



Published in final edited form as:

*Cancer*. 2018 August ; 124(16): 3339–3345. doi:10.1002/cncr.31577.

## Statistical Justification of Expansion Cohorts in Phase I Cancer Trials

Ali A. Mokdad, MD, MS<sup>1,2,3</sup>, Xian-Jin Xie, PhD<sup>4</sup>, Hong Zhu, PhD<sup>2,5</sup>, David E. Gerber, MD<sup>2,5,6</sup>, and Daniel F. Heitjan, PhD<sup>3,5</sup>

<sup>1</sup>Department of Surgery, University of Texas Southwestern Medical Center, Dallas, TX

<sup>2</sup>Harold C. Simmons Comprehensive Cancer Center, University of Texas Southwestern, Dallas, TX

<sup>3</sup>Department of Statistical Science, Southern Methodist University, Dallas, TX

<sup>4</sup>College of Dentistry and College of Public Health, University of Iowa, Iowa City, IA

<sup>5</sup>Department of Clinical Sciences, University of Texas Southwestern Medical Center, Dallas, TX

<sup>6</sup>Department of Internal Medicine, University of Texas Southwestern Medical Center, Dallas, TX

### Abstract

**Background**—Phase I cancer trials increasingly incorporate dose-expansion cohorts (DECs), reflecting a growing demand to acquire more information about investigational drugs. Protocols commonly fail to provide a sample-size justification or analysis plan for the DEC. In this study, we develop a statistical framework for the design of DECs.

**Methods**—We assume the maximum tolerated dose (MTD) for the investigational drug has been identified in the dose-escalation stage of the trial. We use the 80% lower confidence bound and the 90% upper confidence bound for the response and toxicity rates, respectively, as decision thresholds for the dose-expansion stage. We calculate the operating characteristics with reference to prespecified minimum effective response rates and maximum safe DLT rates.

**Results**—We apply our framework to specify a system of DEC plans. The design comprises three components: 1) the number of subjects enrolled at the MTD, 2) the minimum number of responses necessary to indicate provisional drug efficacy, and 3) the maximum number of dose-limiting toxicities (DLTs) permitted to indicate drug safety. We demonstrate our method in an application to a cancer immunotherapy trial.

**Conclusions**—Our simple and practical tool enables creation of DEC designs that appropriately address the safety and efficacy objectives of the trial.

---

**Corresponding author:** Daniel F. Heitjan, PhD, Department of Clinical Sciences, University of Texas Southwestern Medical Center, 5323 Harry Hines Blvd., Dallas, TX 75390; Phone: 214-648-4832; daniel.heitjan@utsouthwestern.edu.

**Disclosures:** None

**Author contributions:** Writing-Original Draft: AM, DH; Writing-Review & Editing: AM, XX, HZ, DG, DH; Conceptualization and Methodology: AM, XX, HZ, DH; Analysis and interpretation of results: AM, XX, HZ, DG, DH.

## Keywords

Dose-finding; dose-expansion; safety; efficacy; design

---

## INTRODUCTION

A typical phase I cancer clinical trial evaluates the safety of an antineoplastic agent not previously studied in humans. It commonly tests the drug initially in a dose-escalation phase in which cohorts of patients receive increasing doses until reaching a maximum tolerated dose (MTD). The cohort receiving the MTD is often small, typically only 6 subjects.<sup>1,2</sup> To further elucidate the drug's properties, investigators may enroll an additional group of subjects, denoted a *dose-expansion cohort* (DEC), once the trial has identified the MTD.<sup>3</sup>

Trialists originally incorporated DEC into phase I trials to enable further study of safety and tolerability at the MTD. More recently, the role of the DEC has evolved to allow for acquiring preliminary efficacy information; characterizing pharmacokinetic and pharmacodynamic properties; and exploring outcomes in designated patient strata, such as those manifesting specific molecular aberrations or clinical features.<sup>3,4</sup>

The expanded role of DEC reflects an impulse to expedite drug development by identifying promising drugs early and eliminating drugs with little chance of approval.<sup>5</sup> As a result, there is increased emphasis on monitoring efficacy in phase I trials and in exploring novel surrogates for efficacy, particularly for molecularly targeted drugs whose toxicity is an unreliable predictor of efficacy.<sup>1,3-7</sup>

The use of DEC is becoming more common. In a systematic review of published phase I cancer trials, the fraction incorporating a DEC increased from 12% in 2006 to 38% in 2011.<sup>4</sup> Extrapolating to all 2,800 phase I cancer trials registered on [clinicaltrials.gov](http://clinicaltrials.gov) in the past 5 years, we estimate that more than 1,000 of these trials included a DEC.<sup>8</sup>

Although there is a rich literature on designs that involve the collection of both safety and efficacy data within the current phase I/phase II taxonomy, few papers address phase I/DEC designs as they are currently implemented.<sup>3,4</sup> Unsurprisingly, an explication of statistical considerations for the DEC design is often absent from the protocol: Justification for DEC sample size is missing; the designation of endpoints is vague; and rules for evaluating trial success are lacking.

In this study, we propose a simple statistical framework to guide the design of DEC for phase I cancer trials. Retaining the currently popular framework for such studies, the key to our approach is the prospective specification of the efficacy and toxicity criteria that define trial success. These together yield a method for selecting the DEC sample size and conducting the data analysis.

## METHODS

### DOSE-ESCALATION AND DETERMINATION OF THE MTD

In the dose-escalation stage of a phase I trial, we monitor subjects for dose-limiting toxicities (DLTs), which are treatment-related adverse events defined *a priori* to satisfy criteria of type, severity, and time of occurrence. The aim is to identify the MTD, if one exists among the proposed doses. We describe our method in the context of a 3+3 design (Appendix 1), which is the most familiar and commonly implemented dose-escalation design,<sup>9,10</sup> but one can use other approaches as well.

### DOSE-EXPANSION: THE DEC DESIGN

When dose escalation is complete, the trial begins enrolling subjects into a DEC at the declared MTD (or the highest planned dose if escalation did not reach the MTD). The aims of a DEC are twofold: 1) to obtain a preliminary assessment of the test drug's efficacy, and 2) to further evaluate the safety of the drug at the MTD. Translating these objectives into statistical hypotheses permits determination of the DEC design. The following sections detail our approach to specifying the DEC plan: Its sample size; the number of responses necessary to deem a drug provisionally "effective"; and the number of DLTs necessary to deem it "unsafe".

### PROVISIONAL EFFICACY

Even at the early phases of drug evaluation, investigators recognize a minimum response probability below which further development would be futile. Thus, one can base a decision to continue development on the observation that the true response rate is likely to exceed a prespecified minimum response fraction. A statistical tool that summarizes the data to address this decision is the lower confidence bound (LCB)—the lower limit of a one-sided confidence interval. If the LCB exceeds the prespecified minimum response fraction, then there is preliminary evidence that the drug is effective, justifying further testing. If the LCB falls below the prespecified minimum response probability, there remains substantial uncertainty that the drug is effective. We use these considerations as a partial basis for the sample size and analysis plan for a DEC.

We assume the response outcome is dichotomous. The sample size, denoted  $n$ , includes subjects who receive the MTD during either the final stage of escalation or the DEC. We report the total sample size at the MTD instead of the DEC size to allow for flexibility in the number of subjects tested at the MTD during the dose-escalation phase, which typically is 6 subjects in a 3+3 design. We calculate the DEC sample size as the targeted  $n$  minus the number of subjects receiving the MTD during the dose escalation.

We calculate the 80% LCB for response counts  $r = 0, 1, \dots, n$  using the Clopper-Pearson method.<sup>11,12</sup> We use an 80% confidence coefficient, rather than the conventional 90% or 95%, because the DEC portion of the trial is intended to be suggestive, not definitive.

We use the LCB for response to determine the minimum number of responses (NR) needed to declare the drug effective. Appendix 2 describes computation of NR.

## A STOPPING BOUND FOR TOXICITY

We propose designating a stopping bound for toxicity based on a DLT probability threshold, specified *a priori* by the investigator, at which we deem the drug unacceptably toxic. In a manner comparable to the efficacy assessment, we use the upper confidence bound (UCB), the upper limit of a one-sided confidence interval, to judge the drug's safety at the MTD. We declare the drug provisionally safe at the identified MTD if UCB is less than the prespecified DLT probability threshold. We declare the MTD provisionally unsafe, and therefore in need of revision downward, if UCB exceeds the threshold.

Defining  $d$  to be the count of DLTs in  $n$  subjects tested, we calculate the 90% UCB for the probability of a DLT at each possible DLT outcome, again using the Clopper-Pearson method.<sup>11,12</sup> With the 3+3 dose-escalation plan, this calculation does not fully account for the design, in that only doses that yield at most 1 DLT (in 6 subjects) can proceed to the dose-expansion phase. Thus, the distribution of DLTs is truncated to exclude values  $d > n-5$ . However, simulations reveal that the Clopper-Pearson UCB is adequate, particularly for DLT probability thresholds commonly used in phase I cancer trials (Appendix 3).

One derives the stopping bound by choosing the minimum number of DLTs, denoted ND, for which the 90% UCB exceeds the prespecified maximum DLT probability. If the number of subjects receiving the MTD and experiencing a DLT reaches ND, one stops enrolling subjects at the MTD and declares the MTD unsafe.

## OPERATING CHARACTERISTICS OF THE DEC DESIGN

We evaluate the operating characteristics of various versions of our DEC designs as means to distinguish among these designs and to guide investigators to choose the DEC plan and sample size that best fits the trial.

A desirable DEC design has the following properties: 1) It rejects with high probability a treatment with an unsatisfactory response rate. 2) It accepts with high probability a treatment with a desirable response rate. 3) It rejects with high probability a treatment with a DLT rate that is higher than the prespecified maximum DLT rate. 4) It accepts with high probability a treatment with a safe DLT rate. Appendix 4 presents the formulas for the operating characteristics of our DEC design.

## RESULTS

### PROVISIONAL EFFICACY

Figure 1 presents 80% LCBs at various response counts for sample sizes 15, 20, 30, and 40. Using 80% LCBs, one can identify, for a specific sample size, the minimum number of responses, denoted NR above, necessary to claim a drug has provisional efficacy. For example, Figure 1C shows that if the minimum response probability is set at 5%, then among 30 subjects tested at the MTD, one declares the drug provisionally effective if there are at least 3 responses. The 80% LCB for 3 responses among 30 patients is 5.2%, which exceeds the prespecified minimum response probability, 5%. One uses the same algorithm to identify NRs at other samples sizes.

## A STOPPING BOUND FOR TOXICITY

Figure 1 also depicts 90% UCBs for sample sizes 15, 20, 30, and 40, superimposed on the same graphs as the LCBs for response rate. One can use Figure 1 in an analogous manner to determine the stopping bound for toxicity, specifically, the number of DLTs, denoted ND above, that marks a drug unsafe. For example, if the maximum DLT probability is set at 30%, one stops for toxicity if the trial reaches 6 DLTs in a cohort of size 30. The 90% UCB for 6 DLTs among 30 patients is 32.5%, which exceeds the prespecified maximum DLT probability, 30%. Thus, upon reaching the 6<sup>th</sup> DLT, one suspends enrollment for lack of safety, potentially abandoning the drug or revising its dose downward.

## OPERATING CHARACTERISTICS OF THE DEC DESIGN

Table 1 summarizes the operating characteristics for DEC designs with a 5% minimum response probability, a 33% maximum DLT probability, and sample sizes between 15 and 40 subjects. The operating characteristics for DEC designs at some other probability combinations appear in Appendix Tables A.1, A.2, and A.3.

Table 1 reveals features of the DEC design:

1. The probability of rejecting a drug that has a 5% response rate is at least 81%, and the probability of declaring the MTD too toxic, given a DLT rate greater than 33% at the MTD, is at least 79%.
2. The probability of declaring a drug effective varies according to sample size: Among sample sizes with the *same* NR, the probability of accepting a drug increases as sample size increases. Notably, however, if the NR is not the same for two different sample sizes, there is no guarantee that a larger  $n$  will give a larger probability of accepting the drug at the same desirable response rate. For instance, the NR for sample size  $n=16$  is 2, and the probability of accepting an “effective” drug is 86%. Whereas for sample size  $n=17$ , NR is 3, but the probability of accepting the drug is only 69%.

The probability of accepting a drug that is safe at the MTD varies markedly across sample sizes, and may be as low as 50%. Moreover, a larger sample size does not guarantee a higher probability of accepting a drug at a safe DLT rate. For instance, for sample sizes that have the same ND, a smaller sample size produces a higher probability of accepting a drug that is safe at the DLT.

## EXAMPLE: NIVOLUMAB FOR REFRACTORY SOLID TUMORS

We demonstrate our method using a phase I trial reported by Brahmer *et al.*<sup>13</sup> Their study evaluated the safety and efficacy of nivolumab (MDX-1106), a monoclonal antibody that enhances the responsiveness of immune cells to tumors by blocking the programmed death-1 coreceptor.<sup>14</sup> Patients with advanced lung, colorectal, melanoma, renal cell, or prostate cancer refractory to prior treatments were eligible. Cohorts of participants were to receive MDX-1106 at 0.3, 1, 3, and 10 mg/Kg according to a 3+3 design, followed by a DEC at the MTD. Investigators evaluated tumor response radiographically at 8 and 12 weeks. They

defined DLT as a grade 3 or 4 treatment-related adverse event or laboratory abnormality occurring within 28 days of drug administration.

We design the DEC for a minimum response rate of 5% and a maximum DLT rate of 33%. With these two inputs, we examine Table 1 to identify the DEC design with the most desirable operating characteristics; “desirable” here is subjective and reflects the investigator’s knowledge of the cancer under study, the properties of the test drug and similar drugs, and the resources available to the trial.

In a setting of constrained resources, for instance, enrolling 14 subjects into the expansion cohort (amounting to  $n=20$  at the MTD—6 from the escalation and 14 from the DEC) produces reasonable operating characteristics. With this sample size, we deem nivolumab effective if 3 or more responses are recorded (NR=3) and excessively toxic if 4 or more subjects suffer a DLT (ND=4) (Figure 1). The probability of rejecting nivolumab is at least 92% if its true response rate is less than 5% and is at least 85% if its true DLT rate is greater than 33% (Table 1). By comparison, the probability of accepting nivolumab—and deeming the trial successful—exceeds 79% if the true response rate exceeds 20%, and is at least 67% if its true DLT rate is less than 17%.

The actual trial enrolled 15 subjects into the DEC, for a total of  $n=21$  receiving the highest planned dose. The study did not reach the MTD, as the highest planned dose produced no DLTs. By including only 14 subjects rather than 15 in the expansion cohort, our hypothetical DEC design requires one less subject, has a similar probability of rejecting the drug if its true response rate is less than a 5%, and has a higher probability of deeming the trial successful if the drug’s true DLT rate is less than 17%. More importantly, the design reflects objective decision criteria defined prior to observing any study data.

## DISCUSSION

We propose a simple method to design and analyze the DEC in a phase I cancer trial where one seeks a preliminary evaluation of drug efficacy and safety. Its components are the planned sample size and targeted response and DLT rates. Our method requires investigators to make explicit the characteristics of the trial design in light of their notions of drug safety and effectiveness.

Our method can accommodate multiple DECs, as would be applicable if one sought to acquire information on treatment performance in strata determined by patient baseline characteristics.<sup>4</sup> For instance, in a trial that enrolls patients with different tumors, one could recruit a separate DEC for each tumor type represented in the escalation cohort. If there is no reason to expect toxicity to differ among cohorts, then one can apply a global stopping bound for toxicity, whereas response thresholds can be DEC-specific.

Numerous designs, many deploying variants of the continual reassessment method (CRM), simultaneously address safety and efficacy in early-phase trials.<sup>15–27</sup> As the DEC is a recent innovation, only a few papers recognize it as a formal design feature.<sup>28–31</sup> The emerging consensus is that 1) with the modest sample sizes of most phase I trials, one cannot estimate the MTD precisely; 2) the CRM is superior to the 3+3 as a general dose-finding tool; and 3)

extending the study beyond its usual stopping point—either by recruiting a DEC or conducting additional CRM iterations—improves the estimated MTD. What then is the rationale for using our proposed approach?

First, compared to other designs it is simple to apply, and one can readily append it to the 3+3 design or any other dose-seeking plan.

Second, our design reflects well-established ideas of how one should conduct the early stages of a clinical research program. In the face of considerable uncertainty about a drug's effects, it is reasonable to begin with a low dose, gradually escalate until toxicities appear, and then treat a few more patients to search for an efficacy signal.

Finally, our design conforms to the principle of “letting the students decide where the sidewalks should go”. The increasing use of DEC designs reflects the desire to unite the preliminary evaluations of safety and efficacy under a single protocol, rather than separate them arbitrarily into “phases”.

One can apply our DEC design with any dose-escalation plan, not just the 3+3. We generally favor the 3+3 for its simplicity and because it sensitively and conservatively detects DLTs that occur in the 15%–25% range.<sup>31</sup> Such rates are likely to be of considerable concern to the patient and the clinician, regardless of the drug or the disease. Indeed, in simulation studies, a 3+3/DEC design was more likely than the CRM to correctly select the MTD with a desired 30% DLT rate for drugs that are highly toxic at low doses.<sup>32</sup> More importantly, for all simulated dose-toxicity curves the 3+3/DEC was less likely to sojourn above the target MTD. The principal advantages of the CRM are that one can tune it to identify any target DLT rate, and that it is efficient, in that it will find the MTD rapidly if one exists. This is particularly important in studies with large numbers of candidate dose levels.

Unsurprisingly, clinicians prefer simple and familiar methods, which likely accounts for the continued popularity of 3+3 even 25 years after the promulgation of CRM.<sup>27</sup> (Similarly, the Simon (1989)<sup>33</sup> and Gehan (1961)<sup>34</sup> phase II designs continue to be popular despite many subsequent advances in the statistical design of such trials.) Our approach superimposes on this framework a way to justify a dose-escalation/DEC study by providing a means to signal a problem with safety or efficacy, and thereby to prompt a formal reconsideration of the drug's usefulness.

What does the future hold for early-phase cancer trial design? The work of O'Quigley and others<sup>31</sup> points to the creation of integrated systems for the collection and analysis of safety and efficacy data, likely founded on the Bayesian paradigm rather than the traditional idea of minimizing long-term error rates of rigidly enforced decision rules. It will be critical for biostatisticians to demonstrate the practical advantages of the integrated approach, as well as its feasibility in modestly funded, single-center, clinical research programs.

In conclusion, phase I trials in cancer increasingly incorporate DEC designs in order to obtain additional safety information together with preliminary drug efficacy data. We have proposed a simple practical tool to generate DEC designs that appropriately address the objectives of the extended phase I study.

## Supplementary Material

Refer to Web version on PubMed Central for supplementary material.

## Acknowledgments

National Institute of Health grants K24 CA201543 (to DEG); 1P30 CA142543; TL1 TR001104; and UL1 RR024982.

**Funding support:** None

## References

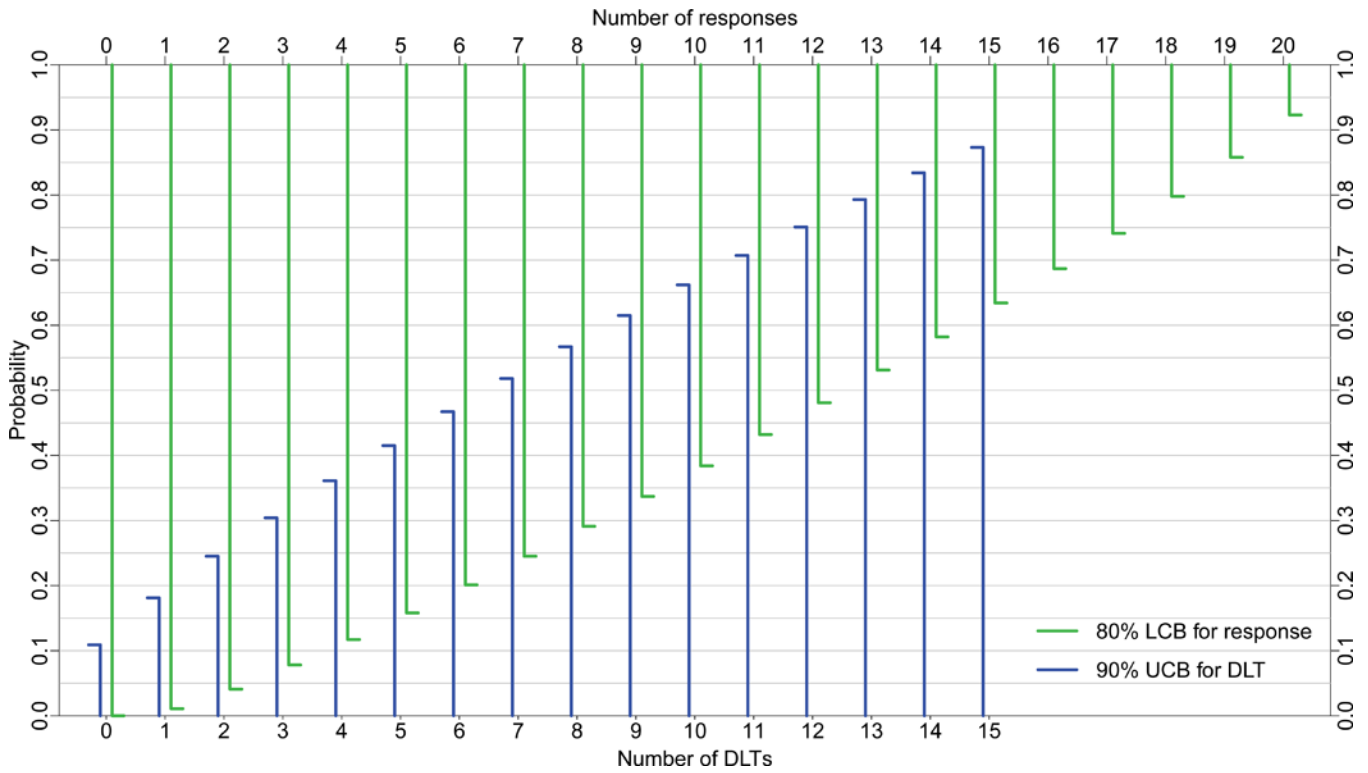
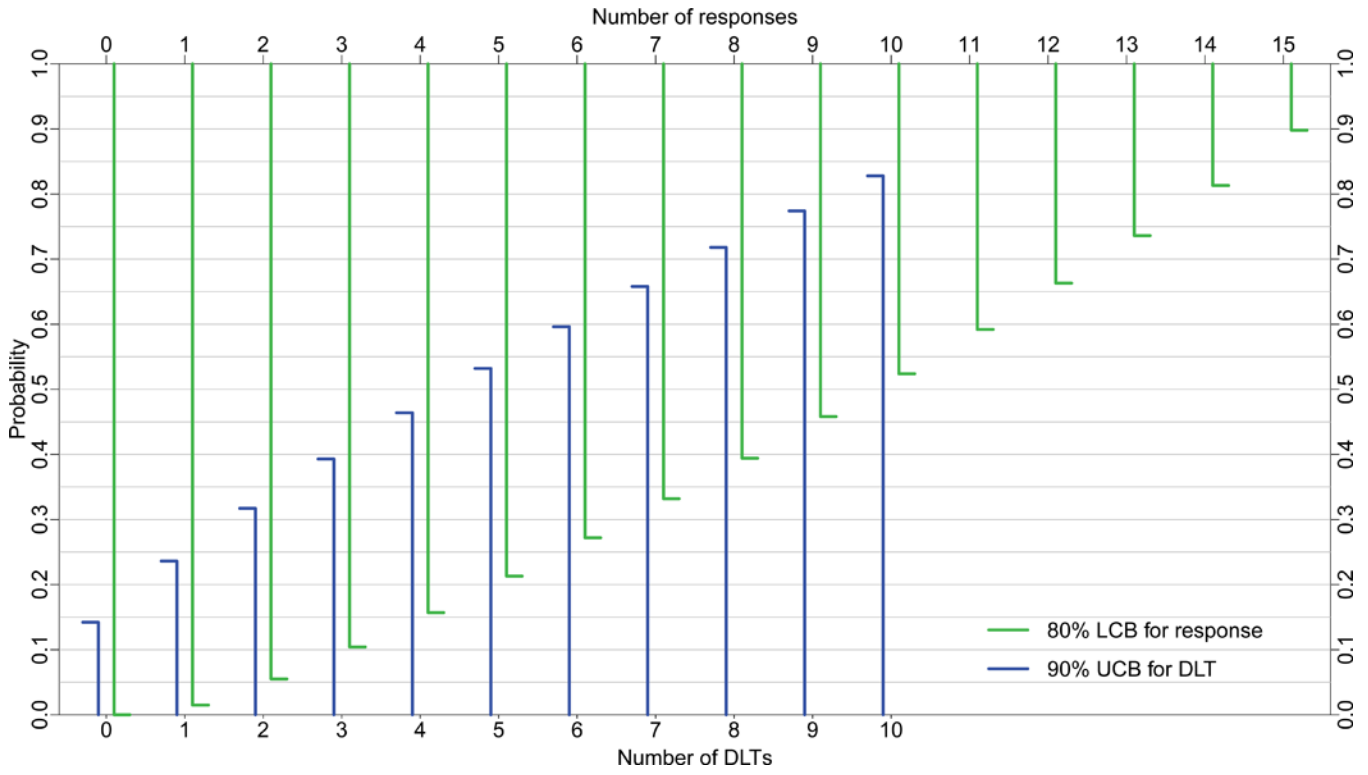
- Ivy SP, Siu LL, Garrett-Mayer E, Rubinstein L. Approaches to phase I clinical trial design focused on safety, efficiency, and selected patient populations: A report from the Clinical Trial Design Task Force of the National Cancer Institute Investigational Drug Steering Committee. *Clin Cancer Res.* 2010; 16(6):1726–1736. DOI: 10.1158/1078-0432.CCR-09-1961 [PubMed: 20215542]
- Korn EL, Midthune D, Chen TT, Rubinstein LV, Christian MC, Simon RM. A comparison of two phase I trial designs. *Stat Med.* 1994; 13(18):1799–1806. [PubMed: 7997713]
- 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. DOI: 10.1200/JCO.2012.47.4957 [PubMed: 24127441]
- Dahlberg SE, Shapiro GI, Clark JW, Johnson BE. Evaluation of statistical designs in phase I expansion cohorts: the Dana-Farber/Harvard Cancer Center experience. *J Natl Cancer Inst.* 2014; 106(7)doi: 10.1093/jnci/dju163
- Wong KM, Capasso A, Eckhardt SG. The changing landscape of phase I trials in oncology. *Nat Rev Clin Oncol.* 2016; 13(2):106–117. DOI: 10.1038/nrclinonc.2015.194 [PubMed: 26552953]
- Workman P, Aboagye EO, Chung Y-L, et al. Minimally invasive pharmacokinetic and pharmacodynamic technologies in hypothesis-testing clinical trials of innovative therapies. *J Natl Cancer Inst.* 2006; 98(9):580–598. DOI: 10.1093/jnci/djj162 [PubMed: 16670384]
- LoRusso PM, Boerner SA, Seymour L. An overview of the optimal planning, design, and conduct of phase I studies of new therapeutics. *Clin Cancer Res.* 2010; 16(6):1710–1718. DOI: 10.1158/1078-0432.CCR-09-1993 [PubMed: 20215546]
- National Institutes of Health. ClinicalTrials.gov. <https://clinicaltrials.gov/>. Accessed February 10, 2016
- Le Tourneau C, Lee JJ, Siu LL. Dose escalation methods in phase I cancer clinical trials. *J Natl Cancer Inst.* 2009; 101(10):708–720. DOI: 10.1093/jnci/djp079 [PubMed: 19436029]
- Rogatko A, Schoeneck D, Jonas W, Tighiouart M, Khuri FR, Porter A. Translation of innovative designs into phase I trials. *J Clin Oncol.* 2007; 25(31):4982–4986. DOI: 10.1200/JCO.2007.12.1012 [PubMed: 17971597]
- Clopper CJ, Pearson ES. The use of confidence or fiducial limits illustrated in the case of the binomial. *Biometrika.* 1934; 26(4):404–413. DOI: 10.2307/2331986
- Leemis LM, Trivedi KS. A comparison of approximate interval estimators for the Bernoulli parameter. *Am Stat.* 1996; 50(1):63–68. DOI: 10.1080/00031305.1996.10473544
- Brahmer JR, Drake CG, Wollner I, et al. Phase I study of single-agent anti-programmed death-1 (MDX-1106) in refractory solid tumors: safety, clinical activity, pharmacodynamics, and immunologic correlates. *J Clin Oncol.* 2010; 28(19):3167–3175. DOI: 10.1200/JCO.2009.26.7609 [PubMed: 20516446]
- Keir ME, Butte MJ, Freeman GJ, Sharpe AH. PD-1 and its ligands in tolerance and immunity. *Annu Rev Immunol.* 2008; 26:677–704. DOI: 10.1146/annurev.immunol.26.021607.090331 [PubMed: 18173375]
- Bryant J, Day R. Incorporating toxicity considerations into the design of two-stage phase II clinical trials. *Biometrics.* 1995; 51(4):1372–1383. DOI: 10.2307/2533268 [PubMed: 8589229]

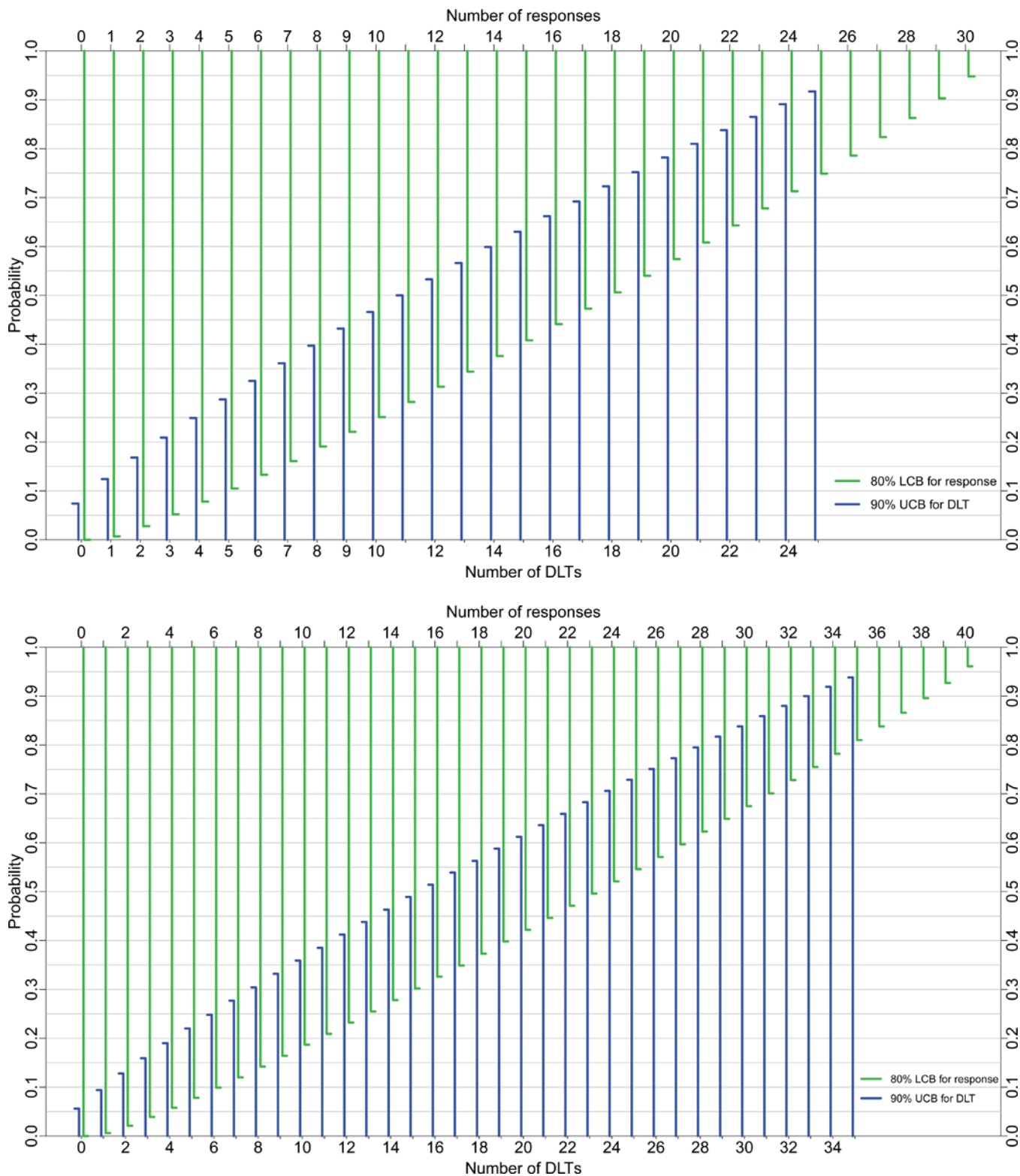


16. Thall PF, Simon RM, Estey EH. Bayesian sequential monitoring designs for single-arm clinical trials with multiple outcomes. *Stat Med*. 1995; 14(4):357–379. DOI: 10.1002/sim.4780140404 [PubMed: 7746977]
17. O’Quigley J, Hughes MD, Fenton T. Dose-finding designs for HIV studies. *Biometrics*. 2001; 57(4):1018–1029. DOI: 10.1111/j.0006-341X.2001.01018.x [PubMed: 11764240]
18. Braun TM. The bivariate continual reassessment method: Extending the CRM to phase I trials of two competing outcomes. *Control Clin Trials*. 2002; 23(3):240–256. DOI: 10.1016/S0197-2456(01)00205-7 [PubMed: 12057877]
19. Ivanova A. A new dose-finding design for bivariate outcomes. *Biometrics*. 2003; 59(4):1001–1007. DOI: 10.1111/j.0006-341X.2003.00115.x [PubMed: 14969479]
20. Zhang W, Sargent DJ, Mandrekar S. An adaptive dose-finding design incorporating both toxicity and efficacy. *Stat Med*. 2006; 25(14):2365–2383. DOI: 10.1002/sim.2325 [PubMed: 16220478]
21. Dragalin V, Fedorov VV, Wu Y. Two-stage design for dose-finding that accounts for both efficacy and safety. *Stat Med*. 2008; 27(25):5156–5176. DOI: 10.1002/sim.3356 [PubMed: 18680164]
22. Yuan Y, Yin G. Bayesian dose finding by jointly modelling toxicity and efficacy as time-to-event outcomes. *J R Stat Soc Ser C Appl Stat*. 2009; 58(5):719–736. DOI: 10.1111/j.1467-9876.2009.00674.x
23. Berry SM, Spinelli W, Littman GS, et al. A Bayesian dose-finding trial with adaptive dose expansion to flexibly assess efficacy and safety of an investigational drug. *Clin Trials*. 2010; 7(2): 121–135. DOI: 10.1177/1740774510361541 [PubMed: 20338905]
24. Houede N, Thall PF, Nguyen H, Paoletti X, Kramar A. Utility-based optimization of combination therapy using ordinal toxicity and efficacy in phase I/II trials. *Biometrics*. 2010; 66(2):532–540. DOI: 10.1111/j.1541-0420.2009.01302.x [PubMed: 19673865]
25. Lei X, Yuan Y, Yin G. Bayesian phase II adaptive randomization by jointly modeling time-to-event efficacy and binary toxicity. *Lifetime Data Anal*. 2011; 17(1):156–174. DOI: 10.1007/s10985-010-9163-z [PubMed: 20364321]
26. Thall PF, Herrick RC, Nguyen HQ, Venier JJ, Norris JC. Effective sample size for computing prior hyperparameters in Bayesian phase I–II dose-finding. *Clin Trials*. 2014; 11(6):657–666. DOI: 10.1177/1740774514547397 [PubMed: 25179541]
27. O’Quigley J, Pepe M, Fisher L. Continual reassessment method: A practical design for phase I clinical trials in cancer. *Biometrics*. 1990; 46(1):33–48. [PubMed: 2350571]
28. Gonen M. A Bayesian evaluation of enrolling additional patients at the maximum tolerated dose in Phase I trials. *Contemp Clin Trials*. 2005; (26):131–140. [PubMed: 15837436]
29. Iasonos A, Ostrovnaya I. Estimating the dose-toxicity curve in completed phase I studies. *Stat Med*. 2011; 30(17):2117–2129. DOI: 10.1002/sim.4206 [PubMed: 21341302]
30. Iasonos A, O’Quigley J. Design considerations for dose-expansion cohorts in phase I trials. *J Clin Oncol*. 2013; 31(31):4014–4021. DOI: 10.1200/JCO.2012.47.9949 [PubMed: 24101039]
31. Iasonos A, O’Quigley J. Sequential monitoring of Phase I dose expansion cohorts. *Stat Med*. 2017; 36(2):204–214. DOI: 10.1002/sim.6894 [PubMed: 26854196]
32. 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)doi: 10.1093/jnci/dju429
33. Simon R. Optimal two-stage designs for phase II clinical trials. *Control Clin Trials*. 1989; 10(1):1–10. DOI: 10.1016/0197-2456(89)90015-9 [PubMed: 2702835]
34. Gehan EA. The determination of the number of patients required in a preliminary and a follow-up trial of a new chemotherapeutic agent. *J Chronic Dis*. 1961; 13(4):346–353. DOI: 10.1016/0021-9681(61)90060-1 [PubMed: 13704181]

### Summary

Phase I cancer trials increasingly incorporate dose-expansion cohorts (DECs) that are not subject to the same high level of statistical scrutiny as other trial components. We describe a statistical framework for the design of a DEC that involves prospectively specifying the efficacy and toxicity criteria that define trial success.





**Figure 1.** 80% lower confidence bounds (LCBs, green) for response and 90% upper confidence bounds (UCB, blue) for toxicity for sample sizes A) 15, B) 20, C) 30, and D) 40 subjects;

DEC sizes are A) 9, B) 14, C) 24, and D) 34 subjects, respectively. DLT= dose-limiting toxicity.

Author Manuscript

Author Manuscript

Author Manuscript

Author Manuscript

Operating characteristics of the DEC design. The minimum number of responses (NR) necessary to claim provisional efficacy, and the number of DLTs (ND) at which the MTD is deemed toxic are estimated according to an 80% lower confidence bound (LCB) that exceeds a prespecified minimum response probability  $p_{Rmir}=0.05$  and a 90% upper confidence bound (UCB) that is less than a prespecified maximum acceptable DLT probability  $p_{Dmax}=0.33$ , respectively.

**Table 1**

Sample size <sup>a</sup>	DEC size <sup>b</sup>	Number of responses (NR)	Number of DLTs (ND, toxicity stopping bound)	Pr[reject] at $p_{Rmir}=0.05$	Pr[accept] at $p_{Rdesirable}=0.20$	Pr[reject] at $p_{Dmax}=0.33$	Pr[accept] at $p_{Dsafe}=0.17$
15	9	2	3	0.83	0.83	0.79	0.66
16	10	2	3	0.81	0.86	0.84	0.6
17	11	3	3	0.95	0.69	0.88	0.55
18	12	3	3	0.94	0.73	0.91	0.5
19	13	3	4	0.93	0.76	0.81	0.71
20	14	3	4	0.92	0.79	0.85	0.67
21	15	3	4	0.92	0.82	0.88	0.63
22	16	3	4	0.91	0.85	0.91	0.58
23	17	3	5	0.89	0.87	0.82	0.75
24	18	3	5	0.88	0.89	0.86	0.72
25	19	3	5	0.87	0.90	0.89	0.68
26	20	3	6	0.86	0.92	0.80	0.82
27	21	3	6	0.85	0.93	0.84	0.79
28	22	3	6	0.84	0.94	0.87	0.76
29	23	3	6	0.82	0.95	0.89	0.72
30	24	3	7	0.81	0.96	0.82	0.84
31	25	4	7	0.93	0.89	0.85	0.81
32	26	4	7	0.93	0.91	0.88	0.79
33	27	4	7	0.92	0.92	0.90	0.76
34	28	4	8	0.91	0.93	0.84	0.86
35	29	4	8	0.90	0.94	0.86	0.84
36	30	4	8	0.90	0.95	0.89	0.81

Sample size <sup>a</sup>	DEC size <sup>b</sup>	Number of responses (NR)	Number of DLTs (ND, toxicity stopping bound)	Pr[reject] at $p_{Rmin}=0.05$	Pr[accept] at $p_{Rdesirable}=0.20$	Pr[reject] at $p_{Dmax}=0.33$	Pr[accept] at $p_{Dsafe}=0.17$
37	31	4	9	0.89	0.96	0.82	0.89
38	32	4	9	0.88	0.96	0.85	0.87
39	33	4	9	0.87	0.97	0.88	0.85
40	34	4	9	0.86	0.97	0.90	0.83

DEC=Dose-expansion cohort, DLT=Dose limiting toxicity,  $p_{Rmin}$ =prespecified minimum response probability,  $p_{Rdesirable}$ =prespecified desirable response probability,  $p_{Dsafe}$ =prespecified safe DLT probability,  $p_{Dmax}$ =prespecified maximum acceptable DLT probability.

<sup>a</sup>Sample size=Total number of subjects to be tested at the MTD; this includes subjects enrolled during the dose-escalation phase and subjects enrolled in the DEC

<sup>b</sup>DEC size=Assuming 6 subjects are tested at the MTD during the dose-escalation phase, DEC size is calculated by subtracting 6 from the sample size