Contemporary Clinical Trials Communications 12 (2018) 40–50

Contents lists available at ScienceDirect

Contemporary Clinical Trials Communications

journal homepage: www.elsevier.com/locate/conctc

Overall success rate of a safe and efficacious drug: Results using six phase 1 designs, each followed by standard phase 2 and 3 designs

Amy S. Ruppert[a,b,*], Abigail B. Shoben[b]

[a] Division of Hematology, College of Medicine, The Ohio State University, Columbus, OH, 43210, USA
[b] Division of Biostatistics, College of Public Health, The Ohio State University, Columbus, OH, 43210, USA

ARTICLE INFO

Keywords:
Dose-finding
Phase 1 designs
Overall success

ABSTRACT

To evaluate the overall success rate of a new drug, phase 1, 2, and 3 trials were simulated using eight toxicity and two non-decreasing efficacy profiles. Six phase 1 designs including the standard 3 + 3, CCD, BOIN, mTPI, mTPI-2, and CRM were considered with standard phase 2 and 3 designs.

Based on our results, phase 1 design recommendations are provided when data informing the general shape of the dose-toxicity curve exist. If a large jump in toxicity between dose levels is expected, the standard 3 + 3 design is recommended; it more often recognized when the MTD was exceeded and had the highest overall success rates. If gradually increasing toxicity is expected, a nonstandard design other than the CRM is recommended. Nonstandard designs were more aggressive in dosing and MTD estimation than the standard 3 + 3 and had higher overall success rates, but the CRM was too aggressive and most frequently overestimated the true MTD. If fairly constant, safe toxicity is expected across dose levels, the BOIN or CRM designs are recommended; they escalated to the highest dose most frequently with superior overall success rates.

Without data informing the shape of the dose-toxicity curve, nonstandard phase 1 designs with a modified excessive toxicity rule more easily eliminating unsafe dose levels are recommended. With this modification, MTD overestimation error decreased and overall success rates were similar or higher with nonstandard designs. Among nonstandard designs, the modified CCD and BOIN perform well and are as transparent and simple to implement as the standard 3 + 3 design.

1. Introduction

The primary objective of a phase 1 clinical trial is to determine a safe and tolerable dose level to recommend for further study of efficacy in subsequent phase 2 and 3 trials. Under the assumption that both efficacy and toxicity increase with increasing dose levels, the recommended phase 2 dose is generally the maximum tolerated dose (MTD), defined as the highest dose level where the percentage of patients experiencing predefined dose limiting toxicity (DLT) is below a specified acceptable level. Selection of a dose level that is at or closely below the true MTD is most desirable.

For the past 25 years, the most common dose-finding phase 1 design has been the rule-based standard 3 + 3 design [1–3]. Many have advocated for the use of the model-based continual reassessment method (CRM) for dose-finding [4], but the CRM has been met with resistance due to its unfamiliarity, assumptions that must be made on the shape of the dose-toxicity curve, statistical complexity, need for specialized software, and increased communication required during trial design and implementation [3]. A new type of phase 1 design, the interval design has emerged, and includes the cumulative cohort design (CCD) [5], the modified toxicity probability interval design (mTPI) [6], the Bayesian optimal interval design (BOIN) [7] and the mTPI-2 design [8].

All of these designs except for mTPI-2 have been directly compared to the standard 3 + 3 design and better estimated the true MTD in most scenarios [9]. The CRM was superior in scenarios with six or eight dose levels, followed by the BOIN and then mTPI [9]. However, the ranking of design performance was less clear for smaller dose-finding studies with fewer dose levels. Further, phase 1 design performance has been primarily measured by estimating the percentage of simulations that correctly identify the true MTD and by estimating the average number of simulated patients treated above the true MTD during phase 1. Evaluations from simulation studies rarely measure the downstream effects due to selecting dose levels above or below the true MTD.

Thus, we herein evaluate the performance of all six phase 1 designs (rule-based standard 3 + 3, CCD, BOIN, mTPI, and mTPI-2 interval designs, and model-based CRM design), in the context of a moderately

* Corresponding author. Division of Hematology, College of Medicine, The Ohio State University, Columbus, OH, 43210, USA.
E-mail address: amy.stark@osumc.edu (A.S. Ruppert).

https://doi.org/10.1016/j.conctc.2018.08.010
Received 27 March 2018; Received in revised form 9 August 2018; Accepted 23 August 2018
Available online 24 August 2018

sized phase 1 trial with four escalation dose levels, using overall success rate as a performance measure. Each phase 1 design is followed by Simon's optimal two-stage phase 2 design [10] and a randomized group sequential phase 3 design [11], with overall success rate defined as the percentage of simulations spanning phase 1, 2, and 3 that identify a new drug as safe and efficacious when it actually is safe and efficacious.

Overall success rates are compared by phase 1 design and clinical scenario defined by different dose-toxicity, dose-response, and dose-survival profiles. The impact of excessive toxicity rules and sample size on overall success rates are investigated. Guidelines for phase 1 statistical design choice in different clinical settings are presented, considering the trade-offs between measures of performance with design complexity and ease of implementation.

2. Materials and methods

2.1. Clinical scenarios

In phase 1, five dose levels of a new drug were considered, including four escalation and one de-escalation dose level. Eight toxicity profiles were evaluated: three with the MTD at dose level 2, three with the MTD at dose level 3, and two with the MTD at dose level 4 (Table 1). All but one of the toxicity profiles were monotonically increasing and mirrored shapes that have been commonly included in other phase 1 simulation studies [12–15]. Linear profiles had toxicity probabilities that increased fairly linearly with increasing dose levels, a typical assumption with standard chemotherapy. Jump profiles had a sharp increase in toxicity probability between dose levels 2 and 3, and represented an increase in dose outside the therapeutic window or target saturation. The Plateau profile had increasing toxicity, with smaller increases in toxicity at higher dose levels, which has been described with orally administered, molecularly targeted agents [16]. Constant toxicity profiles had acceptable toxicity with equal probability across dose levels, and have been described with molecularly targeted agents administered within the therapeutic window [17,18]. One nonmonotonic toxicity profile was included and was Tub-shaped. The Tub-shaped toxicity profile had acceptable but moderately high toxicity probabilities at dose levels −1 and 1, lower toxicity probabilities at dose levels 2 and 3, and a sudden increase in toxicity probability above the acceptable level at dose level 4. This toxicity profile represented a scenario in which disease-related adverse events are observed at low inactive dose levels and called DLTs [19]. As the drug becomes more active at higher dose levels and disease-related adverse events are no longer observed, the DLT rate then decreases. Eventually the drug is delivered at a dose level outside the therapeutic window and the DLT rate increases once again.

Each of the eight toxicity profiles was mapped to a Continuous response/survival profile and a Step response/survival profile (Table 1). Continuous response profiles occurred with Continuous survival profiles and represented therapy that had steadily increasing efficacy with increasing dose levels. Step response profiles occurred with Step survival profiles and represented agents that remained inactive until critical mass was reached between dose levels 1 and 2. In the efficacy profiles evaluated, response rate was not lower than 5% and median survival was not shorter than 6 months at any dose level, the response rate and median survival assumed for the standard of care. Scenarios in which a safe and efficacious drug existed were of primary interest, and so all profiles included the optimal or target response rate and median survival at the true MTD. In this study, the target response rate was 20% and the target median survival was 9 months, corresponding to a hazard ratio of 0.67 when compared to standard of care and assuming exponential survival times. Scenarios with suboptimal response or suboptimal survival at the true MTD were not explored.

Collectively, eight toxicity profiles and two efficacy profiles were simulated, resulting in 16 total scenarios. Six phase 1 designs (i.e., standard 3 + 3, CCD, BOIN, mTPI, mTPI-2, CRM), each followed by Simon's optimal two-stage phase 2 design [10] and a two-arm randomized group sequential phase 3 design [11], were applied to each clinical scenario.

Table 1
Assumed toxicity, response, and survival profiles across dose levels. Toxicity, response, and survival, respectively, are the true proportion of DLT, true response proportion, and true median survival in months at each dose level.

MTD at Dose Level 2

Dose Level	Toxicity Profiles			Continuous Efficacy		Step Efficacy	
	Linear A	Jump A	Jump B	Response	Survival	Response	Survival
−1	0.10	0.05	0.20	0.10	7	0.05	6
1	0.20	0.05	0.20	0.15	8	0.05	6
2	0.30	0.05	0.20	0.20	9	0.20	9
3	0.40	0.60	0.40	0.25	10	0.20	9
4	0.50	0.60	0.40	0.30	11	0.20	9

MTD at Dose Level 3

Dose Level	Toxicity Profiles			Continuous Efficacy		Step Efficacy	
	Linear B	Plateau	Tub	Response	Survival	Response	Survival
−1	0.05	0.05	0.20	0.05	6	0.05	6
1	0.10	0.15	0.20	0.10	7	0.05	6
2	0.20	0.25	0.10	0.15	8	0.20	9
3	0.30	0.30	0.10	0.20	9	0.20	9
4	0.40	0.35	0.35	0.25	10	0.20	9

MTD at Dose Level 4

Dose Level	Toxicity Profiles		Continuous Efficacy		Step Efficacy	
	Constant A	Constant B	Response	Survival	Response	Survival
−1	0.05	0.20	0.05	6	0.05	6
1	0.05	0.20	0.05	6	0.05	6
2	0.05	0.20	0.10	7	0.20	9
3	0.05	0.20	0.15	8	0.20	9
4	0.05	0.20	0.20	9	0.20	9

2.2. Description of phase 1 designs

The standard 3 + 3 design is a rule-based design in which patients are enrolled in cohorts of three, beginning at the starting dose level [1]. If there are no DLTs in the first cohort of three patients treated at a dose level, the dose is escalated. If one DLT is observed in the first cohort of three patients, a second cohort of three patients is treated at the same dose level. If at most one DLT is observed in six patients at a dose level, then escalation to the next highest dose level is permitted. At a dose level with two or more DLTs, the MTD has been exceeded and the dose is de-escalated until at most one DLT is observed in a total of six patients.

The CCD is an interval design in which a target DLT rate (p_t) and small fractions of error (e_1 and e_2) about p_t are specified to form a proper-dosing interval (p_t - e_1, p_t + e_2) [5]. Throughout the trial, the observed DLT rate at a dose level is compared to the proper-dosing interval to make dosing decisions. The decision to escalate, stay at the same dose level, or de-escalate corresponds respectively with whether the observed DLT rate at the current dose level is below, within, or above the proper-dosing interval. As in all interval designs, the MTD is estimated at the end of the trial after applying isotonic regression to estimated DLT probabilities at each dose level and selecting the dose level with estimated DLT probability closest to p_t.

The BOIN is an interval design similar to the CCD [7]. Dosing decisions are based on the observed DLT rate as compared to the proper-dosing interval. However, the recommended proper-dosing interval for a given p_t is different between the CCD and BOIN designs.

The mTPI design is the Bayesian analog of the CCD design [6]. With