Le Tourneau, C., V. Diéras, P. Tresca, W. Cacheux, and X. Paoletti. 2010. Current challenges for the early clinical development of anticancer drugs in the era of molecularly targeted agents. *Targeted Oncology* 5 (1):65–72. doi:10.1007/s11523-010-0137-6.

Lee, S. M., D. Backenroth, Y. K. K. Cheung, D. L. Hershman, D. Vulih, B. Anderson, P. Ivy, and L. Minasian. 2016. Case example of dose optimization using data from bortezomib dose-finding clinical trials. *Journal of Clinical Oncology* 34 (12):1395–1401. doi:10.1200/JCO.2015.66.0662..

Lee, S. M., and Y. K. Cheung. 2009. Model calibration in the continual reassessment method. *Clinical Trials* 6 (3):227–238. doi:10.1177/1740774509105076.

Maki, R. G., D. R. D'Adamo, M. L. Keohan, M. Saulle, S. M. Schuetze, S. D. Undevia, M. B. Livingston, M. M. Cooney, M. L. Hensley, M. M. Mita, et al. 2009. Phase II study of sorafenib in patients with metastatic or recurrent sarcomas. *Journal of Clinical Oncology* 27 (19):3133–3140. doi:10.1200/JCO.2008.20.4495.

Mathijssen, R. H. J., A. Sparreboom, and J. Verweij. 2014. Determining the optimal dose in the development of anticancer agents. *Nature Reviews Clinical Oncology* 11 (5):272–281. doi:10.1038/nrclinonc.2014.40.

Postel-Vinay, S., C. Gomez-Roca, L. Rhoda Molife, B. Anghan, A. Levy, I. Judson, J. De Bono, J.-C. Soria, S. Kaye, and X. Paoletti. 2011. Phase I trials of molecularly targeted agents: Should we pay more attention to late toxicities? *Journal of Clinical Oncology* 29 (13):1728–1735. doi:10.1200/JCO.2010.31.9236.

Postel-Vinay, S., L. Collette, X. Paoletti, E. Rizzo, C. Massard, D. Olmos, C. Fowst, B. Levy, P. Mancini, D. Lacombe, et al. 2014. Towards new methods for the determination of dose limiting toxicities and the assessment of the recommended dose for further studies of molecularly targeted agents – dose-limiting toxicity and toxicity assessment recommendation group for early trials of targeted therapies, an European organisation for research and treatment of cancer-led study. *European Journal of Cancer* 50 (12):2040–2049. doi:10.1016/j.ejca.2014.04.031.

Riviere, M.-K., Y. Yuan, J.-H. Jourdan, F. Dubois, and S. Zohar. 2018. Phase I/II dose-finding design for molecularly targeted agent: Plateau determination using adaptive randomization. *Statistical Methods in Medical Research* 27 (2):466–479. doi:10.1177/0962280216631763.

Wages, N. A., and C. Tait. 2015. Seamless phase I/II adaptive design for oncology trials of molecularly targeted agents. *Journal of Biopharmaceutical Statistics* 25 (5):903–920. doi:10.1080/10543406.2014.920873.

Winther, S. B., T. Lembrecht Jørgensen, P. Pfeiffer, and C. Qvortrup. 2016. Can we predict toxicity and efficacy in older patients with cancer? Older patients with colorectal cancer as an example. *ESMO Open* 1 (3):e000021. doi:10.1136/esmoopen-2015-000021.

Yuan, Y., and G. Yin. 2009. Bayesian dose finding by jointly modelling toxicity and efficacy as time-to-event outcomes. *Journal of the Royal Statistical Society: Series C (Applied Statistics)* 58 (5):719–736. doi:10.1111/j.1467-9876.2009.00674.x.

Zang, Y. J., J. Lee, and Y. Yuan. 2014. Adaptive designs for identifying optimal biological dose for molecularly targeted agents. *Clinical Trials* 11 (3):319–327. doi:10.1177/1740774514529848.

Zhang, W., D. J. Sargent, and S. Mandrekar. 2006. An adaptive dose-finding design incorporating both toxicity and efficacy. *Statistics in Medicine* 25 (14):2365–2383. doi:10.1002/sim.2325.