Clinical Cancer Research

Dose-Finding Methods: Moving Away from the 3 + 3 to Include Richer Outcomes

Xavier Paoletti[1,2], Damien Drubay[1,2], and Laurence Collette[3]

The most commonly used method for dose finding, the 3 + 3, has poor performance. New adaptive designs are more efficient. Nevertheless, they have reached a maximum performance level, and further improvement requires either larger sample sizes or outcomes measures richer than the simplistic severe toxicity measured at cycle 1. *Clin Cancer Res; 23(15); 3977–9.* ©2017 AACR.

See related article by Yan et al., p. 3994

In this issue of *Clinical Cancer Research*, Yan and colleagues (1) propose a new adaptive dose-finding method, termed keyboard, as an alternative to the 3 + 3 design. The performance of the most commonly used dose-finding method, the 3 + 3, is disappointing: The chance of finding the correct dose in a phase I trial is no more than 40% (1). How many active, potentially promising agents are dropped due to incorrect dose selection during development? Few researchers address this question (2), but stakeholders acknowledge the limitations of early-phase trials (3). The statistical community unanimously agrees that the 3 + 3 is not an efficient method, raising ethical concerns according to NIH guidelines (4), and alternatives should be implemented.

Keyboard combines simplicity, performance, and flexibility. With regard to simplicity, there are two reasons: (i) It is based on a natural definition of the MTD, that is, the dose at which the risk of dose-limiting toxicity (DLT) is within a predefined range (typically 20%–35%), and (ii) the decision rules to (de) escalate are driven by the accumulated observations and the observed proportion of DLT at a given dose level. The more patients enrolled at a specific dose, the more confident we can be that this dose is (close to) the MTD, and the more likely the same dose is recommended for the next patients.

However, simplicity is not an objective *per se*, and the key point is performance. The keyboard design is rooted in good and fruitful statistical concepts and properties. It belongs to the semiparametric continual reassessment class of methods (semiparametric CRM; ref. 5), which has been shown to provide good operating characteristics and valid asymptotic properties while being quite flexible. In Yan and colleagues' simulation study (1), the keyboard design outperformed the 3 + 3 but was slightly inferior to the CRM (6). Importantly, these performances were quite close to the maximum achievable performances given the sample size and the scenario. Figure 1 shows the distribution of the recommendations obtained with the optimal benchmark. This benchmark uses so-called complete information, which is estimable only in a simulation framework, as if each patient could be treated at all the doses independently (7). This optimal method serves as a reference.

An important advantage of keyboard over the 3 + 3 design is its flexibility. As with other model-based methods, all collected data contribute to the MTD assessment, not only data from the last six patients. This is crucial for monitoring patient's tolerance throughout the trial, including the dose expansion cohorts. These large cohorts of patients with selected tumor types are aimed to explore preliminary signs of activity. However, they too must be monitored for toxicity, as shown repeatedly by Iasonos and colleagues (8). In cases of excessive toxicity rate, the investigated dose should be modified during the expansion cohort. Following the recommendations of the DLT-TARGETT (9), the expansion cohort should serve to refine the toxicity assessment. The possibility given by keyboard to update the estimated risk of DLT and its confidence interval, based on all accumulated data, provides an additional tool to fulfil these requirements. The analysis combining data from the dose escalation and the dose expansion cohorts could increase the chance of selecting the correct dose (8).

To implement keyboard into practice, we need easy-to-use software with two purposes. First, given previous observations, it should indicate the next dose level to allocate [see Table 2 and Fig. 2 in the article by Yan and colleagues (1)]. Second, and perhaps even more importantly, it should be able to run simulations similar to those presented in Figs. 3 to 5 in the article by Yan and colleagues. Indeed, protocols that use adaptive designs require more preparatory work. There is no equivalent to the "null hypothesis, clinical targeted difference and type I error" used to design phase II or III trials. Each team preparing a phase I trial must simulate possible trials under various scenarios specific to the trial context. This preparatory step is performed by the statistician in collaboration with the investigator to calibrate and tailor the design, and thereby to obtain the best possible operating characteristics. The scenarios of the simulations are designed to reflect the expected relationship between the dose and the risk of DLT for the agent under investigation. For example, the scenario mimicking a trial of a monoclonal antibody will be different from that of a kinase inhibitor, as the former probably has a much

[1]Biostatistics and Epidemiology Department, Gustave Roussy Cancer Center, Villejuif, France. [2]Université Paris 11 Saclay and INSERM U1018 CESP OncoStat, Villejuif, France. [3]European Organization for Research and Treatment of Cancer Headquarters, Brussels, Belgium.

Corresponding Author: Xavier Paoletti, Gustave Roussy, 114 rue Ed. Vaillant, Villejuif 94805 Cedex, France. Phone: 33 142 11 6564; E-mail: xavier.paoletti@gustaveroussy.fr

doi: 10.1158/1078-0432.CCR-17-1306

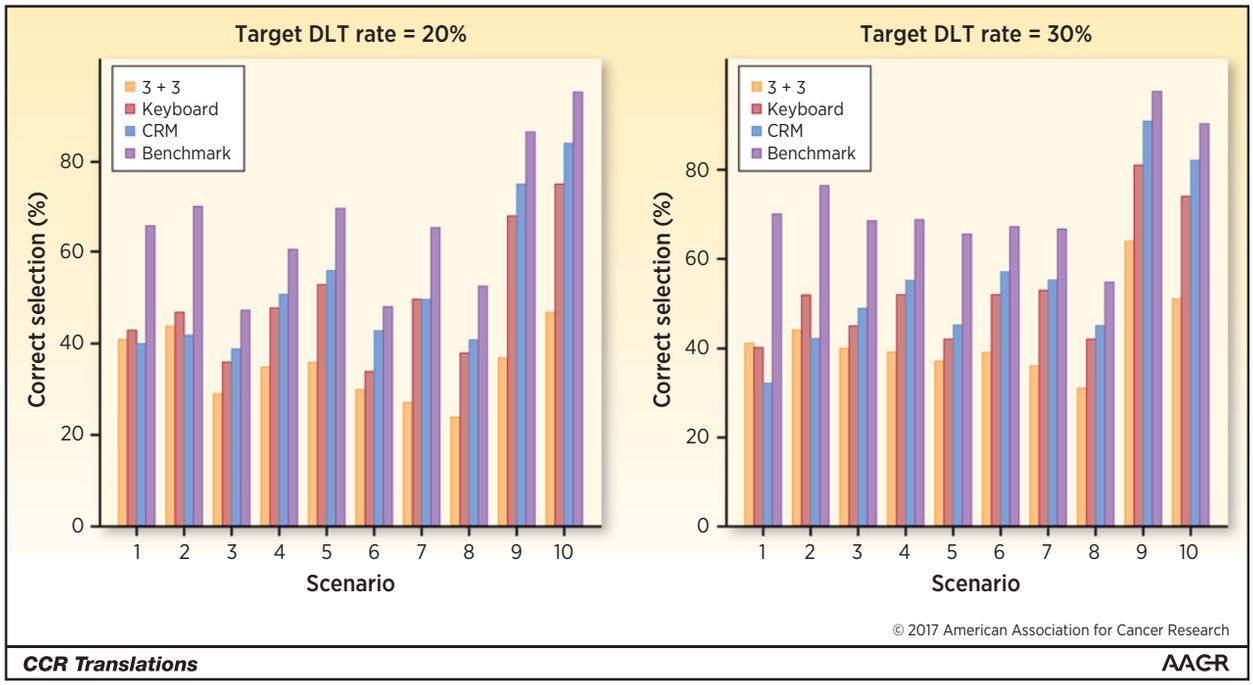

Figure 1.
Percentage of correct selection with the 3 + 3, the CRM, and the optimal benchmark in the 10 scenarios presented by Yan and colleagues (1), with a target rate of either 20% (left) or 30% (right).

larger therapeutic index than the latter. Similarly, to prepare a pediatric trial, we would choose scenarios based on the data collected in adults, as agents are usually tested in adults before being investigated in children (10). The examples from Fig. 4 of Yan and colleagues' work (1) highlight that there is no single method that is the most effective across all scenarios. This process is lengthy but may reconcile the investigators with statistically oriented dose-finding designs.

This combination of relative simplicity, good operating characteristics, and flexibility to tackle various practical situations should motivate investigators and sponsors to move away from the 3 + 3 and to adopt alternative designs.

Will conducting future phase I trials with the keyboard method guarantee the systematic selection of the correct dose? Unfortunately not. As shown in Fig. 1, the chance of selecting the correct dose with the benchmark, (i.e., the highest achievable performance) is below 60% in most of the explored scenarios. Despite the good statistical properties of keyboard, its performances are limited by the type of primary endpoint used in phase I trials: The DLT variable is binary and has irreducible binomial variability for a given (often limited) sample size. Continuous, ordinal, or multiple endpoints would be much more informative.

One may then question the choice of the DLT as the sole endpoint. The DLT-TARGETT database consists of individual patient data from 54 phase I trials of single-targeted agents (9). Of these, a total of 25% completely missed the primary objective, as they failed to identify the MTD and recommend a dose for phase II due to lack of DLT. In the 2,084 treated patients, more than 24,000 graded toxicities, as measured by the NCI common toxicity criteria scale, were recorded as possibly related to the treatment. These toxicities occurred at any of the first six cycles of treatment. Yet, only 164 DLTs were reported, reflecting the huge shrinkage in the available information (99.4% of all the events were discarded to define the MTD). A large fraction (50%) of the first grade 3 or 4 toxicity occurred after the DLT assessment period, and dose intensity was reduced in more than 20% of all treatment cycles, raising the question of tolerability of the treatments administered in the long run. To define the recommended dose, we need richer outcomes, possibly reflecting both toxicity and activity. Continuous endpoints have a better sensitivity to agent effect and a better discriminatory value to rank dose activity.

There is an urgent need for innovative designs that are both efficient and simple and that could reduce the high failure rate of dose finding (2). This "well-tempered" keyboard can help synthesize complex information and may stimulate statisticians and principal investigators to work together to provide a good design for a particular agent. Nevertheless, more efficient use of the multiple sources of data that are collected in most phase I trials is needed. Currently pharmacokinetics, pharmacodynamic biomarkers, late-onset toxicity, and complex imaging are not formally integrated in the dose-finding analysis (11); this is a waste of resources.

Disclosure of Potential Conflicts of Interest

No potential conflicts of interest were disclosed.

Authors' Contributions

Conception and design: X. Paoletti, L. Collette
Development of methodology: X. Paoletti, L. Collette

Acquisition of data (provided animals, acquired and managed patients, provided facilities, etc.): X. Paoletti
Analysis and interpretation of data (e.g., statistical analysis, biostatistics, computational analysis): X. Paoletti, D. Drubay, L. Collette
Writing, review, and/or revision of the manuscript: X. Paoletti, D. Drubay, L. Collette
Administrative, technical, or material support (i.e., reporting or organizing data, constructing databases): X. Paoletti, D. Drubay
Study supervision: X. Paoletti

Grant Support

X. Paoletti was partly supported by the ITN-Marie Curie IDEAS project, X. Paoletti and D. Drubay were partly supported by the French NCI (INCA-Optidose program), and L. Collette was partly supported by FOCA (Fonds cancer) from Belgium.

Received May 21, 2017; accepted June 5, 2017; published OnlineFirst June 19, 2017.

References

1. Yan F, Mandrekar SJ, Yuan Y. Keyboard: a novel Bayesian toxicity probability interval design for phase I clinical trials. Clin Cancer Res 2017;23:3994–4003.
2. Kola I, Landis J. Can the pharmaceutical industry reduce attrition rates? Nat Rev Drug Discov 2004;3:711–5.
3. Wong KM, Capasso A, Eckhardt SG. The changing landscape of phase I trials in oncology. Nat Rev Clin Oncol 2016;13:106–17.
4. Emanuel EJ, Wendker D, Grady C. What makes clinical research ethical? JAMA 2000;283:2701–11.
5. Clertant M, O'Quigley J. Semi-parametric dose finding methods. J R Stat Soc Series B Stat Methodol 2017. doi:10.1111/rssb.12229.
6. 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.
7. Paoletti X, O'Quigley J, Maccario J. Design efficiency in dose finding studies. Comput Stat Data Anal 2004;45:197–214.
8. Iasonos A, O'Quigley J. Design considerations for dose-expansion cohorts in phase I trials. J Clin Oncol 2013;31:4014–21.
9. Postel-Vinay S, Collette L, Paoletti X, Rizzo E, Massard C, Olmos D, et al. 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. Eur J Cancer 2014;50:2040–9.
10. Paoletti X, Geoerger B, Doz F, Baruchel A, Lokiec F, Le Tourneau C. A comparative analysis of paediatric dose-finding trials of molecularly targeted agent with adults' trials. Eur J Cancer 2013;49:2392–402.
11. Postel-Vinay S, Soria JC. Phase I trials in oncology: a new era has started. Ann Oncol 2015;26:7–9.