INNOVATIONS JOURNALS

Research Article

MC-Keyboard: A Practical Phase I Trial Design for Targeted Therapies and Immunotherapies Integrating Multiple-Grade Toxicities

Liyun Jiang,[1] Zhulin Yin,[2] Fangrong Yan,[1] Ying Yuan[3]

[1]Research Center of Biostatistics and Computational Pharmacy, China Pharmaceutical University, Nanjing, China
[2]Clinical Trials and Statistics Unit, Institute of Cancer Research, London, UK
[3]Department of Biostatistics, The University of Texas MD Anderson Cancer Center, Houston, TX, USA

Address correspondence to Ying Yuan (yyuan@mdanderson.org).

Sources of Support: The research of Drs Jiang and Yan was supported in part by National Natural Science Foundation of China (funding nos. 82304252 and 82273735).

Conflicts of Interest: None.

Submitted: Jan 25, 2024; First Revision Received: Jan 25, 2024; Accepted: Jan 25, 2024

Jiang L, Yin Z, Yan F, Yuan Y. MC-Keyboard: a practical phase I trial design for targeted therapies and immunotherapies integrating multiple-grade toxicities. *J Immunother Precis Oncol*. 2024; 7:159–167. DOI: 10.36401/JIPO-23-35.

ABSTRACT

Introduction: In targeted therapies and immunotherapies, the occurrence of low-grade (e.g., grade 1–2) toxicities (LGT) is common, while dose-limiting toxicities (DLT) are relatively rare. As a result, conventional phase I trial designs, solely based on DLTs and disregarding milder toxicities, are problematic when evaluating these novel therapies. **Methods:** To address this issue, we propose a novel phase I design called a multiple-constraint keyboard (MC-Keyboard) that integrates multiple toxicity constraints, accounting for both DLT and LGT, for precise dose escalation and de-escalation, and identification of the maximum tolerated dose (MTD). As a model-assisted design, an important feature of MC-Keyboard is that its dose-escalation or de-escalation rule can be pretabulated and incorporated into the trial protocol before the initiation of the trial, greatly simplifying its implementation. **Results:** The simulation study showed that the MC-Keyboard had high accuracy in identifying the MTD and is safer than some existing designs. **Conclusion:** The MC-Keyboard provides a novel, simple, and safe approach to assessing safety and identifying the MTD for targeted therapies and immunotherapies.

Keywords: immunotherapy, targeted therapies, dose finding, model-assisted designs, toxicity grade, multiple-toxicity constraints

INTRODUCTION

Traditionally, phase I trial designs have focused on severe toxicities referred to as dose-limiting toxicities (DLT) and perform dose escalation solely based on DLT occurrences. Treatment-related toxicities are classified into five grades by the National Cancer Institute Common Terminology Criteria for Adverse Events: grade 0, indicating no toxicity; grades 1 and 2, indicating mild to moderate toxicity; and grades 3 and higher, indicating severe toxicity.[1] DLT is typically defined as grade 3 or higher toxicity.[2] The primary objective of conventional phase I trial designs was to establish the maximum tolerated dose (MTD), which was defined as the highest dose that yields a DLT rate equal to or closest to a predefined target, such as 25% or 30%.[3]

Numerous phase I designs based on DLT have been proposed, including algorithm-based, model-based, and model-assisted designs. An example of an algorithm-based design is the widely used the 3 + 3 design,[4] which, while easy to understand and implement, often exhibits poor performance in identifying the MTD accurately. Model-based designs, such as the continual reassessment method (CRM),[5] rely on statistical models for the dose-toxicity relationship and continually update these models based on accumulated data to guide dose escalation and de-escalation. Although model-based designs generally outperform the 3 + 3 design, their implementation is hindered by complex statistical models and estimation procedures.[6]

The model-assisted design was introduced to combine the simplicity of the 3 + 3 design with the superior performance of model-based approaches.[6,7] An eminent example of this category is the Bayesian optimal interval (BOIN) design,[8] which derives its decision rules through rigorous statistical optimization theory, akin to model-based designs. At the same time, the resulting dose-escalation and de-escalation rules can be pretabulated and incorporated into the trial protocol, mirroring the straightforward

implementation of algorithm-based designs. The BOIN design has garnered increasing attention for its ease of implementation, high efficiency, and flexibility, leading the United States Food and Drug Administration (FDA) to designate it as a fit-for-purpose tool for drug development.[9] Additionally, the Keyboard design represents another model-assisted approach that has found practical use.[10] However, most existing designs, including those mentioned above, primarily focus on DLT, overlooking the significance of low-grade (e.g., grade 1–2) toxicities (LGT).

The emergence of molecularly targeted therapies and immunotherapies has posed a challenge to the conventional dose-finding paradigm. These innovative treatments exhibit distinct safety profiles compared with traditional chemotherapies. Often, toxicities induced by targeted therapies and immunotherapies manifest as LGT, with rare instances of DLT. For instance, treatment-related adverse events of grades 3 and higher were observed in less than 10% of patients treated with pembrolizumab.[11] Consequently, applying the conventional DLT-based designs to phase I trials of targeted therapies and immunotherapies often fails to establish the MTD.

Furthermore, traditional designs neglect to account for the impact of LGT, which, despite being less severe, hold significant clinical relevance and bear substantial implications for treatment. The high frequency of LGTs can greatly diminish the patient's quality of life and lead to tolerance issues, such as dose reduction, interruption, or termination, ultimately affecting the treatment's effectiveness. An example is ceritinib, which initially received approval by the FDA at a daily dose of 750 mg. However, an excessive number of low-grade gastrointestinal toxicities were observed, leading to dose modifications occurring in 38% of patients and approximately 60% of patients requiring at least one dose reduction.[12] As a result, the FDA modified the approved dose to 450 mg.[13] This is one of the key issues that motivated the FDA to launch Project Optimus to reform the dose selection paradigm.

To address these limitations, we propose a novel phase I design called the multiple-constraint keyboard (MC-Keyboard), which integrates multiple toxicity constraints, accounting for both DLT and LGT, for precise dose escalation and de-escalation. Considering both toxicity types, the proposed approach aimed to identify the MTD more accurately, providing valuable insights for optimal treatment strategies. The MC-Keyboard is an extension of the Keyboard design. As a model-assisted design, the MC-Keyboard offers a crucial advantage in that its dose-escalation and de-escalation rules can be tabulated and included in the trial protocol even before commencing the trial. This feature greatly simplifies the implementation of the MC-Keyboard, making it a practical and efficient approach for dose finding.

Several designs have been proposed to integrate LGT into the dose-finding process. Bekele and Thall[14] introduced the concept of severity weights and defined the MTD as the dose associated with a prespecified expected total toxicity burden, which aggregates severity weights for different toxicities. Yuan et al[15] developed the quasi-CRM design, wherein toxicity grades are converted into the number of DLT events and integrated into the CRM using the quasi-Bernoulli likelihood. Lee et al[16,17] summarized the toxicity profiles from cancer trials by employing a toxicity burden score, determined through a weighted sum, and established DLT criteria based on the toxicity burden score. Furthermore, they proposed an extension of CRM known as the multiple-toxicity constraints CRM (MC-CRM), which accommodates multiple-toxicity thresholds.[18] Mu et al[19] proposed the generalized BOIN, an extension of BOIN that handles multiple-toxicity grades and incorporates all existing toxicity burden scoring systems within a unified framework.

METHODS

A Review of Keyboard Design

We first briefly reviewed the Keyboard design,[10] which was the foundation of the MC-Keyboard. Let ϕ_{DLT} denote the prespecified target DLT rate, and δ_{DLT} denote the indifference margins. Keyboard design starts by specifying a target dosing interval $(\phi_{DLT} - \delta_{DLT}, \phi_{DLT} + \delta_{DLT})$, referred to as the "target key." Any dose with a DLT rate in that range is considered as the MTD. For example, when $\phi_{DLT} = 0.25$ and $\delta_{DLT} = 0.05$, the target key or interval is (0.2, 0.3), implying a dose with a DLT rate within this range could be deemed the MTD. Afterward, the design constructs a series of keys, equally spaced on both sides of the target key. These keys serve as potential intervals where the true DLT rate of the current dose is likely to lie. With the observed data from the current dose level j, the "strongest key" is identified, which corresponds to the key with the largest posterior probability. The strongest key offers crucial information regarding the probable location of the true DLT rate for the current dose level.

Based on the target key and the strongest key, the decision of dose escalation and de-escalation was made as follows:

- If the strongest key is on the left side of the target key, escalate the dose level to $j + 1$;
- If the strongest key is on the right side of the target key, de-escalate the dose level to $j - 1$;
- Otherwise, stay at the current dose, j.

Simulation studies have demonstrated that the Keyboard design performs on par with the model-based CRM design and the model-assisted BOIN design while outperforming the modified toxicity probability interval design.[20] However, one of the limitations of the Keyboard design is that it makes dose escalation and de-escalation decisions solely based on DLT, disregarding the importance of LGT in early-phase clinical trials. To address this limitation, we introduced the MC-Keyboard design to incorporate the relevance of both DLT and LGT.

MC-Keyboard Design

The MC-Keyboard determines dose escalation or de-escalation based on both DLT and LGT. Let ϕ_{LGT} denote the prespecified target for the LGT rate, representing the highest acceptable LGT rate, and δ_{LGT} denote the corresponding indifference margin. Suppose that the current dose level is j, to determine the dose for the next cohort of patients, the MC-Keyboard forms two sets of keys independently based on $(\phi_{DLT}, \delta_{DLT})$ and $(\phi_{LGT}, \delta_{LGT})$, respectively. The indifference margins δ_{DLT} and δ_{LGT} determine the width of the key. They should be chosen in a way that any deviation from the target ϕ_{DLT}, no more than δ_{DLT}, and any deviation from the target ϕ_{LGT}, no more than δ_{LGT}, are deemed clinically not different from the targets. The values of δ_{DLT} and δ_{LGT} are typically selected within the range of 0.05 to 0.1 and may differ from each other. The margin within the range of 0.05 to 0.1 might seem large, but it is actually reasonable considering small sample size (e.g., 3–9 patients at each dose) and thus the very limited power to distinguish a difference within the range of 0.05 to 0.1. In application, the values of ϕ_{DLT}, δ_{DLT}, ϕ_{LGT}, and δ_{LGT} should be chosen carefully based on the treatment's characteristics, clinicians' prior knowledge of risks and benefits of the treatment, and patients' inputs. They should also be calibrated by simulation to ensure desirable operating characteristics. These values may vary from trial to trial.

We apply the same posterior probability calculation as the Keyboard design to DLT data and LGT data independently to determine the strongest key for each toxicity type. The strongest key for LGT indicates the most likely location of the true LGT rate, while the strongest key for DLT presents where the true DLT rate is most likely located.

Given the identified strongest keys for LGT and DLT, the dose-escalation and de-escalation rule of MC-Keyboard is as follows and illustrated in Figure 1.

- If the strongest key of LGT is on the left side of its target key AND the strongest key of DLT is on the left side of its target key, escalate the dose level to $j + 1$;
- If the strongest key of LGT is on the right side of its target key OR the strongest key of DLT is on the right side of its target key, de-escalate the dose level to $j - 1$;
- Otherwise, stay at the current dose j.

The trial proceeds with dose escalation or de-escalation using the above rule until a predefined stopping criterion is met (e.g., stopping when a predetermined maximum sample size is achieved). Subsequently, based on the accumulated DLT and LGT data across all doses, the MC-Keyboard identifies the MTD as follows. The design first identifies the dose, denoted as d_{DLT}, whose estimate of the DLT rate is closest to the prespecified target DLT rate ϕ_{DLT}. Similarly, it identifies another dose, denoted as d_{LGT}, whose estimate of the LGT rate is closest to the prespecified target LGT rate ϕ_{LGT}. The MTD is then determined as the minimum of d_{DLT} and d_{LGT}, ensuring that the MTD accounts for both DLT and LGT considerations. To

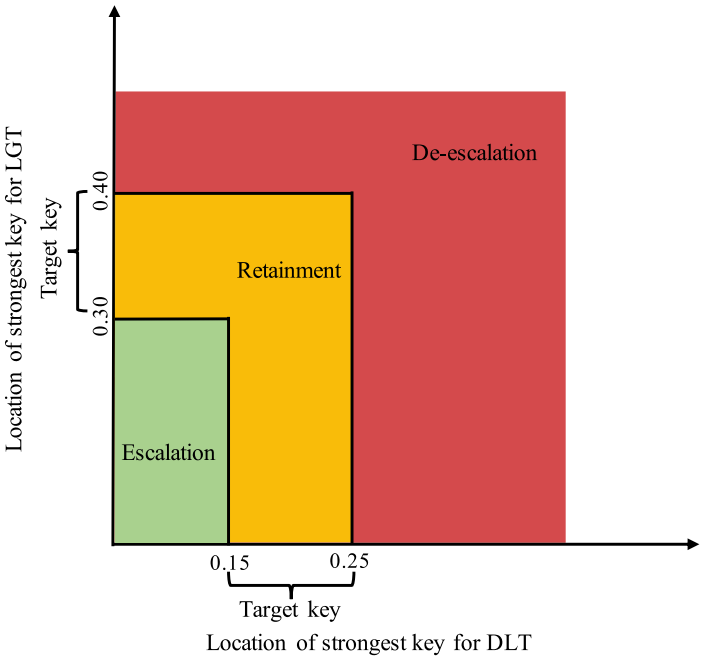

Figure 1. Dose escalation, retainment, and de-escalation region of the MC-Keyboard. MC-Keyboard: multiple-constraint keyboard.

differentiate this MTD, which considers both DLT and LGT, from the traditional definition of MTD based solely on DLT, we will henceforth refer to the latter as the DLT-based MTD. The MC-Keyboard design employs a statistical method known as isotonic regression to obtain the estimates of the DLT rate and LGT rate for each dose. The details of this method can be found in the works of Liu and Yuan[8] and Yan et al.[10] The technical details of the MC-Keyboard are provided in the Supplementary Materials (available online). Here, we select the MTD as the dose that has the estimate closest to the target LDT and LGT rates, which could be outside of the target intervals, i.e., $(\phi_{DLT} - \delta_{DLT}, \phi_{DLT} + \delta_{DLT})$ and $(\phi_{LGT} - \delta_{LGT}, \phi_{LGT} + \delta_{LGT})$. This selection rule aligns with the majority of dose-finding methods. Nevertheless, if desired, it is straightforward to impose the additional condition that the estimate of the DLT rate and/or LGT rate for the MTD must be located within the target intervals. However, it is important to note that this additional condition may be excessively restrictive, given the small sample size. There is a substantial likelihood that the estimates of DLT and LGT rates are outside the target intervals, even when the true DLT and LGT rates are within the target intervals and vice versa.

As a model-assisted design, the MC-Keyboard offers a crucial advantage in that its dose-escalation and de-escalation rule described above can be tabulated and included in the trial protocol before commencing the trial. This feature greatly simplifies the implementation of the MC-Keyboard, making it a practical and efficient approach for dose finding. Table 1 presents the MC-Keyboard decision rule for a cohort size of 3, with a target LGT rate $\phi_{LGT} = 0.35$ and a target DLT rate $\phi_{DLT} = 0.2$. For example, suppose that among three patients treated

Table 1. Dose-escalation and de-escalation decision rule table for the MC-Keyboard with a cohort size of 3, a target DLT rate of 0.2, and a target LGT rate of 0.35

No. of Patients Treated at Current Dose Level	DLTs (*n*)	LGTs (*n*)	Decision		
			Escalate	Stay	De-escalate
3	0	0	Y		
3	0	1		Y	
3	0	2			Y
3	0	3			Y&Elim
3	1	0–1			Y
3	1	2			Y&Elim
3	2–3	0–1			Y&Elim
6	0	0–1	Y		
6	0	2		Y	
6	0	3–4			Y
6	0	5–6			Y&Elim
6	1	0–2		Y	
6	1	3–4			Y
6	1	5			Y&Elim
6	2	0–3			Y
6	2	4			Y&Elim
6	3–6	0–3			Y&Elim
9	0	0–2	Y		
9	0	3		Y	
9	0	4–5			Y
9	0	6–9			Y&Elim
9	1	0–2	Y		
9	1	3		Y	
9	1	4–5			Y
9	1	6–8			Y&Elim
9	2	0–3		Y	
9	2	4–5			Y
9	2	6–7			Y&Elim
9	3	0–5			Y
9	3	6			Y&Elim
9	4–9	0–5			Y&Elim
12	0	0–3	Y		
12	0	4		Y	
12	0	5–6			Y
12	0	7–12			Y&Elim
12	1	0–3	Y		
12	1	4		Y	
12	1	5–6			Y
12	1	7–11			Y&Elim
12	2	0–4		Y	
12	2	5–6			Y
12	2	7–10			Y&Elim
12	3	0–6			Y
12	3	7–9			Y&Elim
12	4	0–6			Y
12	4	7–8			Y&Elim
12	5–12	0–7			Y&Elim
15	0	0–4	Y		
15	0	5		Y	
15	0	6∼8			Y
15	0	9∼15			Y&Elim
15	1	0∼4	Y		
15	1	5		Y	
15	1	6∼8			Y
15	1	9∼14			Y&Elim
15	2	0∼4	Y		
15	2	5		Y	
15	2	6∼8			Y
15	2	9∼13			Y&Elim
15	3	0∼5		Y	
15	3	6∼8			Y
15	3	9∼12			Y&Elim
15	4	0∼8			Y
15	4	9∼11			Y&Elim
15	5	0∼8			Y
15	5	9∼10			Y&Elim
15	6–15	0∼9			Y&Elim

Table 1 continues on next column

Table 1. Continued

Note: If a patient experiences multiple different-grade toxicities (e.g., both DLT and LGT), only the highest grade (e.g., DLT) is counted for that patient to guide dose escalation (i.e., the patient is only counted once).
MC-Keyboard: multiple-constraint keyboard; DLTs: dose-limiting toxicities; LGTs: low-grade toxicities; Y: yes; Y&Elim: de-escalating the current dose and eliminating this dose and higher doses.

at the current dose, none experienced DLT and LGT (without DLT); by looking up the table, the dose should be escalated. Consider another case where nine patients have been treated at the current dose, one patient experienced DLT, and four experienced LGT without DLT; then the dose should be de-escalated. Of note, if a patient experiences multiple different-grade toxicities, only the highest grade is counted for that patient to guide dose escalation. For example, if a patient experienced grade 2 hypotension and grade 4 nausea, that patient is counted as a DLT, not LGT. In addition, it is important to mention that Table 1 assumes a cohort size of 3; however, the MC-Keyboard is flexible and can be applied for any prespecified cohort size.

The MC-Keyboard's simplicity stands in contrast to competitive methods like the model-based MC-CRM, which also considers multiple toxicity constraints but requires complex real-time model fitting and estimation for dose assignment in each cohort. Simulation studies demonstrate that the MC-Keyboard offers accuracy similar to the MC-CRM in identifying the MTD but is safer.

For patient safety, the MC-Keyboard includes a dose elimination and early stopping rule as follows. When the observed data at the current dose level indicates a high posterior probability (e.g., 95% chance) that the dose is above the MTD, the current dose and higher doses are eliminated from the trial. Subsequently, the next cohort of patients is treated at the next lower dose. Note that at least three patients must be treated before a dose can be eliminated. If the lowest dose is eliminated, the trial is terminated early, and no MTD is selected. This dose-elimination rule is applied after each cohort and is displayed in Table 1 under the column titled "De-escalate," represented as "Y&Elimi," signifying the de-escalation of the current dose and elimination of it and higher doses. For instance, when six patients have been treated at the current dose, among them, three patients experienced DLT, and three patients had LGT without DLT. The MC-

Keyboard would then eliminate the current and higher doses, and the next cohort would be treated at the next lower dose, ensuring patient safety. The MC-Keyboard design is summarized in Box 1.

Box 1. MC-Keyboard Design

1. Start the trial by treating the first cohort of patients at the lowest dose;
2. Count the number of patients treated, the number of patients who experienced DLT, and the number of patients who experienced LGT (without DLT) at the current dose. Then, select the dose for the next cohort of patients based on the MC-Keyboard decision rule table (e.g., Table 1);
3. Repeat step 2 until the maximum sample size is reached or the trial is terminated early for safety;
4. Select the MTD based on both DLT and LGT data using isotonic regression as described in the main text.

Trial Example

We present a hypothetical example to illustrate using the MC-Keyboard in practice. Consider a phase I trial that aims to find the MTD with a target LGT rate of 0.35, a target DLT rate of 0.2, and an indifference margin of 0.05 for both LGT and DLT. The target rates were chosen based on the input of clinicians involved in the trial to reflect the clinical consideration that the desirable DLT rate and LGT rate should not exceed 0.25 (i.e., 0.2 + 0.05) and 0.4 (i.e., 0.35 + 0.05), respectively. The total sample size is 30, and patients are treated in cohorts of three. There are five doses. Figure 2 shows the dose-escalation or de-escalation process by applying the MC-Keyboard design. The trial started by assigning the first cohort of patients to the lowest dose, dose level 1. No DLTs and LGTs were observed among the three patients. Based on the decision rule presented in Table 1, the next three patients were treated at dose level 2. As none experienced DLT and LGT at this dose level, the dose was escalated to dose level 3 for the next cohort. Among the three patients treated at dose level 3, no DLTs and LGTs were observed, and the next cohort was treated at dose level 4. Because one DLT and one LGT were observed at dose level 4, the MC-Keyboard de-escalated the dose to level 3 for the next cohort. However, in this case, the 3 + 3 design would retain dose level 4 because of ignoring LGT information. Based on all observed data at the current dose and by continuously applying the decision rule in Table 1, the trial ended when the maximum sample size of 30 was reached. Finally, observed DLT rates and LGT rates for five doses were (0/3, 0/3, 2/18, 1/6, 0/0) and (0/3, 0/3, 6/18, 3/6, 0/0), respectively. The MC-Keyboard selected dose 3 as the MTD with an estimated LGT rate of 33% and DLT rate of 11%, whereas the 3 + 3 design would select dose 4 (overly toxic, with an estimated LGT rate of 50%) or higher as MTD because of ignoring LGT information.

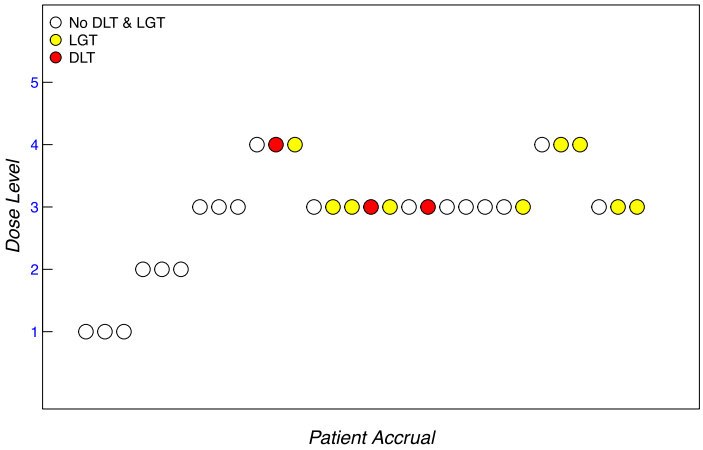

Figure 2. A hypothetical trial to illustrate the MC-Keyboard, with a total sample size of 30, cohort size of 3, target DLT rate of 20%, and target LGT rate of 35%. Dose level 3 was selected as the MTD based on DLT and LGT. In contrast, the 3 + 3 design, based only on DLT, would select dose level 4 or higher as the MTD. MC-Keyboard: multiple-constraint keyboard; DLT: dose-limiting toxicity; LGT: low-grade toxicity; MTD: maximum tolerated dose.

RESULTS

Simulation Configuration

We evaluated the operating characteristics of the MC-Keyboard through computer simulation. We compared the MC-Keyboard with the 3 + 3 and Keyboard designs,[10] which only consider DLT and ignore LGT. We considered five doses with a target LGT rate of 0.35 and a target DLT rate of 0.2. The total sample size is 30, and patients are treated in cohorts of 3. We considered eight scenarios that are displayed in Table 2. Under each scenario, we simulated 1000 trials. Because the 3 + 3 design often stops the trial

Table 2. Eight scenarios for true DLT and LGT rates, with the target DLT rate of 0.2 and the target LGT rate of 0.35

		Dose Level				
Scenario	**Toxicity**	**1**	**2**	**3**	**4**	**5**
1	DLT rate	0.10	**0.20**	0.27	0.38	0.42
	LGT rate	0.19	**0.35**	0.42	0.44	0.45
2	DLT rate	0.03	0.10	**0.20**	0.26	0.37
	LGT rate	0.10	0.18	**0.35**	0.43	0.50
3	DLT rate	0.03	0.06	0.10	**0.20**	0.26
	LGT rate	0.05	0.09	0.18	**0.35**	0.45
4	DLT rate	0.01	0.02	0.04	0.08	**0.20**
	LGT rate	0.05	0.06	0.10	0.18	**0.35**
5	DLT rate	**0.12**	0.20	0.28	0.34	0.43
	LGT rate	**0.35**	0.42	0.46	0.50	0.52
6	DLT rate	0.03	0.06	0.08	**0.12**	0.20
	LGT rate	0.06	0.10	0.19	**0.35**	0.46
7	DLT rate	0.03	0.05	**0.11**	0.20	0.33
	LGT rate	0.09	0.20	**0.35**	0.45	0.49
8	DLT rate	0.04	**0.10**	0.20	0.35	0.40
	LGT rate	0.18	**0.35**	0.46	0.52	0.53

The true maximum tolerated dose is in **bold**. DLT: dose-limiting toxicity; LGT: low-grade toxicity.

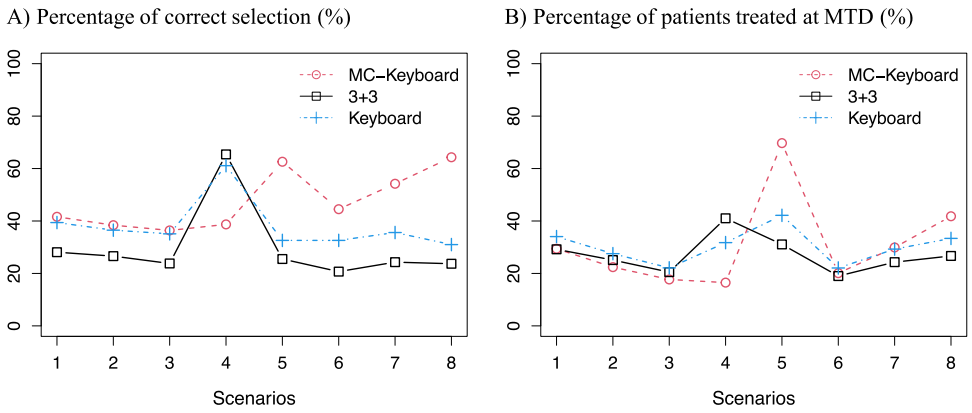

Figure 3. A) PCS of MTD and B) a percentage of patients treated at MTD for the 3 + 3, Keyboard, and MC-Keyboard designs, with a target DLT rate of 0.2 and target LGT rate of 0.35. PCS: percentage of correct selection; MTD: maximum tolerated dose; MC-Keyboard: multiple-constraint keyboard; DLT: dose-limiting toxicity; LGT: low-grade toxicity.

early (e.g., when 2/3 patients have DLT) before the total sample size is reached, to have a comparable sample size, the remaining patients are assigned to the selected MTD as a dose expansion.

Performance Metrics

Following Zhou et al,[21] we used the following performance metrics to summarize the operating characteristics of the designs.

(a) Accuracy

1. Percentage of correct selection (PCS) of the MTD
2. Percentage of patients treated at the MTD

(b) Safety

3. Percentage of trials that select over toxic doses (above the MTD)
4. Percentage of patients assigned to over-toxic doses (above the MTD)
5. Percentage of trials terminated early because of toxicity

(c) Reliability

6. The risk of overdosing, defined as the percentage of trials with 60% or more patients assigned to doses above the MTD
7. The risk of poor allocation, defined as the percentage of trials where less than six patients are assigned to the MTD
8. The risk of irrational dose assignment, defined as the percentage of trials that fail to de-escalate the dose when two out of the first three patients had DLTs at any dose

Simulation Results

Accuracy

Figure 3 presents the results for PCS and the percentage of patients treated at the MTD for the 3 + 3, Keyboard, and MC-Keyboard designs. Both the Keyboard and MC-Keyboard surpass the 3 + 3 design, exhibiting higher PCS rates and larger percentages of patients treated at the MTD. In scenarios 1–4, where the MTD aligns with the DLT-based MTD, the MC-Keyboard performs comparable to the Keyboard, with slightly lower percentages of patients treated at the MTD (e.g., 4.7% and 4.4% lower in scenarios 1 and 3) but slightly higher PCS rates (e.g., 2.2% and 1.9% higher in scenarios 1 and 2). However, in scenarios 5–8, where the MTD falls below the DLT-based MTD, the MC-Keyboard substantially outperforms the Keyboard. By ignoring LGT, the Keyboard, tends to select overly toxic doses, resulting in significantly lower PCS rates and percentages of patients treated at the MTD. For instance, in scenario 5, the Keyboard exhibits a 30% lower PCS rate and a 27.5% lower percentage of patients treated at the MTD compared with the MC-Keyboard. This outcome underscores the critical importance of considering both LGT and DLT in dose-finding trials.

Safety

Figure 4 shows the safety results, including the percentages of trials selecting doses above the MTD, the percentages of patients assigned to overly toxic doses, and the percentages of trials that were stopped early because of toxicity. The MC-Keyboard consistently demonstrates superior safety compared with the 3 + 3 and Keyboard designs. It exhibits lower percentages of overdosing selection, fewer patients overdosed in most scenarios, and higher percentages of trials stopped for toxicity in scenarios 5–8.

Reliability

Figure 5 displays the results of reliability metrics, including risks related to overdosing, poor allocation, and irrational dose assignment. Of note, all the designs exhibit no risk of irrational dose assignment. In terms of the risk of overdosing 60% or more patients, the MC-Keyboard stands out with a lower risk compared with the 3 + 3 and Keyboard designs, especially in scenarios 5–8. For instance, in scenario 8, the MC-Keyboard presents a 33.1% lower chance of overdosing 60% or more patients than the Keyboard. The risk of poor allocation with the MC-Keyboard is

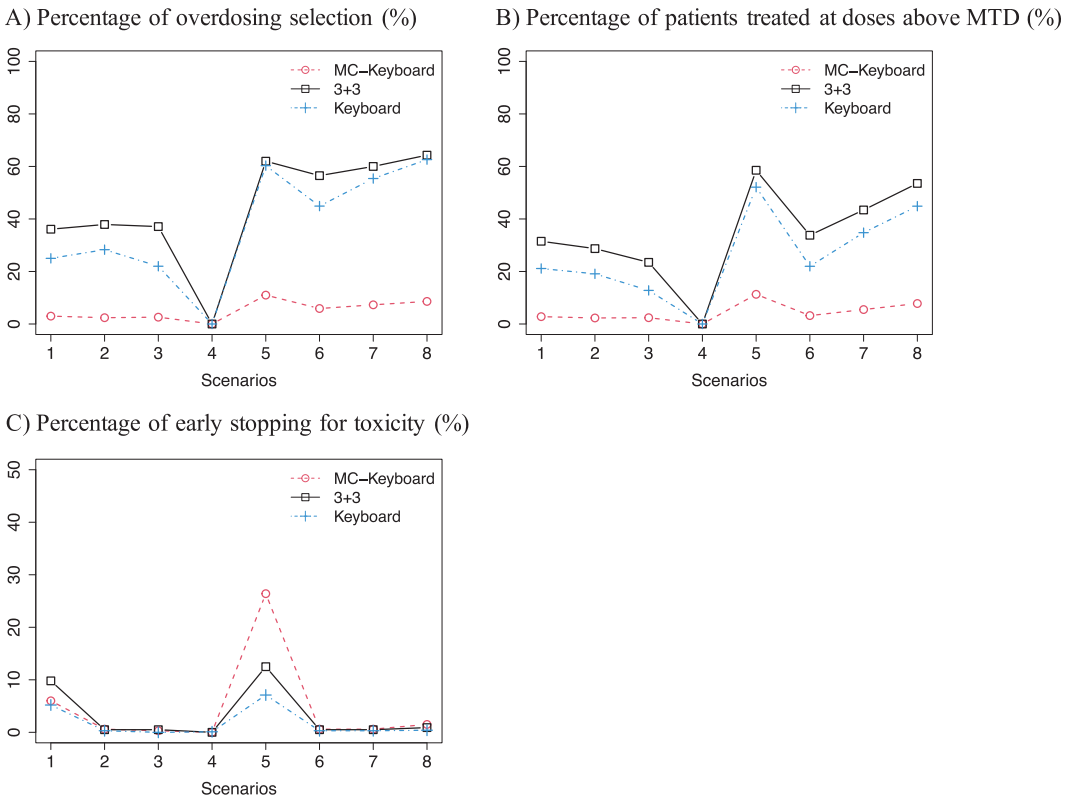

Figure 4. A) The percentage of trials selecting doses above MTD, B) a percentage of patients treated at doses above MTD, and C) a percentage of trials early stopped for toxicity for the 3 + 3, Keyboard, and MC-Keyboard designs, with the target DLT rate of 0.2 and target LGT rate of 0.35. MTD: maximum tolerated dose; MC-Keyboard: multiple-constraint keyboard; DLT: dose-limiting toxicity; LGT: low-grade toxicity.

higher than that of the Keyboard in scenarios 1–4, but it becomes lower in scenarios 5 through 8. Overall, the MC-Keyboard demonstrates a good performance in terms of reliability.

Sensitivity Analysis

We also examined different target DLT rates of 0.3 and target LGT rates of 0.4. The corresponding dose-escalation and de-escalation decision tables and true scenarios are displayed in Table S1 and Table S2, respectively (Supplementary Materials are available online). The simulation results (refer to Supplemental Figs. S1–S3) generally align with those previously described. Furthermore, we conducted a comparison between the MC-Keyboard and the model-based MC-CRM.[18] The simulation results are depicted in Supplemental Figures S4–S6. Overall, the MC-Keyboard's performance is either comparable or superior to that of the MC-CRM. The most notable advantage of the MC-Keyboard, in comparison to the MC-CRM, lies in its simplicity. The MC-CRM necessitates the specification of a complex statistical model and repeated estimation after each cohort, whereas implementing the MC-Keyboard merely requires referencing the decision table (Table 1).

DISCUSSION

Generalized from the Keyboard design, the MC-Keyboard accounts for both DLT and LGT, precisely guiding dose escalation or de-escalation and determining the MTD for immunotherapy or molecularly targeted agent trials. When compared with the 3 + 3 and Keyboard designs, the MC-Keyboard offers the same convenience for practical implementation but boasts higher accuracy in finding the MTD and enhanced safety. In contrast to the complex model-based design of the MC-CRM, the MC-Keyboard exhibits comparable accuracy in identifying the MTD while being safer, more reliable, and easier for clinicians to implement. Additionally, unlike the 3 + 3 design with a fixed cohort size of three patients, the MC-Keyboard can be applied with any specified cohort size, providing greater flexibility for clinicians who can select the cohort size based on practical considerations.

The MC-Keyboard primarily focuses on assessing DLT and LGT, defined as binary endpoints. In practice, it may be interesting to further distinguish between grade 1 and grade 2 toxicities, as well as the severity of different types of toxicities. An approach to incorporate this consideration is by assigning different weights to various grades and types of toxicities. For example, grade 1 increased amylase could be counted as 0.6 LGT, and grade 2 anorexia could be counted as 1 LGT. This approach can also be used to account for the larger uncertainty related to rating the attribution of LGT compared with DLT. For instance, an LGT with clear evidence of being drug-related could be counted as a full LGT, while an LGT with uncertainty regarding whether it is drug-

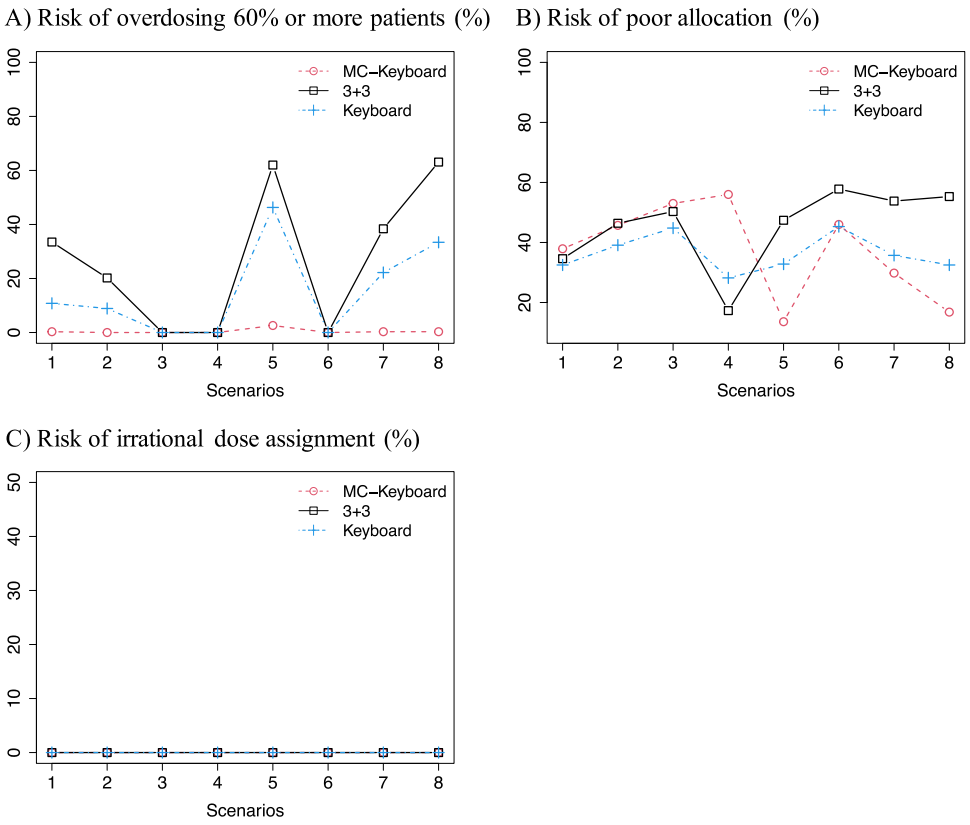

Figure 5. A) The risk of overdosing 60% or more patients, B) the risk of poor allocation, and C) the risk of irrational dose assignment for the 3 + 3, Keyboard, and MC-Keyboard designs, with the target DLT rate of 0.2 and target LGT rate of 0.35. MC-Keyboard: multiple-constraint keyboard; DLT: dose-limiting toxicity; LGT: low-grade toxicity.

related is counted as a fraction of LGT (eg, an LGT that is considered to have a 30% chance of being drug-related is counted as 0.3 LGT). Depending on the trial setting, for patients who have experienced multiple grades and types of toxicities, we can either count the most severe LGT they experienced or sum up all toxicities as the total toxicity burden. Then, the MC-Keyboard can be applied to guide dose escalation. It is worth noting that in this case, the number of LGT and DLT may not be integers, but this poses no issue based on a statistical method known as the quasi–binomial-likelihood approach.[15,19]

In addition, the MC-Keyboard can be readily extended to accommodate more than two categories of toxicities. For instance, in certain trials, it is useful to consider three levels of toxicities (low, medium, high) and incorporate three corresponding constraints to guide dose escalation and determine the MTD. Moreover, DLT and LGT tend to be late onset for many immunotherapy agents. This poses logistical challenges when implementing the MC-Keyboard because the design necessitates that patients treated at the current dose must complete DLT and LGT assessments before enrolling the next cohort. This can result in infeasibly long trials. To address this issue, an approach similar to that of TITE-BOIN[22] or Lin and Yuan [23] may be adopted to extend the MC-Keyboard's applicability and accommodate late-onset toxicities.

Lastly, it is worth noting that immunotherapy agents are frequently administered in combination with other agents. The MC-Keyboard can be further extended to determine MTDs in drug combination trials, following a similar approach to the Keyboard combination design by Pan and colleagues.[24]

Supplemental Material

Supplemental materials are available online with the article.

References

1. Linardou H, Gogas H. Toxicity management of immunotherapy for patients with metastatic melanoma. *Ann Transl Med*. 2016;4:272–282.
2. Paoletti X, Le Tourneau C, Verweij J, et al. Defining dose-limiting toxicity for phase 1 trials of molecularly targeted agents: results of a DLT-TARGETT international survey. *Eur J Cancer*. 2014;50:2050–2056.
3. Tighiouart M, Cook-Wiens G, Rogatko A. A Bayesian adaptive design for cancer phase I trials using a flexible range of doses. *J Biopharm Stat*. 2018;8:562–574.
4. Storer BE. Design and analysis of phase I clinical trials. *Biometrics*. 1989;45:925–937.

5. 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.
6. Yuan Y, Lee JJ, Hilsenbeck SG. Model-assisted designs for early-phase clinical trials: simplicity meets superiority. *JCO Precis Oncol*. 2019;3:PO.19.00032.
7. Yuan Y, Lin R, Lee JJ. *Model-assisted Bayesian Designs for Dose Finding and Optimization: Methods and Applications*. CRC Press; 2022.
8. Liu S, Yuan Y. Bayesian optimal interval designs for phase I clinical trials. *J R Stat Soc Series C*. 2015;64:507–523.
9. Drug development tools: Fit-for-purpose initiative. US Food and Drug Administration. Aug 19, 2022. Accessed Aug 5, 2022. www.fda.gov/drugs/development-approval-process-drugs/drug-development-tools-fit-purpose-initiative
10. 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.
11. Garon EB, Rizvi NA, Hui R, et al. Pembrolizumab for the treatment of non–small-cell lung cancer. *N Engl J Med*. 2015;372:2018–2028.
12. USPI ZYKADIA. US Food and Drug Administration.
13. Cho BC, Kim DW, Bearz A, et al. ASCEND-8: a randomized phase 1 study of ceritinib, 450 mg or 600 mg, taken with a low-fat meal versus 750 mg in fasted state in patients with anaplastic lymphoma kinase (ALK)-rearranged metastatic non–small cell lung cancer (NSCLC). *J Thorac Oncol*. 2017;12:1357–1367.
14. Bekele BN, Thall PF. Dose-finding based on multiple toxicities in a soft tissue sarcoma trial. *J Am Stat Assoc*. 2004;99:26–35.
15. Yuan Z, Chappell R, Bailey H. The continual reassessment method for multiple toxicity grades: a Bayesian quasi-likelihood approach. *Biometrics*. 2007;63:173–179.
16. Lee SM, Hershman DL, Martin P, et al. Toxicity burden score: a novel approach to summarize multiple toxic effects. *Ann Oncol*. 2012;23:537–541.
17. Lee SM, Hershman D, Martin P, et al. Validation of toxicity burden score for use in phase I clinical trials. *J Clin Oncol*. 2009;27(15_suppl):2514–2514.
18. Lee SM, Cheng B, Cheung YK. Continual reassessment method with multiple toxicity constraints. *Biostatistics*. 2011;12:386–398.
19. Mu R, Yuan Y, Xu J, et al. gBOIN: a unified model-assisted phase I trial design accounting for toxicity grades, and binary or continuous end points. *J R Stat Soc Series C*. 2019;68:289–308.
20. Ji Y, Wang SJ. Modified toxicity probability interval design: a safer and more reliable method than the 3+ 3 design for practical phase I trials. *J Clin Oncol*. 2013;31:1785.
21. Zhou H, Yuan Y, Nie L. Accuracy, safety, and reliability of novel phase I trial designs. Clinical Cancer Research, 2018;24:4357–4364.
22. Yuan Y, Lin R, Li D, et al. Time-to-event Bayesian optimal interval design to accelerate phase I trials. *Clin Cancer Res*. 2018;24:4921–4930.
23. Lin R, Yuan Y. Time-to-event model-assisted designs for dose-finding trials with delayed toxicity. *Biostatistics*. 2020;21:807–824.
24. Pan H, Lin R, Zhou Y, Yuan Y. Keyboard design for phase I drug-combination trials. *Contemp Clin Trials*. 2020;92:105972.