Current Problems in Cancer 44 (2020) 100583

Contents lists available at ScienceDirect

Current Problems in Cancer

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

Untenable dosing: A common pitfall of modern DLT-targeting Phase I designs in oncology

Paul H. Frankel[a,*], Vincent Chung[b], Yan Xing[b], Jeff Longmate[a], Susan Groshen[c], Edward M. Newman[d]

[a] Division of Biostatistics, City of Hope, Duarte, California
[b] Department of Medical Oncology, City of Hope, Duarte, California
[c] Biostatistics Core, University of Southern California/Norris Cancer Center, Los Angeles, California
[d] Division of Molecular Pharmacology, Department of Medical Oncology, City of Hope, Duarte, California

ABSTRACT

Background: There is increasing use of Phase I statistical designs to find a dose that causes rapidly emerging and particularly concerning severe or life-threatening toxicities (dose-limiting toxicities, DLTs) in a specified percent of patients most commonly 25%. While a convenient statistical framework, the foundation for selecting any specified target DLT rate, and its relevance to the recommended Phase II dose is generally lacking. *Method:* We surveyed 78 medical oncologists, most (69%) with experience as a principal investigator on a Phase I study, to ascertain their opinions related to this approach to Phase I studies and the targets often chosen. *Results:*Eighty-seven percent of respondents preferred severe toxicities in only 5%-10% of patients, consistent with 58% of respondents noting that 10% or fewer patients experience severe toxicities in the first cycle with standard outpatient treatments. The survey also documented in an example that the majority (62%) of physicians modify their patient selection during the conduct of the study based on observed toxicity and 78% note that higher toxicity is acceptable in patients where a cure is more likely. *Conclusion:* DLT-target rate designs search for a single target that is rarely well-supported in a patient population that is not stable. The most common target used is inconsistent with the toxicity of most clinically used drugs and investigator preference and can lead to the pursuit of unacceptable doses. Use of Phase I

☆ This work is supported in part by NCI grants UM1CA186717, U01CA62505, N01CM62209.

☆☆ Research reported in this publication included work performed in the Biostatistics and Mathematical Modeling Core supported by the National Cancer Institute of the National Institutes of Health under grant number P30CA033572. The content is solely the responsibility of the authors and does not necessarily represent the official views of the National Institutes of Health.

★ Stephen Shibata, MD, Deceased, assisted with the concept and design of the survey.

* Correspondence to: Paul H. Frankel, PhD, Division of Biostatistics, Department of Information Sciences, City of Hope, Duarte, CA 91010, USA.

E-mail address: pfrankel@coh.org (P.H. Frankel).

https://doi.org/10.1016/j.currproblcancer.2020.100583
0147-0272/© 2020 Elsevier Inc. All rights reserved.

trial designs with a target DLT rate should be limited to settings with a well-justified target and should specify how the target relates to the recommended Phase II dose.

© 2020 Elsevier Inc. All rights reserved.

ARTICLE INFO

Keywords: Phase I designs; Survey; Medical oncology; Dose-limiting toxicity (DLT); Target DLT rate

Introduction

Phase I cancer trials are designed to test a new drug or new combination of drugs in a patient population primarily to develop dosing guidelines for subsequent clinical trials. This is the leading edge of cancer clinical trials—a first step in the clinical evaluation of a new treatment. It is generally reserved for patients whose options are the fewest, in order to balance the risks of unexpected severe or life-threatening toxicities with yet-to-be-demonstrated offsetting therapeutic benefit. As a result, most Phase I designs in oncology limit the number of patients at risk for rapid and severe adverse events and limit the starting dose and dose increments. In this setting, the most frequently used design over the last half a century allows dose escalation if 0 of 3 or at most 1 of 6 patients experience a particularly concerning severe adverse event or a life-threatening adverse event in the first cycle of therapy (called dose-limiting toxicities [DLTs]) and requires de-escalation if 2 DLTs are encountered when 6 or fewer patients are treated. The specific definition of a DLT allows investigators to adjust to different clinical settings. This design, the 3+3 dose escalation design[1] arrives by its rules at a dose denoted as the "Maximum Tolerated Dose (MTD)" allowed per those rules. Simulations suggest that this design will select a dose on average that has a chance of DLT in the range of 10%-20% assuming the escalation proceeds until a DLT-level is reached.[1,2] This empirical 3+3 design, after extensive use over many decades, has evolved to become the tradition due to the fact it is simple, transparent, can be completed in a timely fashion and is easy to implement and provides clinical experience to help guide future doses and changes to the regimen and supportive care. It also limits risk to patients during escalation and is particularly popular for combination studies where the goal is to achieve the single agent recommended doses when used in combination and where no DLTs would usually be welcome.

More recent variations of the classical 3+3 design include allowing more than three patients to be at risk for pediatric studies testing doses already established as safe in adult trials, allowing smaller cohorts, intrapatient dose escalation or larger increments in the absence of a moderate toxicity signal, or by refining the rules to accommodate the queuing process.[3-5] In practice, these designs, when properly implemented, allow the principal investigator (PI), based on their clinical judgement, to choose a more conservative action than the design recommends (just not a more aggressive action), such as not escalating if 3 of 3 patients have grade 2 rising creatinine that does not, by rule, qualify as a DLT. Additionally, multiple protocol amendments are expected from PIs as part of adapting to new information as experience is gained with the novel treatment (eg, modifications of the schedule, changes to premedications and supportive care, changes to dose modification rules, and changes to monitoring tests). Formally, however, Monte Carlo simulations are used to evaluate the statistical properties of proposed Phase I designs. When viewed through the rigid lens of simulation, assuming random patient selection and decisions restricted to the simple binary DLT call (yes or no), the rules for these 3+3-based designs described above have readily apparent weaknesses in terms of providing a wide range of possible MTDs selected with associated wide ranges in DLT rates.[6] As a result, while staying within the DLT yes/no paradigm, a new class of Phase I designs was introduced[7] in 1990 that proposes a very different and specific goal: *To find the dose such that probability of a DLT at that dose is a pre-set value (a target DLT rate). The resulting dose is defined as the MTD.*

This represents a subtle but important change from the traditional concepts of limiting toxicity to "acceptable" levels based on tradition as we test ever higher doses. In the traditional setting, not only are the Phase I patient risks limited, but the MTD will have at most 1 of 6 patients with a DLT (16.7%); we will accept limited observed toxicity, and discuss the chance of inadvertently finding an MTD with a hidden true DLT rate that is too high, and would not get flagged without an appropriate expansion cohort. In these newer designs, however, this is replaced with the concept of a "target DLT rate" which needs to be prespecified and where the dose associated with the DLT rate is denoted the MTD. A successful dose finding would then find that the true DLT rate at the MTD is neither much lower nor much higher than the target DLT rate.

Framing dose finding based on a target DLT rate and providing a relevant technique provides a potentially useful tool. However, the use of such a tool is not well discussed nor is the importance of realistically defining the target DLT rate. The most commonly used target DLT rate is 25%,[8] which clinically means these designs aim for a dose where a quarter of the future patients should experience a DLT—a rapidly emerging (eg, within 21 or 28 days) life-threatening adverse event or a rapidly emerging and especially concerning severe adverse event. Unfortunately, there is rarely a justification for this target or any specific target chosen.

While we value the development of new methodologies to increase the design armamentarium of cancer clinical scientists, we are concerned at the lack of discussion of what constitutes an appropriate target DLT rate and especially with how this is related to the recommended Phase II dose (RP2D). We conducted a survey of oncologists to ask a few simple questions that relate to Phase I designs in general, and specifically relate to the challenge of selecting a target DLT rate for this newer class of Phase I designs if considered as a primary candidate for the RP2D.

Methods

To facilitate this dialogue, medical oncologists at the California Cancer Consortium (CCC) were surveyed using SurveyMonkey in March, 2009, with later expansion to NCI-registered researchers at the National Comprehensive Cancer Network institutions. This survey was approved by the City of Hope Institutional Review Board as an anonymized survey, and the 2 groups of physicians had similar responses and are pooled in this report.

The 5 questions included in this report are listed in Table 1 below.

Table 1
Five questions and possible responses

(a) For a new cytotoxic agent (with nonhematologic toxicity) tested in a Phase II study on solid tumors, choose your preferred toxicity profile (not knowing the effect of dose on response):
☐ Grade 1 (20%), Grade 2 (75%), Grade 3 (5%)
☐ Grade 1 (5%), Grade 2 (85%), Grade 3 (10%)
☐ Grade 1 (5%), Grade 2 (75%), Grade 3 (20%)
☐ Grade 1 (10%), Grade 2 (85%), Grade 3 (5%)
☐ Other (specify):_____
(b) For standard outpatient treatment of first-line metastatic cancer patients, in your experience, estimate the frequency of grade 3 or higher nonhematological toxicity in the first course of therapy.
☐ <5% ☐ 10% ☐ 15% ☐ 20% ☐ 25% ☐ 30% ☐ above 30% ☐ Other:_____
(c) If in a Phase I study, a novel agent had already demonstrated reversible grade 3 hematologic toxicity (non-DLT) in 3/3 patients, and the drug was just dose-escalated to the next level would this impact the profile of eligible patients you might recommend for this study?
☐ No ☐ Yes—somewhat or occasionally ☐ Yes—usually
(d) For patients with refractory metastatic disease where cure is unlikely, should chemotherapy dosing be less toxic than in patients where cure is more likely (eg, previously untreated disease, due to potential for conversion to a surgical candidate, etc)?
☐ Yes ☐ No ☐ Other (specify):_____
(e) Does finding the dose where 20% of the patients experience a DLT represent your concept of the ideal dose to take to subsequent trials to evaluate anti-tumor activity?
☐ Yes ☐ Usually ☐ Rarely ☐ No

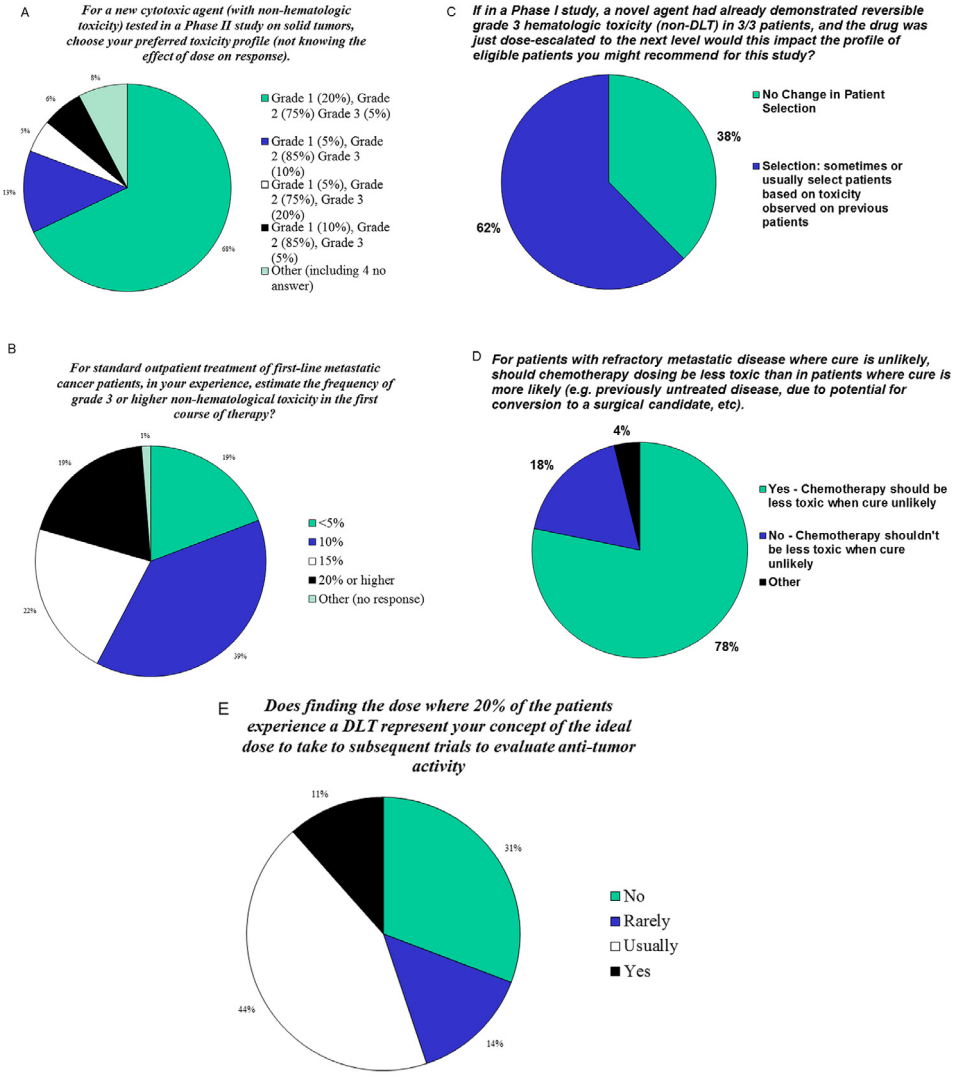

Fig. 1. Responses of 78 medical oncologists to 5 different questions. The specific question is listed above each pie chart, and the options for the respondent to choose from is provided on the right-hand side.

Results

Seventy-eight (78) medical oncologists completed the questionnaire. We report on the 5 key questions, with the full survey and results for the medical oncologists available in the supplement. The median practice experience of respondents was 15 years (range 2-45). Years of practice was not assessed on 24 respondents as this question was not included on the CCC questionnaire. Of the 78 medical oncologists, 54 had been a PI on a Phase I study.

Only 5% of respondents preferred to see a 20% incidence of grade 3 toxicities for a cytotoxic drug with nonhematologic toxicity moving into Phase II studies (Fig. 1A). More commonly (74%), investigators preferred rare grade 3 toxicity (5%), and more frequent (75%-85%) grade 2 toxicity and 87% preferred grade 3 toxicity in 10% or fewer patients. For drugs that are in use (see

Fig. 1B) we found that a majority (58%) of the physicians observe grade 3 or higher first-cycle nonhematologic toxicity in only 5%-10% of patients. Fig. 1C shows that 62% of the respondents would modify patient selection based on patterns of toxicity in previously treated patients. In Fig. 1D, 78% of the respondents thought that the toxicity should relate to the likelihood of cure (eg, converting a patient to be a surgical candidate). Finally question (e) asks in words "Is a 20% chance of DLT at the MTD reasonable?" In Fig. 1E, we see that 44% responded "Usually" and 11% "Yes," while 31% responded "No," and 14% "Rarely." Of those that selected "Usually" or "Yes" (N = 43) to that question, only 3 of 43 preferred a toxicity profile with a 20% frequency of grade 3 adverse events for a new cytotoxic drug with nonhematologic toxicity. Of the 4 of 78 total respondents that preferred a 20% frequency of grade 3 adverse events, 3 of those 4 responded "Usually" or "Yes" to the 20% DLT rate question (Fig 1E).

Discussion

The purpose of a Phase I trial is to determine the dosing of a new drug or new drug combination; in oncology this is often primarily based on toxicity and not activity. There is a dearth of dialog regarding the method or rationale for choosing a specific target DLT rate when implementing Phase I designs aiming for such a target. Oncologists, patients, and especially patients who enroll in these Phase I trials in hopes of receiving a novel therapeutic to control their disease, should expect a well-discussed and well-considered rationale for the toxicity targets as should those involved in the studies carrying the selected dose forward. To start to fill this void, we conducted a nationwide survey of the National Comprehensive Cancer Network- and CCC-affiliated oncologists using clinically familiar language to provide some insight for using toxicity as a guide for dose finding and the limitations involved.

The responses to our survey (eg, Fig 1A and 1B) align poorly with the target DLT rates used in the model-based methods where 20%-30% target DLT rates are common. This is instructive for the statistician and medical oncologist dialogue. The drift in patient selection noted in Fig. 1C has been noted in previous physician surveys[9,10] and unfortunately this selection bias calls into question the existence of a stable population necessary to rigorously define the operating characteristics of any Phase I design. Fig. 1D also suggests that one target dose (one size fits all) may not be sensible, a concept lacking from the Phase I statistical literature focused solely on toxicity.

Most Phase I medical oncologists found target DLT rates of 20%-25% too high to carry forward, yet the model-based designs both use such targets or higher and often highly value the number of patients treated at the MTD defined by such a target DLT rate.[11] In light of the survey, valuing the percent of patients treated at the MTD is not necessarily a positive feature and can (unpublished and personal communications) lead subsequent Phase II studies astray and cause unnecessary human suffering and halt interest in a new agent. Even in the traditional designs, physicians are careful to consider the possibility that the MTD may exceed the RP2D. Certainly the MTD defined and selected by a more aggressive design will usually be above the RP2D, but when the design focuses on finding the MTD based on a target DLT rate and seeks to enroll as many patients as possible at the MTD or near, researchers may be more likely to mistake the MTD for the RP2D. The RP2D selection is very complex, context specific, and depends on data far beyond the simple DLT (yes/no) consideration. However, after several failed experiences of high-dose chemotherapy in adult solid tumors and with the focus on targeted therapy, eagerness for treating at an MTD with high probabilities of DLT has been decreasing.[12] If this trend changes with chimeric antigen receptor T-cell therapy or other new therapies, or the specific clinical setting dictates a more aggressive approach, Phase I designs, and goals need to appropriately and thoughtfully adjust to the clinical setting.

We note some limitations to the survey: (1) the survey is not a formally validated instrument; (2) there was no cognitive testing separate from the multiple questions; (3) we have not demonstrated that the respondents are representative of a larger population of PIs in medical oncology; (4) more extensive and up-to-date surveys are certainly in order, especially as DLT

target-based designs have increased in popularity. New surveys of physicians may also want to consider more specific patient populations and specific therapies separately along with issues surrounding the length of the DLT window; and (5) future surveys of investigators should also be accompanied by patient surveys; as the intent is to move therapies developed in clinical trials forward into practice, the patient perspective on what level of toxicity is "acceptable" is critical. Future surveys also need to continue to evaluate the known influence of the wording of questions,[13] as the results in our survey (Fig 1A vs Fig 1E) provide a striking contrast of the responses of physicians to what appears to be the same question asked in a more clinically interpretable manner.

However, the issue is not just semantics. If the target DLT rate is supposed to determine the suggested RP2D, these designs only trade one unknown (the recommended target DLT rate) for another unknown (the recommended dose). Asking the PI to divine the recommended target DLT rate of the RP2D in a Phase I study is not a validated approach to dose finding. However, if the target DLT rate is interpreted to be the MTD better understood as the "maximum conceivable dose," the dose associated with the highest conceivably justifiable toxicity in case there is an astonishing dose–response relationship—then these targets are easier to defend even though the methods will not, in general, provide a guide for the RP2D.

The results of our survey highlight gaps between statistical models and clinical goals. Defining a desired dose in the context of Phase I studies by specifying the percent of future patients that we hope experience a severe or life-threatening adverse event in the first cycle (often 21 or 28 days) is an uncomfortable starting point. At the very least, in all Phase I trials that aim for a target DLT rate, the target DLT rate justification should be explicit in the protocol, and should be well-considered. There needs to be a justification for the 25% DLT rate commonly chosen when so few drugs are in use where 25% of the patients experience such rapid and concerning adverse reactions and often need to terminate treatment. It should also be clear and explicit if such a defined MTD is unlikely to be the RP2D.

We realize that some DLT-targeting designs are selected based on escalation/de-escalation rules that are felt to be reasonable (which often means they resemble the 3+3), rather than because a specific rate is felt to be the best choice. This may reflect the impact subtle differences in rules can have on the likelihood of a higher dose being selected. In fact, the BOIN design[14] with a target DLT rate of 25% de-escalates with 1 DLT in 1, 2, or 3 patients, while the 3+3 does not, making the BOIN more conservative at that decision point. In many settings, this specific decision could be supported as an improvement over the 3+3 in light of modern therapies and the survey results. However, the BOIN design with a 25% DLT rate target also escalates with 2 DLTs out of 11 (18.2%) or 3 DLTs of 16 (18.8%) and does not de-escalate with 5 DLTs in 17 patients (29.4%), which is more aggressive. Ultimately, if a DLT-targeting rate design is employed when the stated statistical target is not the clinical target, this represents a disconnect between the statistician and the PI, and is a signal for improved communication.

Such detailed communication is also critical when considering the extent of the DLT window, and future surveys, as noted above, should better explore this topic. Specific therapies may require different windows, and this can complicate the design selection and determination of acceptable DLT rates. Even beyond the *a priori* DLT window, late or cumulative toxicities may indicate that the Recommended Phase 2 Dose should be lower than the first cycle determined MTD. This is even the case with "classical" chemotherapy agents, where cumulative toxicities may mean that the "MTD" is not in fact tolerable.[15] Likewise, late-developing radiation side-effects are difficult to account for in a "reasonable" DLT window and the expanding use of checkpoint inhibitors and other immune modulators present additional challenges as immune-related adverse events are often after the first cycle. These and other considerations prevent a one-size-fits all answer to the best approach for dose-finding, but increased communication can only improve our clinical trial designs and better protect patients in the process.

Our survey, even with its limitations, provides the largest and most recent survey of Phase I medical oncologists and provides an understanding that may help frame the discussion surrounding the appropriate Phase I design and agreed upon goals for a given situation.

Declaration of competing interest

All authors have confirmed that there are no conflicts of interest associated with this work.

Supplementary materials

Supplementary material associated with this article can be found, in the online version, at doi:10.1016/j.currproblcancer.2020.100583.

CRediT authorship contribution statement

Paul H. Frankel: Conceptualization, Methodology, Investigation, Writing - original draft, Visualization, Supervision. **Vincent Chung:** Conceptualization, Methodology, Writing - review & editing. **Yan Xing:** Writing - review & editing. **Jeff Longmate:** Conceptualization, Methodology, Writing - review & editing. **Susan Groshen:** Conceptualization, Methodology, Writing - review & editing. **Edward M. Newman:** Conceptualization, Methodology, Writing - review & editing, Investigation, Supervision, Funding acquisition.

References

1. Storer BE. Design and analysis of phase I clinical trials. *Biometrics*. 1989;45:925–937.
2. Sposto R, Groshen S. A wide-spectrum paired comparison of the properties of the rolling 6 and 3+3 phase I study designs. *Contemp Clin Trials*. 2011;32:694–703.
3. Skolnik JM, Barrette JS, Jayaraman B, et al. Shortening the timeline of pediatric phase I trials: the rolling six design. *J Clin Oncol*. 2008;26:190–195.
4. Simon R, Friedlin B, Rubinstein L, et al. Accelerated titration designs for phase I clinical trials in oncology. *J Natl Cancer Inst*. 1997;89:1138–1147.
5. Frankel P, Chung V, Tuscano J, et al. Model of a queuing approach for patient accrual in phase 1 oncology studies. *JAMA Network Open*. 2020;3(5) e204787. doi:10.1001/jamanetworkopen.2020.4787.
6. Ahn C. An evaluation of phase I cancer clinical trial designs. *Stat Med*. 1998;17:1537–1549.
7. O'Quigley J, Pepe M, Fisher L. Continual reassessment method: a practical design for phase 1 clinical trials in cancer. *Biometrics*. 1990;46:33–48.
8. Iasonos A, O'Quigley J. Adaptive dose-finding studies: a review of model-guided phase I clinical trials. *J Clin Oncol*. 2014;32:2505–2511.
9. Mick R, Lane N, Daugherty C, et al. Physician-determined patient risk of toxic effects: impact on enrollment and decision making in phase I cancer trials. *J Natl Cancer Inst*. 1994;86:1685–1693.
10. Daugherty C, Ratain MJ, Grochowski E, et al. Perceptions of cancer patients and their physicians involved in phase I trials. *J Clin Oncol*. 1995;13:1062–1072.
11. Zhou H, Ying Y, Nie L. Accuracy, safety and reliability of novel phase I trial designs. *Clin Cancer Res*. 2018;24:4357–4364.
12. Ratain M. Shrinking oncology phase I clinical trials: If the drug is a breakthrough, why would you ever want to utilize the maximally tolerated dose? *Cancer Letter*. 2018;44:14–15.
13. Chao C, Studts JL, Abell T, et al. Adjuvant chemotherapy for breast cancer: how presentation of recurrence risk influences decision-making. *J Clin Oncol*. 2003;21:4299–4305.
14. Yuan Y, Hess KR, Hilsenbeck SG, Gilbert MR. Bayesian optimal interval design: a simple and well-performing design for phase I oncology trials. *Clin Cancer Res*. 2016;22:4291–4301.
15. Muggia FM, Synold TW, Newman EM, et al. Failure of pretreatment with intravenous folic acid to alter the cumulative hematologic toxicity of lometrexol [letter]. *J Natl Cancer Inst*. 1996;88:1495–1496.