Keyboard: A Novel Bayesian Toxicity Probability Interval Design for Phase I Clinical Trials
Fangrong Yan1, Sumithra J. Mandrekar2, and Ying Yuan3

Abstract

The primary objective of phase I oncology trials is to find the MTD. The 3 + 3 design is easy to implement but performs poorly in finding the MTD. A newer design, such as the modified toxicity probability interval (mTPI) design, provides better accuracy to identify the MTD but tends to overdose patients. We propose the keyboard design, an intuitive Bayesian design that conducts dose escalation and de-escalation based on whether the strongest key, defined as the dosing interval that most likely contains the current dose, is below or above the target dosing interval. The keyboard design can be implemented in a simple way, similar to the traditional 3 + 3 design, but provides more flexibility for choosing the target toxicity rate and cohort size. Our simulation studies demonstrate that compared with the 3 + 3 design, the keyboard design has favorable operating characteristics in terms of identifying the MTD. Compared with the mTPI design, the keyboard design is safer, with a substantially lower risk of treating patients at overly toxic doses, and has the better precision to identify the MTD, thereby providing a useful upgrade to the mTPI design. Software freely available at http://www.trialdesign.org facilitates the application of the keyboard design. Clin Cancer Res; 23(15); 1-10. ©2017 AACR. See related commentary by Paoletti et al., p. 3977

Introduction

With increasing awareness of the limitation and poor operating characteristics of the traditional 3 + 3 design (1-5), clinical researchers, pharmaceutical companies, and regulatory agencies are more open to using innovative phase I trial designs that more accurately identify the MTD while maintaining patient safety. Phase I dose-finding designs can be generally classified into algorithm-based designs and model-based designs (6, 7). The algorithm-based design uses a set of simple, prespecified rules (or algorithm) to determine the dose escalation and de-escalation. Examples include the 3 + 3 design (1) and up-and-down design (8). Although performing poorly in identifying the MTD (5-7), the 3 + 3 design is still the dominant phase I trial design in practice because of its simplicity for implementation. The model-based design assumes a parametric model for the dose-toxicity relationship. As information accrues during the trial, the dose-toxicity relationship is reevaluated by updating the estimates of the model parameters and then used to guide the dose allocation for subsequent patients. Examples of model-based designs include the continuous reassessment method (CRM; ref. 9) and dose escalation with overdose control (10). Although the model-based design, such as the CRM, yields better performance than the algorithm-based design (6, 7, 11), it is considered by many to be statistically and computationally complex, leading practitioners to perceive dose allocations as coming from a “black box.” As a result, the use of the model-based design has been limited in practice (2).

Recently, there is increasing interest in a new class of designs that combines the simplicity of algorithm-based designs and the good performance of model-based designs. This class of designs utilizes a model for efficient decision making, similar to the model-based design, but its rule of dose escalation and de-escalation can be pretabulated before the onset of the trial in a fashion similar to the algorithm-based design. We refer to this class of new designs as the “model-assisted” design. Examples of model-assisted designs include the modified toxicity probability interval (mTPI) design (12, 13) and Bayesian optimal interval (BOIN) design (14, 15).

Ji and Wang (13) reported a simulation study to compare the mTPI with the 3 + 3 design, and concluded that a major advantage of the mTPI design is that it is safer and ‘the 3 + 3 design has higher risks of exposing patients to toxic dose above the MTD than the mTPI’ (in Abstract; ref. 13) with matched sample sizes. In what follows, however, we show that the mTPI design actually has a high risk of overdosing patients and elucidate the reason behind it. Motivated by that finding, we propose the keyboard design, a Bayesian design that uses a series of dosing intervals (referred to as keys) to direct dose escalation and de-escalation. The keyboard design provides a useful upgrade of the mTPI design. It shares the same simplicity as the mTPI design, which can be implemented in a simple way like the 3 + 3 design, but yields better overdose control and higher accuracy to identify the true MTD.

Methods

An overview of the mTPI design

The mTPI design starts by defining three dosing intervals, namely the underdosing interval, proper dosing interval, and overdosing interval. For example, given the target dose-limiting
toxicity (DLT) rate of 0.2, the three intervals may be, respectively, defined as (0, 0.15), (0.15, 0.25), and (0.25, 1) in terms of the DLT rate. The mTPI design makes the decision of dose escalation and de-escalation based on the unit probability mass (UPM) of the three intervals. Given a specific dosing interval, the UPM is defined as the posterior probability of the interval divided by the length of the interval, where the posterior probability of the interval represents the probability that the true DLT rate of the current dose is located within the interval given the observed data. For example, the posterior probability of the overdosing interval represents the probability that the current dose is an overdose (or located in the overdosing interval). For treating the next patient, the mTPI design escalates the dose if the UPM of the underdosing interval is the largest, de-escalates the dose if the UPM of the overdosing interval is the largest, and stays at the same dose if the UPM of the proper dosing interval is the largest.

An attractive feature of the mTPI design is that its decision rule can be pretabulated and thus is easy to implement in practice. Table 1 shows the decision rule of the mTPI design when the target DLT rate is 0.2 or 0.3, with (0.15, 0.25) and (0.25, 0.35), respectively, as the proper dosing interval, given that 3 or 6 patients have been treated at the current dose. Following Ji and Wang (13), two variations of the 3 + 3 design are also displayed in Table 1 as the comparators: the 3 + 3\(^L\) design targeting the MTD with no more than one DLT out of 6 patients, and the 3 + 3\(^H\) design targeting the MTD with no more than two DLTs out of 6 patients. Supplementary Fig. S1 provides the dose escalation and de-escalation schemas of these two 3 + 3 designs.

**Comparison of the mTPI design with the 3 + 3 designs**

It is immediately clear that it is not theoretically possible for the mTPI design to be safer than the 3 + 3 design, because the dose escalation rule of the mTPI is more aggressive than that of the 3 + 3 design. Specifically, when 2 of 6 (or 3/6) patients have DLTs, the 3 + 3\(^L\) (or 3 + 3\(^H\)) design will de-escalate the dose, whereas the mTPI design will continue treating patients at the same dose. In other words, given the target DLT rate of 30\%, even when we observe that 3 of 6 = 50\% of patients experienced DLTs, the mTPI design will continue to treat the next cohort of patients at the same dose, which is intuitively excessively risky. The aggressiveness of the mTPI design stems from its foundation, that is, using the UPM as the criterion to determine dose escalation. To see this problem, consider a trial for which the target toxicity rate is 0.2 and the underdosing, proper dosing, and overdosing intervals are (0, 0.17), (0.17, 0.23), and (0.23, 1), respectively. Suppose that at a certain stage of the trial, the observed data indicate that the posterior probabilities of the underdosing interval, proper dosing interval, and overdosing interval are 0.01, 0.09, and 0.9, respectively. That is, there is a 90% chance that the current dose is overdosing patients and only a 9% chance that the current dose provides proper dosing. Despite such dominant evidence of overdosing, the mTPI design retains the same dose for treating the next new patient because the UPM for the proper dosing interval is the largest. Specifically, the UPM for the proper dosing interval is 0.09/(0.23 – 0.17) = 1.5, and the UPM for the overdosing interval is 0.9/(1 – 0.23) = 1.7. This example demonstrates that the UPM cannot appropriately quantify the evidence of overdosing. The fundamental issue is that the UPM nonproportionally weighs the evidence of underdosing, proper dosing, and overdosing due to different lengths of the three intervals (recall that the UPM is defined as the posterior probability of the interval divided by the length of the interval). As the overdosing interval is typically the longest, the UPM has bias in suppressing the evidence of overdosing. Consequently, the mTPI design tends to keep treating patients at a toxic dose even when there is actually strong evidence for that dose being overly toxic.

The simulation results of Ji and Wang (13), however, showed that the mTPI design is safer than the 3 + 3 design, with substantially fewer patients treated above the MTD (see Fig. 5 in Ji and Wang ref. 13). This is because they used the average sample size of the 3 + 3 design (across thousands of simulated trials) as the sample size of the mTPI as a way to match the sample size between the two designs. Such sample size matching seems reasonable but actually is problematic and leads to a biased comparison. The key is that the sample size of the 3 + 3 design is not fixed and demonstrates a bell-shaped distribution. Supplementary Fig. S2 shows the sample size distribution of the 3 + 3\(^L\) design in 10,000 simulated trials when the true DLT rates for five dose levels are 0.01, 0.12, 0.30, 0.41, and 0.55, respectively. Using the mean sample size of the 3 + 3 design (i.e., 14.5 patients) as the sample size of the mTPI would truncate all larger sample sizes and thus largely forbid the mTPI to reach overly toxic doses. In other words, it makes the mTPI design artificially safe, simply because there are not enough patients to reach overly toxic doses. In the numerical study described later, we show that with a sample size that yields reasonable precision (e.g., 40\% or more chance) of correctly selecting the MTD, the mTPI has a substantially higher risk of treating a large percentage of patients above the MTD than the 3 + 3 design.

A fundamental characteristic, which is also a limitation of the 3 + 3 design, is that its sample size is not fixed and is too small to provide adequate information to learn the dose–toxicity relationship. In contrast to what many might believe, the poor operating characteristics of the 3 + 3 design are not caused by its dose escalation/de-escalation rule, which indeed is sensible for the target DLT rate of around 20\% to 30\%, but by the excessively small sample size that provides too little information to reliably estimate the true DLT rate. Specifically, under the 3 + 3 design, the number of patients treated at any dose cannot be more than 6. If 1 of 6 patients experiences a DLT, the estimate of the toxicity rate, 1/6 = 16.7\%, seems low, but the 95\% exact confidence interval (CI) for that estimate is (0.004 – 0.641), indicating that the

<table>
<thead>
<tr>
<th>Table 1. Dose escalation rules for 3 + 3 and mTPI designs</th>
</tr>
</thead>
<tbody>
<tr>
<td>No. of patients</td>
</tr>
<tr>
<td>-----------------</td>
</tr>
<tr>
<td>3 + 3(^L)</td>
</tr>
<tr>
<td>mTPI (p_T = 20%)</td>
</tr>
<tr>
<td>3 + 3(^H)</td>
</tr>
<tr>
<td>mTPI (p_T = 30%)</td>
</tr>
</tbody>
</table>

Abbreviations: D, de-escalation; E, escalation; p_T, target DLT rate; S, stay at the same dose; Se, stop the trial and select the current dose as the MTD.
true toxicity rate can be as high as 64.1%. Conversely, if 3 of 6 patients experience DLTs, the estimate of the toxicity rate, $3/6 = 50\%$, seems very high, but the 95% CI for that estimate is $(0.118–0.88)$, and the true toxicity rate can be as low as 11.8%. This is the fundamental reason why the $3 + 3$ design has poor accuracy when seeking the true MTD. An important improvement of novel designs over the $3 + 3$ design is that they allow investigators to use a larger, more reasonable sample size to continuously learn the dose–toxicity relationship, thereby improving the MTD identification accuracy. Thus, besides causing the aforementioned bias, it is generally not appropriate to use the sample size of the $3 + 3$ design for evaluating the performance of a novel design, which would be analogous to evaluating the performance of a racing car in a street with a speed limit of 40 miles per hour.

**Keyboard design: Definition and framework**

We propose the keyboard design that retains the simplicity of the mTPI design, but yields better overdose control and higher accuracy to identify the true MTD. Similar to the mTPI design, the keyboard design is a Bayesian model-based design that relies on the posterior distribution of the toxicity probability to guide dose escalation and de-escalation. The innovation is that the keyboard design defines a series of equal-width dosing intervals (or keys) to present the potential locations of the true toxicity of a dose and guide the dose escalation and de-escalation, whereas the mTPI design uses the UPMs of three dosing intervals (i.e., underdosing, proper dosing, and overdosing) to determine the dose transition. As shown in Fig. 1, the keyboard design starts by specifying a proper dosing interval, referred to as the "target key," and then

![Figure 1](https://example.com/figure1.png)

**Figure 1.**
The dose escalation/de-escalation rule of the keyboard design. If the strongest key (red) is on the right side of the target key (blue), de-escalate the dose (top); if the strongest key is on the left side of the target key, escalate the dose (middle); and if the strongest key is the target key, retain the current dose (bottom). The curves in the panels are the posterior distributions of the DLT rate of the current dose.
populates this interval toward both sides of the target key, forming a series of keys of equal width that span the range of 0 to 1. For example, given the proper dosing interval or target key of (0.25, 0.35), on its left side, we form two keys of width 0.1, that is, (0.15, 0.25) and (0.05, 0.15); and on its right side, we form six keys of 0.1, that is, (0.35, 0.45), (0.45, 0.55), (0.55, 0.65), (0.65, 0.75), (0.75, 0.85), and (0.85, 0.95). We require all keys to have the same length to avoid nonproportional weights for different intervals, which leads to the problematic behavior of the mTPI design described previously. In addition, using equal-length keys also simplifies the practical implementation of the design because users only need to specify the target key based on which the other keys are automatically populated. All keys must be within the range of 0 to 1 because the DLT rate must be located between 0 and 1. Some values are automatically populated. All keys must be within the range of 0 to 1 only need to specify the target key based on which the other keys are automatically populated. All keys must be within the range of 0 to 1 because the DLT rate must be located between 0 and 1.

The keyboard design stops the trial when the preselected maximum sample size is reached or the dose elimination rule is evaluated after every cohort is enrolled, and a minimum of 3 patients must be treated before a dose can be eliminated.

Table 2. Dose escalation and de-escalation rules for target DLT rates of 0.2 and 0.3 under the keyboard design

<table>
<thead>
<tr>
<th>Action</th>
<th>Target DLT rate = 20%&lt;sup&gt;a&lt;/sup&gt;</th>
<th>Target DLT rate = 30%&lt;sup&gt;b&lt;/sup&gt;</th>
</tr>
</thead>
<tbody>
<tr>
<td>Escalate if number of DLTs ≤</td>
<td>0</td>
<td>0</td>
</tr>
<tr>
<td>De-escalate if number of DLTs ≥</td>
<td>1</td>
<td>1</td>
</tr>
<tr>
<td>Eliminate if number of DLTs ≥&lt;sup&gt;c&lt;/sup&gt;</td>
<td>NA</td>
<td>NA</td>
</tr>
</tbody>
</table>

Abbreviation: NA, not applicable.

<sup>a</sup>The proper dosing intervals are (15%, 23%) and (25%, 35%) for target DLT rates of 20% and 30%, respectively, as in the simulation study.

<sup>b</sup>When the current dose is eliminated from the trial, the higher doses should also be eliminated and the dose is automatically de-escalated to the next lower level for treating the next new patient. A minimum of three patients must be treated before a dose can be eliminated.

For patient safety, a dose elimination rule is added to the keyboard design, which says that if the observed data indicate that there is more than a 95% chance that the current dose is above the MTD, we eliminate that and higher doses from the trial to prevent exposing future patients to these overly toxic doses. Such a safety dose elimination rule is displayed in Table 2 in the rows starting with "Eliminate if number of DLT ≥." For example, when the target DLT rate is 30%, if 3 of 6 patients experienced DLTs, the keyboard design will de-escalate the dose, whereas the mTPI will stay at the same dose for treating the next cohort of patients.
Figure 2.
Implementation of the keyboard design using a hypothetical trial. The numbers indicate the order of patient accrual.

Dose–toxicity monotonicity assumption. Such violation is possible due to the randomness of observations, even the truth that dose level 3 is more toxic than dose level 2. The isotonic regression identifies these two violators and replaces their DLT estimates with their average, that is, \( (1/3 + 0/3)/2 = 1/6 \), resulting in the isotonic regression DLT estimates \((0/3, 1/6, 1/6, 3/15, 2/4) = (0, 16.7\%, 16.7\%, 20\%, 50\%)\), which monotonically increase with the dose level. On the basis of this isotonic estimate, assuming that the trial goal is to find the dose with the DLT rate of 20%, dose level 4 will be selected as the MTD. In the case that there is no violator, for example, when the observed DLT data at five doses are \((0/3, 0/3, 1/6, 3/15, 2/3)\), the isotonic regression directly uses the observed DLT rates as the final estimates for the MTD selection. The statistical procedures to determine the strongest key are provided in the Supplementary Data. The decision table of the keyboard design can be easily generated using the computer program discussed in the Supplementary Data and available at http://www.trialdesign.org.

Keyboard design versus the mTPI and the 3 + 3 designs
Both the keyboard design and the mTPI design allow for predetermining the decision table for easy implementation; however, the designs use different statistical criteria and principles. Specifically, the mTPI design makes dose transitions based on the “highest UPM,” which lacks explicit interpretation and biases the estimate of true toxicity, and the keyboard design makes dose transitions based on the “strongest key,” which has an intuitive interpretation as the dosing interval where the current dose is most likely located. As a result, the two designs have different operating characteristics.

Compared with the 3 + 3 design, the keyboard design enjoys several advantages. First, the keyboard design is more flexible and can target any prespecified DLT rate. For example, for cancer that does not have effective treatment available, a higher target DLT rate, such as 30%, may be an acceptable trade-off to achieve higher treatment efficacy; whereas for cancer with effective treatment, a lower target DLT rate, for example, 20%, may be more appropriate. Second, unlike the 3 + 3 design, for which the dose escalation and de-escalation decisions can be made only when we have 3 or 6 evaluable patients, the keyboard design provides a decision rule for all possible sample sizes, which allows for more efficient, fully sequential decision making (i.e., making the decision of dose escalation/de-escalation after each patient), as well as handling
the common situation that some patients are not evaluable for toxicity (e.g., among 6 enrolled patients, only 5 are evaluable). Third, the keyboard design allows for prespecification of the sample size, thereby providing users the opportunity to calibrate and choose an appropriate sample size to achieve desirable operating characteristics (e.g., attain a certain MTD selection accuracy). In the 3 + 3 design, we do not have a fixed sample size for a trial, and often, the number of patients treated at each dose level is too small to reliably estimate the true toxicity rate at each dose level as well as the DLT rate.

As a side note, the keyboard design is different from the BOIN design (14, 15), another novel Bayesian interval design, in several aspects. The two designs are based on different statistical frameworks: The keyboard design relies on the posterior distribution of the toxicity probability to make the decision of dose escalation and de-escalation, whereas the BOIN design does not involve the calculation of posterior distributions and directly uses the observed toxicity rate for dose escalation and de-escalation (i.e., more similar to the 3 + 3 design). Operatively, the two designs are also different. The BOIN design compares the observed toxicity rate with a pair of fixed escalation and de-escalation cutoffs to direct the dose escalation and de-escalation, whereas the keyboard design compares the number of DLTs observed at the current dose with the escalation and de-escalation boundaries that vary according to the number patients treated. In other words, the BOIN design is slightly simpler.

Evaluation of design performance

Computer simulations were used to compare the operating characteristics of the keyboard design with those of the mTPI and 3 + 3 designs. We considered a phase I trial with five dose levels and a maximum sample size of 30 patients. We investigated target DLT rates of 20% and 30%, with the proper dosing interval of (15%, 23%) and (25%, 35%), respectively. For each of the target DLT rates, 10 representative toxicity scenarios (i.e., the true toxicity rates of the five investigational doses) were examined, which vary in the location of the MTD and the shape of the dose-toxicity curve (see Supplementary Fig. S3; Supplementary Table S1). Under each scenario, 10,000 trials were simulated to compare the designs. We used the 3 + 3 design when the target DLT rate was 20% and the 3 + 3H design when the target DLT rate was 30%.

To illustrate the flexibility of the mTPI and keyboard designs, these two designs were implemented in a fully sequential way where patients are treated one by one, that is, the cohort size is 1. We here mainly focus on the comparison of the 3 + 3, mTPI, and keyboard designs to demonstrate that the keyboard design addresses the issues of the mTPI design. The comparison of the keyboard design with other novel designs, such as CRM and BOIN designs, is reported in the Supplementary Data.

Four metrics are used to evaluate the performance of the designs:

1. The percentage of correct selection (PCS), defined as the percentage of times that the MTD is correctly selected in 10,000 simulated trials.
2. The average percentage of patients treated at the MTD across 10,000 simulated trials.
3. The average percentage of patients treated above the MTD across 10,000 simulated trials.
4. The risk of overdosing, defined as the percentage of simulated trials in which a large percentage (e.g., 60% or 80%) of patients are treated at doses above the MTD.

As expected, the average sample size of the 3 + 3 design is substantially smaller than the mTPI and keyboard designs. The average sample size of the 3 + 3 design is about 30 patients (see Supplementary Fig. S4). As described previously, although one might regard that as an advantage, the excessively small sample size of the 3 + 3 design renders it a high risk for misidentifying the MTD.

Results

PCS of the MTD

As shown in Fig. 3, overall, the 3 + 3 design performs the worst, with the lowest PCS of the MTD, and the keyboard design performs the best, with the highest PCS of the MTD. The PCS of the MTD for the keyboard design is often 10% to 15% higher than that for the 3 + 3 design, and 9% to 12% higher than that for the mTPI design. As expected, the average sample size of the 3 + 3 design is substantially smaller than the mTPI and keyboard designs. The average sample size of the 3 + 3 design is about 15 patients in most scenarios, whereas that of the mTPI and keyboard designs is generally close to 30 patients (see Supplementary Fig. S4). As described previously, although one might regard that as an advantage, the excessively small sample size of the 3 + 3 design renders it a high risk for misidentifying the MTD.
resulting in very undesirable consequences, for example, exposing a large number of patients to an overly toxic or subtherapeutic dose in subsequent cohort expansion or phase II trials because of misidentification of the MTD.

Percentage of patients allocated to the MTD and above the MTD

In terms of the percentage of patients allocated to the MTD, the performance of the 3 + 3 design depends on the location of the MTD (see Fig. 4A). When the MTD is located at the lowest dose (i.e., scenarios 1 and 2), the 3 + 3 design performs well and allocates a large percentage of patients to the MTD; however, when the MTD is located at higher doses (doses 4 and 5, corresponding to scenarios 7 – 10), the 3 + 3 design performs worse than the mTPI and keyboard designs. The performance of the keyboard design is slightly better than that of the mTPI design. In terms of the percentage of patients allocated above the MTD (see Fig. 4B), the 3 + 3 design performs best, whereas the keyboard and mTPI designs are comparable.

Risk of overdosing

For the risk of treating patients with overly toxic doses, the mTPI design has a substantially higher risk of overdosing patients than the 3 + 3 and keyboard designs. The mTPI design often has more than 35% chance of assigning more than 60% of patients to overly toxic doses (Fig. 5A) and more than 30% chance of assigning more than 80% of patients to overly toxic doses (Fig. 5B). The keyboard design is substantially safer than the mTPI design. The risk of overdosing patients under the keyboard design is often 50% lower than that under the mTPI design, especially for the risk of overdosing 80% of the patients. The 3 + 3 design generally has the lowest risk of overdosing, which is consistent with previous research that found the 3 + 3 design to be overly conservative (4–6) and thus has low accuracy to identify the MTD. The keyboard design achieves a good balance in safety (i.e., the risk of overdosing) and identifying the MTD. Compared with the 3 + 3 design, the keyboard design has higher PCS of the MTD (see Fig. 3), and compared with the mTPI, the keyboard design has a lower risk of overdosing patients.

Comparison to CRM and BOIN

Supplementary Figs. S5–S7 summarize the results of comparing the keyboard design with the CRM and BOIN designs. Overall, the performance of the keyboard design is roughly comparable to CRM and BOIN.
with that of the CRM, noting that the keyboard design is simpler to implement. Surprisingly, the keyboard and BOIN designs yield nearly identical performance, although they are based on different statistical approaches, as described previously. For practitioners who have used and are comfortable with the mTPI design, the keyboard design provides a useful alternative and seamless upgrade of the mTPI design.

**Discussion**

We have proposed the Bayesian keyboard design for phase I clinical trials. The keyboard design is easy to implement in a manner similar to the 3 + 3 design but provides much more flexibility in choosing the target toxicity rate and higher precision to correctly select the MTD. Compared with the mTPI design, the keyboard design is safer, with a lower chance of overdosing patients and with a higher precision to identify the MTD. The keyboard design provides a useful upgrade of the mTPI design. Compared with other model-based designs, such as the CRM, the key advantage of the keyboard design is that it is much simpler to implement, with little sacrifice of performance. Using the keyboard design, dose escalation and de-escalation can be easily determined on the basis of the pretabulated decision rule (see Table 2) in a way similar to the 3 + 3 design. There are no complicated model estimation and statistical calculation for the trial conduct. In some sense, the keyboard design bridges the 3 + 3 design and more complicated model-based design.

The keyboard design requires specification of the proper dosing interval or keys. The sensitivity analysis (see Supplementary Fig. S8) shows that the performance of the keyboard design is not sensitive to specification of the keys. In practice, as a safeguard, a simulation study should be carried out to confirm that the specified keys yield desirable operating characteristics prior to the launch of the actual trial. This can be easily done using the software provided at http://www.trialdesign.org and http://odin.mdacc.tmc.edu/~yyuan/index_code.html.

Unlike the 3 + 3 design, which mandates treating patients in cohort sizes of 3, the keyboard design is flexible and can accommodate any prespecified cohort size. Although the fully sequential approach with cohort size of 1 (i.e., treating patients one by one) is
generally more efficient, the choice of cohort size in practice depends on practical considerations, for example, trial duration. Given a fixed total sample size and time window for DLT assessment, the fully sequential approach is appropriate when the accrual is slow such that enrolled patients mostly have completed their DLT assessment before enrolling the next new patient. However, when the accrual is relatively fast, the fully sequential approach becomes logistically difficult because we have to wait until each patient completes his/her DLT assessment before enrolling the next new patient, which leads to a long trial duration and overly frequent interruption of the accrual. In this case, it is preferable to treat patients by cohort, such as 3 patients per cohort, such that waiting for the DLT assessment is required only after 3 patients are enrolled. This reduces the trial duration and the frequency of interrupting the accrual. For patient safety, a large cohort size, such as 6 patients, should be avoided because it may expose a larger number of patients to an overly toxic dose (e.g., 6 of 6 patients all experienced DLT) before the trial design can perform dose de-escalation.

The keyboard design is suitable for finding the MTD for agents for which toxicity and efficacy both monotonically increase with the dose, for example, cytotoxic agents. For noncytotoxic agents (e.g., antibodies and vaccines), for which efficacy may not increase with the dose, the objective of the dose-finding trial is often to identify the optimal biological dose (OBD), a dose that is safe and has the highest efficacy. In this case, the keyboard design is not recommended. Some novel designs have been proposed to find the OBD for noncytotoxic (or molecularly targeted) agents (17, 18). Nonetheless, the keyboard design is still preferable to the 3 + 3 design. For example, in the scenario that no DLTs are expected, as is often the case in antibody and vaccine trials, by using a cohort size of 1, the keyboard design allows for quicker dose escalation than the 3 + 3 design, which mandates the use of a cohort size of 3. There are several important directions in which to extend the keyboard design to handle more complicated clinical trials. For example, the keyboard design cannot be directly used for drug combination trials. Given the importance of drug combination trials (19), it is of great practical interest to extend the keyboard design to accommodate drug combination trials. One possible approach is to combine the keyboard design with the sequential dose-finding strategy to find the MTD(s) in drug combination trials (20), that is, to test each combination sequentially each time using the keyboard method. The sequential process is efficient because it uses accumulating trial data to adaptively shrink the MTD search space, but it may lead to a longer trial duration. If that is a concern, an alternative approach is to concurrently test different doses of the drugs in combination with multiple “arms” running in parallel. In addition, to speed up drug development, it is also of interest to incorporate the keyboard design into seamless early phase trials. A simple approach is to perform cohort expansion after the keyboard design identifies the MTD. However, a more efficient and flexible approach is to take the phase I to phase II paradigm (21) and simultaneously consider toxicity and efficacy for optimizing the dose. Finally, the keyboard design requires that the toxicity outcome be quickly ascertainable to inform the dose escalation and de-escalation for the next new patient. If late-onset toxicity is likely, the keyboard design is not suitable for finding the MTD, because it requires repeated suspension of accrual, resulting in unfeasibly long trials. Some model-based designs (22–24) are available to handle late-onset toxicity but are statistically complicated. Extending the keyboard design for late-onset toxicity while maintaining the simplicity of the design is an interesting topic for future research.

Disclosure of Potential Conflicts of Interest
No potential conflicts of interest were disclosed.

Authors’ Contributions
Conception and design: F. Yan, Y. Yuan
Development of methodology: F. Yan, S.J. Mandrekar, Y. Yuan
Acquisition of data (provided animals, acquired and managed patients, provided facilities, etc.): F. Yan, S.J. Mandrekar, Y. Yuan
Analysis and interpretation of data (e.g., statistical analysis, biostatistics, computational analysis): F. Yan, S.J. Mandrekar, Y. Yuan
Writing, review, and/or revision of the manuscript: F. Yan, S.J. Mandrekar, Y. Yuan
Administrative, technical, or material support (i.e., reporting or organizing data, constructing databases): Y. Yuan
Study supervision: Y. Yuan

Acknowledgments
The authors thank Lee Ann Chastain for editorial help on the manuscript.

Grant Support
Y. Yuan was supported in part by the NIH under award numbers P50CA098258 and RO1CA154591.

The costs of publication of this article were defrayed in part by the payment of page charges. This article must therefore be hereby marked advertisement in accordance with 18 U.S.C. Section 1734 solely to indicate this fact.

Received January 25, 2017; revised March 21, 2017; accepted May 22, 2017; published OnlineFirst May 25, 2017.

References

www.aacjournals.org
Keyboard: A Novel Bayesian Toxicity Probability Interval Design for Phase I Clinical Trials
Fangrong Yan, Sumithra J. Mandrekar and Ying Yuan

Access the most recent version of this article at:
doi:10.1158/1078-0432.CCR-17-0220

Access the most recent supplemental material at:
http://clincancerres.aacrjournals.org/content/suppl/2017/05/25/1078-0432.CCR-17-0220.DC1

Sign up to receive free email-alerts related to this article or journal.

To order reprints of this article or to subscribe to the journal, contact the AACR Publications Department at pubs@aacr.org.

To request permission to re-use all or part of this article, use this link
http://clincancerres.aacrjournals.org/content/early/2017/07/11/1078-0432.CCR-17-0220.
Click on "Request Permissions" which will take you to the Copyright Clearance Center's (CCC) Rightslink site.