Published in final edited form as:

Contemp Clin Trials. 2013 September ; 36(1): 218–227. doi:10.1016/j.cct.2013.07.002.

Up-and-down designs for phase I clinical trials

Suyu Liu^{a,*}, Chunyan Cai^b, and Jing Ning^a

^a Department of Biostatistics The University of Texas M. D. Anderson Cancer Center

^b Biostatistics/Epidemiology/Research Design Core Center for Clinical and Translational Sciences
The University of Texas Health Science Center at Houston

Abstract

Various up-and-down designs have been proposed to improve the operating characteristics of the traditional “3+3” design, but they have been of limited use in practice. A major impediment to the adoption of the improved up-and-down designs is a lack of general guidance and a comprehensive assessment of the operating characteristics of these designs under practical clinical settings. To fill this gap, we review six up-and-down designs: the “3+3” design, accelerated titration design, biased coin design, k-in-a-row design, group up-and-down design and cumulative group up-and-down design. We conduct comprehensive simulation studies to evaluate their operating characteristics under various practical settings, and compare their performance to a theoretical optimal bound of nonparametric designs. The results show that the cumulative group up-and-down design has the best overall performance in terms of selecting the maximum tolerated dose (MTD), assigning patients to the MTD and patient safety. Its performance is generally close to the upper bound of nonparametric designs, but improvement seems possible in some cases.

Keywords

Optimal design; upper bound; dose finding; maximum tolerated dose

1 Introduction

The primary objective of a phase I clinical trial is to identify the maximum tolerated dose (MTD) of a new drug, which is defined as the dose with a dose-limiting toxicity (DLT) probability that is closest to the target toxicity rate. A phase I clinical trial is important because it determines the MTD that will be further investigated in the ensuing phase II or III trials. Misidentification of the MTD could result in misleading results and serious consequences in the subsequent larger-scale trials.

A class of phase I trial designs widely used in practice are the so-called algorithm-based designs, which are also known as up-and-down designs. The major advantage of this class of designs is their simplicity of implementation. Algorithm-based designs do not require any parametric assumptions on the dose-toxicity curve and strictly conduct dose escalation and deescalation according to prespecified algorithms. Because clinicians know under what

*syliu@mdanderson.org.

precise circumstances dose escalation and deescalation will occur *a priori*, before the onset of the trial, it is often easy for them to understand and evaluate up-and-down designs based on their clinical experience.

The most well-known up-and-down design is the “3+3” design [1]. Although dominant in practice, the “3+3” design has been widely criticized for its poor operating characteristics [1, 2, 3]. Examples of this include a tendency for the resulting estimators of the MTD to be biased or inconsistent and for a large percentage of patients to be treated at doses below the MTD, and severe restrictions as to the choice of the targeted toxicity probability. Other up-and-down designs have been proposed to achieve better operating characteristics. Simon et al. [4] proposed accelerated titration designs to reduce the number of patients treated at subtherapeutic dose levels in the “3+3” design. Durham and Flournoy [5] proposed the biased coin design (BCD), in which the decision of dose escalation and deescalation is based on the toxicity outcome from the most recently treated patient. Wetherill [6] and Gezmu [7] investigated a *k*-in-a-row design to use the toxicity outcomes from the *k* most recently treated patients to determine dose escalation and deescalation. Lin and Shih [8] studied statistical properties of general “A+B” designs. Leung and Wang [9] proposed an up-and-down design based on isotonic regression without making any parametric assumptions on the dose-toxicity curve. Stylianou and Flournoy [10] studied dose finding using the BCD and isotonic regression. Ivanova et al. [11] discussed several up-and-down designs that used more information from the most recently treated patients to make the dose assignment. Stylianou and Follmann [12] extended the BCD to handle delayed toxicities. Gezmu and Flournoy [13] developed the group up-and-down design in which the patients are treated in cohorts. Ivanova, Flournoy and Chung [14] proposed a cumulative cohort design to utilize the cumulative information at the current treating dose to make dose assignment. Ivanova and Kim [15] proposed a more general up-and-down design for dose finding with continuous and ordinal outcomes based on a *t*-statistic. Comprehensive reviews of dose-finding methods for phase I clinical trials have been provided by Chevret [16] and Ting [17].

Although most of the aforementioned up-and-down designs provide better operating characteristics than the “3+3” design, they have had limited use in practice. Rogatko et al. [18] reviewed 1,235 phase I cancer trials published between 1991 and 2006, and found that an overwhelming 98.4% of the clinical trials used the “3+3” design. A major impediment to the adoption of the improved up-and-down designs in practice is a lack of general guidance and a comprehensive assessment of the operating characteristics of these designs under practical clinical settings. Ivanova [19] provided an excellent review of up-and-down designs, but mainly from the methodological perspective.

To fill this gap, herein we review six up-and-down designs and conduct comprehensive simulation studies to evaluate their operating characteristics under various practical settings. In particular, we compare the performance of these designs to a theoretical optimal bound of nonparametric designs, which elucidates not only the relative but also the absolute performance of the designs. In addition, this comparison sheds light on the potential for improving the existing up-and-down designs.

The remainder of this article is organized as follows. In Section 2, we review six up-and-down designs and the nonparametric optimal design. In Section 3, we present comprehensive simulation studies to assess the operating characteristics of these designs under a wide range of practical settings. We conclude with a brief discussion in Section 4.

2 Methods

Let $(d_1, \dots, d_J)$ denote a set of J prespecified doses for the drug under investigation with corresponding toxicity probabilities $(p_1, \dots, p_J)$. Let φ denote the target toxicity rate specified by physicians, and Y_i denote the binary toxicity outcome for the i th subject with $Y_i = 1$ indicating that the patient experiences a DLT. We assume that a total of $n = m \times s$ patients will be treated in the trial with a cohort size of s , $s \in \{1, 2, \dots, n\}$, and the first cohort of patients receives the lowest dose d_1 . When $s = 1$, patients are assigned one at a time to a dose.

2.1 The “3+3” design

In the “3+3” design, patients are treated in cohorts of size $s = 3$. The dose escalation rule for the “3+3” design can be described as follows: among three patients treated at the current dose, if none experiences DLT, then the next cohort of three patients is treated at the next higher dose; if two or more patients experience DLT, then the next cohort of patients is treated at the next lower dose unless six patients have already been treated at that dose; and if one out of the three patients experiences DLT, then three more patients are treated at that same dose level. In general, if fewer than one of six patients treated at a dose level experiences DLT, then the next cohort is treated at the next higher dose; and if two or more of the six patients treated at a dose level experience DLT, then the MTD is considered to have been exceeded. The MTD is defined as the highest dose at which fewer than two out of six patients experience DLT. Although not explicitly defined in the design, the target toxicity rate of the “3+3” design is approximately $\varphi = 0.25$.

2.2 The accelerated titration design

The accelerated titration design (ATD) [4] is an extension of the “3+3” design that aims to speed up the trial and reduce the number of patients assigned to low doses by adding an accelerated dose-assignment phase. The ATD is conducted in two phases: the trial starts with the titration phase, in which one patient is treated per dose level until one patient exhibits DLT, and then the trial switches to the second phase, which is the traditional “3+3” design phase described previously. Similar to the “3+3” design, one limitation of the ATD is that it targets only a toxicity rate of $\varphi = 0.25$.

2.3 The biased coin design

Unlike the “3+3” design and the ATD, the biased coin design (BCD) is more flexible and can target any prespecified toxicity rate φ [5]. The BCD is based on the theory of random walk and assigns patients to a dose level one at a time, i.e., $s = 1$. Suppose that the i th patient is treated at dose level j . To determine a dose for the next patient,

- if $Y_i = 1$, we deescalate the dose level to $j - 1$;

- if $Y_i = 0$, we escalate the dose level to $j + 1$ with a probability of $\phi/(1 - \phi)$, otherwise we retain the current dose level j .

One drawback of the BCD is its low efficiency as it uses only the outcome of the most recently treated patient to determine the dose assignment, discarding the treatment information for all the other previously treated patients.

2.4 The k-in-a-row design

The k -in-a-row (KIR) design was proposed to address the low efficiency of the BCD by utilizing the treatment information from the last k ($k > 1$) patients treated, rather than only the last one patient treated, to make the decision of dose assignment [6, 7]. In the KIR design, patients are also assigned to a dose level one at a time with $s = 1$. Suppose that the i th patient is treated at dose level j . To determine a dose for the next patient,

- if $Y_i = 1$, we deescalate the dose level to $j - 1$;
- if $Y_i = Y_{i-1} = \dots = Y_{i-k+1} = 0$, that is, the k most recently treated patients were all treated at dose level j and none of them experienced toxicity, we escalate the dose level to $j + 1$; otherwise we retain the current dose level j .

Compared to the BCD, another advantage of the KIR design is that it does not require randomization for dose assignment, which simplifies the implementation of the trial design in practice. However, in terms of the toxicity probability that the design can target, the KIR design is less flexible. The BCD can target any arbitrary toxicity probability ϕ , whereas the KIR design can target only certain toxicity probabilities, given by $\psi = 1 - (0.5)^{1/k}$, $k = 1, 2, 3, \dots$. For example, the KIR design cannot exactly target toxicity probability $\phi = 0.25$ since $\psi = 0.293$ when $k = 2$ and $\psi = 0.206$ when $k = 3$. In practice, in order to target a prespecified toxicity rate ϕ , we can choose the value of k to minimize the absolute distance between ψ and ϕ , i.e., $k = \text{argmin}_{k \in \{1, \dots, n\}} |\psi - \phi|$.

2.5 Group up-and-down design

The BCD and the KIR design both assign patients one at a time to a dose level (i.e., $s = 1$). However, in practice, it is often more convenient to treat patients by cohorts (e.g., $s = 3$) for logistic reasons. The group up-and-down (GUD) design was devised to accommodate this practical issue [13]. The GUD design can be described as follows. Let c_L and c_U denote lower and higher cutoffs, respectively, which are integers between 0 and s with $0 < c_L < c_U < s$; and suppose that the most recently enrolled cohort of s patients were treated at dose level j and that m of them experienced toxicity. To determine a dose for the next cohort,

- if $m > c_U$, we deescalate the dose level to $j - 1$;
- if $m < c_L$, we escalate the dose level to $j + 1$;
- if $c_L < m < c_U$, we retain the current dose level j .

In the GUD design, the cutoffs c_L and c_U should be chosen so that the dose assignment clusters around the target toxicity probability ϕ . Gezmu and Flournoy [13] showed that, given values of c_L and c_U , the dose assignment of the GUD design asymptotically peaks around the dose level whose toxicity probability is ψ , where ψ is the solution of

$$pr \{Bin(s, \psi) \leq c_L\} = pr \{Bin(s, \psi) \geq c_U\} \quad (1)$$

with $Bin(s, \psi)$ denoting a binomial distribution with size parameter s and probability parameter ψ . Therefore, in order to target a prespecified toxicity probability ϕ , the values of cutoffs c_L and c_U should be chosen so that $\psi = \phi$ is the solution of equation (1). Such values of c_L and c_U , however, may not always exist because c_L and c_U can take only integer values between 0 and s with $0 \leq c_L < c_U \leq s$ and sometimes none of these values can lead to the solution of $\psi = \phi$. For example, given a cohort size of 2 (i.e., $s = 2$), c_L and c_U take values only from the set $\{0, 1, 2\}$, and the GUD design cannot target toxicity probabilities other than 0.293, 0.500 and 0.707. In practice, to target a toxicity probability ϕ , we can choose the values of c_L and c_U that minimize the difference between ψ and ϕ , that is,

$$(c_L, c_U) = \underset{0 \leq c_L < c_U \leq s}{\operatorname{argmin}} |\psi - \phi|. \quad (2)$$

As an example, given the cohort size $s = 3$, this approach leads to $(c_L, c_U) = (0, 1)$, $(0, 1)$ and $(0, 2)$ when $\phi = 0.2$, 0.25 and 0.3 , respectively.

2.6 Cumulative group up-and-down design

The cumulative group up-and-down (cGUD) design provides an extension of the GUD design by using the toxicity information from the cumulative number of subjects treated at the current dose to make the decision of dose escalation and de-escalation [14]. To determine a dose for a new cohort, the cGUD design uses the same decision rule as the GUD design, but takes the “cohort” size to be the cumulative number of subjects treated at the current dose. More precisely, let s^* be the number of patients that have been assigned to the current dose level j , and assume that m^* out of the s^* patients experience toxicity. The decision rule of the cGUD design is the same as that of the GUD design described previously, with s^* and m^* respectively substituting for s and m . Equivalently, this decision rule can be more conveniently expressed using the observed toxicity $\hat{p} \doteq m^*/s^*$ as follows,

- if $\hat{p} \geq \phi + \delta_U$, we deescalate the dose level to $j - 1$;
- if $\hat{p} \leq \phi - \delta_L$, we escalate the dose level to $j + 1$;
- if $\phi - \delta_L < \hat{p} < \phi + \delta_U$, we retain the current dose level j ,

where $\delta_L = \phi - c_L/s^*$ and $\delta_U = c_U/s^* - \phi$, and c_L and c_U are determined based on equations (1) and (2) with $s = s^*$. Similar to the GUD design, the cGUD design can target only certain toxicity probabilities determined by equation (1). In practice, in order to target a prespecified toxicity probability ϕ , we can set $\delta_L = \delta_U = \delta$ and choose the value of δ to minimize the distance between ψ and ϕ . The recommended value for δ is 0.09 for $\phi = 0.1, 0.15, 0.20$, or 0.25 and $\delta = 0.1$ for $\phi = 0.3$ or 0.35 .

2.7 Nonparametric optimal design

The nonparametric optimal (NPO) design proposed by O'Quigley, Paoletti and Maccario [20], while not applicable in a real trial, provides a theoretical upper bound for the performance of any nonparametric dose-finding method such as the up-and-down designs.

The NPO design is built upon a theoretical concept called “complete information.” Let $Y_i(j)$ denote the binary toxicity outcome for the i th patient when he/she is treated at dose level j . The complete information is defined as $\{Y_i(j), i = 1, \dots, n, j = 1, \dots, J\}$, i.e., a collection of the toxicity outcomes for each of the n patients treated at each of the J dose levels. The complete information represents the maximum information that we could possibly obtain from n patients regarding the rate of toxicity they experience when treated at J doses. Obviously, in a real trial, the complete information is not available and only one of $\{Y_i(j), j = 1, \dots, J\}$ is observable for the i th patient because in practice a patient receives only one of the J doses. Nevertheless, the theoretical concept of complete information can be used to derive the upper bound for the performance of nonparametric designs in simulation studies, where the complete information is available.

Based on the complete information, the NPO design estimates the toxicity probability of the j th dose level by

$$\hat{p}_j = \frac{\sum_{i=1}^n Y_i(j)}{n}, \quad j=1, \dots, J.$$

This estimate is optimal in the sense that it is unbiased and its variance attains the Cramer-Rao lower bound under the monotonicity assumption that a patient who experiences a DLT at a particular dose level would necessarily experience a DLT at any higher level (i.e., if $Y_i(j) = 1$, then $Y_i(j') = 1$ for $j' > j$); and similarly, if a patient tolerates a given dose level, he or she would also tolerate any lower dose level (i.e., if $Y_i(j) = 0$, then $Y_i(j') = 0$ for $j' < j$) [20]. Unless we are prepared to make some parametric assumptions on the dose-toxicity relationship, we cannot improve upon this estimator. Based on the $\hat{p}_j$ s, the MTD is selected as the dose level at which the estimate of the toxicity probability is closest to the target ϕ . Because $\hat{p}_j$ is optimal, the selection of the MTD under the NPO design is also optimal and provides the upper bound of the MTD selection for any nonparametric dose-finding method. Note that the NPO design is a theoretical design. It does not assign patients sequentially to the doses, as in a real trial, and thus cannot be used to assess the other aspects of a trial design, such as the number of patients treated at each dose level. To implement the NPO design, we simulate a random number u_i from a uniform distribution to represent the toxicity tolerance threshold for each subject i , and then determine $Y_i(j) = 0$ if $\sum_{k=1}^j p_k < u_i$ and 1 otherwise.

2.8 Dose selection based on isotonic regressions

We have described the dose escalation and deescalation rules for the BCD and the KIR, GUD and cGUD designs. When the phase I trial is completed, a method is needed to select the MTD based on the observed data. One such method is isotonic regression, which has been shown to yield a superior performance [9, 10]. Assume that at the end of the trial, at the dose level j , m_j out of n_j patients experience toxicity. The isotonic estimates of the toxicity probabilities, say $\hat{p}_j, j = 1, \dots, J$, are obtained by applying the pool adjacent violator algorithm [21] to the observed toxicity rate $p_j = m_j/n_j$. We then select as the MTD the dose j^* with an isotonic estimate of toxicity probability closest to ϕ , i.e.,

$$j^* = \operatorname{argmin}_{j \in (1, \dots, J)} |\tilde{p}_j - \phi|.$$

If there are ties to p_{j^*} , we select from the ties the highest dose level when $p_{j^*} < \phi$ and the lowest dose level when $p_{j^*} > \phi$.

To protect the patients in the trial, we also impose the following safety rule: if $\operatorname{pr}(p_j > \phi | \text{data}) > 0.95$, the dose level j and higher doses are eliminated from the trial and are not used to treat any subsequently enrolled patients; and the trial is terminated if the first dose level is eliminated. The posterior probability $\operatorname{pr}(p_j > \phi | m_j, n_j) > 0.95$ can be easily evaluated based on a beta-binomial model, assuming that m_j follows a binomial distribution (with size and probability parameters n_j and p_j) and p_j follows a vague beta prior, e.g., $p_j \sim \text{beta}(0.1, 0.1)$. This safety rule is important because in practice once there is adequate evidence to support that a dose is excessively toxic, no additional patients should receive that dose.

3 Simulation studies

3.1 Simulation descriptions

We compared the operating characteristics of the aforementioned up-and-down designs through extensive simulation studies. A literature review of 157 phase I clinical trials conducted between 1999 and 2005 showed that the median number of dose levels was 5 and the median sample size was 25 [22]. Following that, in our simulation, we assumed five dose levels and a maximum sample size of 9 cohorts of size 3 (i.e., 27 patients). We note that in principle the maximum sample size required by the “3+3” design could be 10 cohorts for 5 dose levels. However, it is very rare for a phase I trial using the “3+3” design to use up all 10 cohorts because the trial is typically terminated before reaching the maximum sample size. Therefore, we do not expect that using a sample size of 9 cohorts will have noticeable effects on the operating characteristics of the “3+3” design. The BCD and the KIR design cannot handle patients in cohorts, and thus for these two designs, patients are treated one by one.

Under the standard assumption that toxicity monotonically increases with the dose, the difficulty of dose finding is essentially controlled by the toxicity probability differences (i.e., distances) between the MTD and the two adjacent doses because these two doses are the most difficult ones to distinguish from the MTD. In our simulation, we considered four configurations of distances around the MTD: (0.07, 0.07), (0.15, 0.15), (0.07, 0.15) and (0.15, 0.07), where the first and second numbers within each set of parentheses are the toxicity probability distances between the MTD and the adjacent lower and higher doses. That is, the first and second configurations consider small and large distances on both sides of the MTD, respectively; and the third and fourth configurations consider the cases in which the distance is small on one side and large on the other side of the MTD. Note that controlling the distance around the MTD is equivalent to controlling the shape of the dose-toxicity curve around the MTD. For example, (0.07, 0.07) indicates that the dose-toxicity curve increases slowly both below and above the MTD, while (0.07, 0.15) indicates that the curve increases slowly below the MTD but rapidly above the MTD.

Under each of the four distance configurations, we varied the location of the MTD sequentially from the first to the last dose levels, resulting in a total of 16 toxicity scenarios. Table 1 shows the 16 toxicity scenarios when the target toxicity rate $\varphi = 0.25$. In particular, scenarios 1 to 5 correspond to (0.07, 0.07); scenarios 6-10 correspond to (0.15, 0.15); scenarios 11-13 correspond to (0.07, 0.15); and scenarios 14-16 correspond to (0.15, 0.07). Besides $\varphi = 0.25$, we considered two other target toxicity probabilities commonly encountered in oncology: $\varphi = 0.2$ and 0.3 . We constructed 16 toxicity scenarios for $\varphi = 0.2$ (or 0.3) by adding (or subtracting) 0.05 to (or from) the scenarios listed in Table 1. For example, under scenario 3, by subtracting 0.05 , we obtained the toxicity scenario (0.06, 0.13, 0.20, 0.27, 0.34) with the target toxicity probability of 0.2 located at dose level 3; and by adding 0.05 , we obtained the toxicity scenario (0.16, 0.23, 0.30, 0.37, 0.44) with the target toxicity probability of 0.3 located at dose level 3. Under each scenario, we conducted 10,000 simulations for each of the designs, including the NPO design.

3.2 Simulation results

Table 2 shows the simulation results under 16 toxicity scenarios with the target toxicity probability $\varphi = 0.25$, including four summary statistics of trial performance: the MTD selection percentage, the percentage of patients treated at the MTD, the percentage of patients treated above the MTD (i.e., overly toxic doses) and the selection percentage for doses above the MTD. The last two summary statistics measure the safety of the design.

To gauge the average performance of the designs across 16 toxicity scenarios, for each of the summary statistics, we reported two measures: average rank and average loss, defined as follows. Let t_{ij} denote the value of a summary statistic t for the i th design under the j th toxicity scenario, and r_{ij} denote the rank of t_{ij} among $\{t_{ij}, i = 1, \dots, 6\}$ under the j th toxicity scenario, with $r_{ij} = 1$ representing the best performance (e.g., the highest MTD selection percentage, or the lowest selection percentage for doses above the MTD). The average rank of the j th design for summary statistic t across 16 toxicity scenarios is defined as

$$\text{Average rank of the } j\text{th design} = \sum_{i=1}^{16} r_{ij} / 16.$$

Letting t_{ij}^* denote the value of t_{ij} associated with $r_{ij} = 1$, we define the average loss of the j th design for summary statistic t as

$$\text{Average loss of the } j\text{th design} = \sum_{i=1}^{16} |t_{ij}^* - t_{ij}| / 16,$$

which represents the average loss in t under the j th design when compared to the best performer in each scenario. Because the nonparametric optimal design is a theoretical device for providing the upper bound of the MTD selection, it was not included in the calculation of the average rank and loss. We depict a comparison of the average rank and average loss of each design in Figures 1 and 2. Tables 3 and 4 show the simulation results when the target toxicity probability $\varphi = 0.2$ and 0.3 , respectively.

3.2.1 MTD selection percentage—Among the six up-and-down designs in this comparison, the “3+3” design and the ATD performed the worst, with MTD selection percentages substantially lower than those of the other designs, although the ATD performed slightly better than the “3+3” design. For example, when the target toxicity rate was 0.25, in 15 out of 16 scenarios, these two designs led to the two lowest MTD selection percentages and their average loss in the MTD selection ranged from 11% to 25%. The BCD and the KIR design outperformed the “3+3” design and the ATD, with 10-20% higher MTD selection percentages in most scenarios. Their average loss in the MTD selection was about 6% when the target toxicity ϕ was 0.25 and 0.3, and about 9% when $\phi = 0.2$.

The GUD and cGUD designs had the best performances and dominated the other designs in selecting the MTD. For example, out of the 16 toxicity scenarios, the GUD and cGUD designs yielded the highest MTD selection percentages in 5 and 9 scenarios, respectively, when the target toxicity rate was 0.25 (Table 2), and yielded the highest selection percentages in 9 and 7 scenarios, respectively, when the target toxicity rate was 0.3 (Table 4). In addition, the average loss in the MTD selection for these two designs was typically less than 3%. The relative performance of the GUD and cGUD designs varied with the target toxicity rate ϕ . It appears that the cGUD design performed slightly better when $\phi = 0.25$; whereas the GUD design had a small performance edge when $\phi = 0.3$.

The absolute performances of the GUD and cGUD designs were also impressive (see Figure 3). When the distances around the MTD were (0.07, 0.07), i.e., scenarios 1 to 4, or (0.15, 0.07), i.e., scenarios 14-16, the selection percentages of the GUD and cGUD designs were rather close to the upper bound obtained under the nonparametric optimal design; and when the distances around the MTD were (0.15, 0.15), i.e., scenarios 6-10, or (0.07, 0.15), i.e., scenarios 11-13, the MTD selection percentages of the GUD and cGUD designs were about 10 to 15% lower than the upper bound. These results suggest that, under the nonparametric framework, substantial improvement over the GUD and cGUD designs may be limited. If we want to improve these two designs, we should focus on the case in which the distances around the MTD are large and the case in which the distances below and above the MTD are respectively small and large. These two cases correspond to the dose-toxicity curve that increases rapidly around the MTD and the dose-toxicity curve that increases slowly below the MTD then rapidly above the MTD. Note that occasionally, for example in scenario 2 with $\phi = 0.25$, the GUD design may yield an MTD selection percentage that is higher than that of the nonparametric optimal design. This is because the nonparametric optimal design provides an upper bound only on the average performance and does not guarantee that the design always yields the highest selection percentage in each individual scenario.

3.2.2 Percentage of patients treated at the MTD—In terms of the percentage of patients treated at the MTD, the cGUD and GUD designs again were generally superior to the other designs. These two designs were often the best or the second best performers across the simulation scenarios. For example, when the target rate was 0.25, the cGUD design assigned the highest or the second highest percentage of patients to the MTD in 14 out of 16 scenarios, with an average ranking of 1.6; and when the target rate was 0.3, the GUD design assigned the highest or the second highest percentage of patients to the MTD in 13 out of 16 scenarios, with an average ranking of 2. Between the GUD and cGUD designs,

the cGUD design was the winner. Compared to the GUD design, the average loss of the cGUD design in the percentage of patients treated at the MTD was 5.5%, 2.6% less when $\varphi = 0.25$ and 0.2, and comparable when $\varphi = 0.3$. The performances of the other four designs ("3+3", ATD, BCD and KIR) were generally comparable in terms of assigning patients to the MTD, and their average losses in the percentage of patients treated at the MTD were between 5-11%.

3.2.3 Design safety—We considered two safety measures for the trial designs: the percentage of patients treated above the MTD (or at overly toxic doses) and the selection percentage for doses above the MTD. The first measure focuses on the safety of the patients treated in the current phase I trial; and the second measure gauges the safety of the patients who will be treated in future trials (e.g., subsequent phase II or III trials) based on the dose selected in the phase I trial.

In terms of protecting the patients enrolled in the current trial, the ATD performed the worst, with the highest percentage of patients receiving doses above the MTD. This may be due to the accelerated dose-assignment strategy of the ATD. Overall, the cGUD design performed the best, with average losses below 2.5% across different target toxicity rates. The other designs demonstrated some variations for different target toxicity rates. For example, the KIR design performed extremely well when $\varphi = 0.25$, but not that well when $\varphi = 0.3$; and the GUD design had an average loss of 2.1% when $\varphi = 0.25$, but more than 6% when $\varphi = 0.2$ or 0.3.

In terms of the selection percentage for doses above the MTD, that is, the safety of future patients, the "3+3" and cGUD designs appear to have performed better than the other designs. However, it is worth noting that all six up-and-down designs were generally safer than the nonparametric optimal design, with lower selection percentages of the doses above the MTD. This indicates that the six designs were actually overly conservative and the likelihood of selecting doses above the MTD may not be of concern for these six up-and-down designs.

In summary, among the six up-and-down designs in this comparison, the cGUD design yields the best overall performance, with the highest MTD selection percentage and the highest percentage of patients treated at the MTD. The cGUD design is also generally safer than the other designs in that it treated a lower percentage of patients at doses above the MTD. Although the cGUD design is generally favorable, the GUD design is a very competitive alternative when the target toxicity probability is 0.3. The comparison with the theoretical upper bound of nonparametric designs reveals that although there is some space for improvement, dramatic improvements to the cGUD design may not be possible within the nonparametric framework.

The up-and-down designs reviewed here are easy to use in practice as they are all based on prespecified fixed dose-transition rules. Once we determine these rules based on design parameters (i.e., the target toxicity probability, cohort and sample size), during the trial conduct, we only need to count the number of observed DLTs and make the decision of dose escalation/deescalation based on the rule. All the six up-and-down designs assume that

toxicity is binary and quickly ascertainable. From the viewpoint of trial and accrual planning, the cGUD and GUD designs are convenient because they prespecify the sample size, which can be done through calibrating the operating characteristics. Although the BCD also prespecifies the sample size, the implementation of that design requires randomization, which can be a logistic burden in practice. Comparatively, the sample sizes for the “3+3” design and the ATD are random. In some cases, that can be less convenient for planning the trial and accrual.

4 Conclusions

We have reviewed six up-and-down phase I clinical trial designs and conducted a comprehensive comparative study to evaluate their relative performances. We have also compared these designs to the theoretical upper bound of nonparametric designs to examine their relative and absolute performances. Our study shows that the cGUD design has the overall best performance: it is safe, yields a high MTD selection percentage and assigns a high percentage of patients to the MTD. The comparison with the theoretical upper bound indicates that the performance of the cGUD design is generally close to optimal, and under the nonparametric framework, substantial improvements over the cGUD design may be limited. Nevertheless, in some cases, for example when the dose-toxicity curve increases rapidly or increases slowly then rapidly, there seems to be some potential for further improving the cGUD design. In these cases, the cGUD design seems overly conservative and tends to select the doses below the MTD, compared to the optimal NPO design. One possible approach to addressing this issue is to use asymmetric cutoffs c_L and c_U (e.g., decreasing c_U or/and increasing c_L) in the cGUD design to encourage more aggressive dose escalations. Spreading patient data across doses will increase the power of distinguishing the MTD from its near neighbors, and thus improve the selection percentage of the MTD.

We have focused on the up-and-down designs, but note that a rich class of model-based designs also exists. The model-based designs may be more efficient than the nonparametric up-and-down designs because they use a parametric dose-response model to borrow information across dose levels. However, the model-based designs are often more complicated to implement due to complex model fitting and are also more sensitive to model misspecifications. Although we do not show the results, in preliminary comparisons of the up-and-down designs and model-based designs, such as the continual reassessment method [23], the efficiency loss when using good up-and-down designs is actually minor. Taking into account the robustness and simplicity of implementation, the up-and-down designs are competitive designs worthy of consideration in practice, and worthy of using in place of the “3+3” design.

References

1. Storer BE. Design and analysis of phase I clinical trials. *Biometrics*. 1989; 45:925–937. [PubMed: 2790129]
2. Korn EL, Midthune D, Chen TT, Rubinstein LV, Christian MC, Simon RM. A comparison of two phase I trial designs. *Statistics in Medicine*. 1994; 13:1799–1806. [PubMed: 7997713]
3. Ahn C. An evaluation of phase I cancer clinical trial designs. *Statistics in Medicine*. 1998; 17:1537–1549. [PubMed: 9699228]

4. Simon R, Friedlin B, Rubinstein LV, Arbuck SG, Collins J, Christian MC. Accelerated titration designs for phase I clinical trials in oncology. *Journal of the National Cancer Institute*. 1997; 89:1138–1147. [PubMed: 9262252]
5. Durham, SD.; Flournoy, N. Random walks for quantile estimation.. In: Berger, JO.; Gupta, SS., editors. *Statistical Decision Theory and Related Topics*. V. Springer; New York: 1994. p. 467476
6. Wetherill G. Sequential estimation of quantal response curves, 1963. *Journal of Royal Statistical Society Series B*. 25:1–48.
7. Gezmu, M. Dissertation. American University; 1996. The Geometric Up-and-Down Design for Allocating Dosage Levels..
8. Lin Y, Shih WJ. Statistical properties of the traditional algorithm-based designs for phase I cancer clinical trials. *Biostatistics*. 2001; 2:203–215. [PubMed: 12933550]
9. Leung D, Wang YG. Isotonic designs for phase I trials. *Controlled Clinical Trials*. 2001; 22:126–138. [PubMed: 11306151]
10. Stylianou M, Flournoy N. Dose finding using the biased coin up-and-down design and isotonic regression. *Biometrics*. 2002; 58:171–177. [PubMed: 11890313]
11. Stylianou M, Montazer-Haghighi A, Mohanty S, Durham S. Improved up-and-down designs for phase I trials. *Statistics in Medicine*. 2003; 22:69–82. [PubMed: 12486752]
12. Stylianou M, Follmann N. The accelerated biased coin up-and-down design in phase I trials. *Journal of Biopharmaceutical Statistics*. 2004; 14:249–260. [PubMed: 15027512]
13. Gezmu M, Flournoy N. Group up-and-down designs for dose-finding. *Journal of Statistical Planning and Inference*. 2006; 136:1749–1764.
14. Ivanova A, Flournoy N, Chung Y. Cumulative cohort design for dose finding. *Journal of Statistical Planning and Inference*. 2007; 137:2316–2317.
15. Ivanova A, Kim S. Dose-finding for binary ordinal and continuous outcomes with monotone objective function: a unified approach. *Biometrics*. 2009; 65:307–315. [PubMed: 18479486]
16. Chevret, S. *Statistical Methods for Dose-Finding Experiments*. John Wiley and Sons Ltd; England: 2006.
17. Ting, N. *Dose Finding in Drug Development*. Springer; Cambridge: 2006.
18. Rogatko A, Schoeneck D, Jonas W, Tighiouart M, Khuri FR, Porter A. Translation of innovative designs into phase I trials. *Journal of Clinical Oncology*. 2007; 25:4982–4986. [PubMed: 17971597]
19. Ivanova, A. *Wiley Encyclopedia of Clinical Trials*. John Wiley & Son; New York: 2007. Escalation and up-and-down designs..
20. O'Quigley J, Paoletti X, Maccario J. Non-parametric optimal design in dose finding studies. *Biostatistics*. 2002; 3:51–56. [PubMed: 12933623]
21. Barlow, RE.; Bartholomew, DJ.; Bremner, JM.; Brunk, HD. *Statistical Inference under Order Restrictions*. John Wiley & Sons; London, New York: 1972.
22. Zohar S, Lian Q, Levy V, Cheung K, Ivanova A, Chevret S. Quality assessment of phase I dose-finding cancer trials: proposal of a checklist. *Clinical Trials*. 2008; 2008; 5:478–485. [PubMed: 18827040]
23. O'Quigley J, Pepe M, Fisher L. Continual reassessment method: a practical design for phase I clinical trials in cancer. *Biometrics*. 1990; 46:33–48. [PubMed: 2350571]

Figure 1.

Average rank for the six designs under different target toxicity rates.

Figure 2.

Average loss for the six designs under different target toxicity rates.

Figure 3. MTD selection percentages under the group up-and-down (GUD) design (denoted by circles) and cumulative group up-and-down (cGUD) design (denoted by triangles). The line segments denote the upper bound.

Table 1

Sixteen toxicity scenarios for the simulation study. The target toxicity rates are in boldface.

Dose level	Scenario															
	1	2	3	4	5	6	7	8	9	10	11	12	13	14	15	16
1	.25	.18	.11	.04	.01	.25	.10	.01	.01	.01	.18	.10	.06	.10	.01	.01
2	.32	.25	.18	.11	.04	.40	.25	.10	.05	.03	.25	.18	.12	.25	.10	.05
3	.39	.32	.25	.18	.11	.55	.40	.25	.10	.06	.40	.25	.18	.32	.25	.10
4	.46	.39	.32	.25	.18	.70	.55	.40	.25	.10	.55	.40	.25	.40	.32	.25
5	.53	.46	.39	.32	.25	.80	.70	.55	.10	.25	.70	.55	.40	.48	.40	.32

Table 2

Simulation results with a target toxicity rate of $\phi = 0.25$.

Summary statistics		Design	Scenario															Average rank		Average loss	
			1	2	3	4	5	6	7	8	9	10	11	12	13	14	15				
Selection % of MTD	3+3	32.1	24.1	21.2	21.3	34.2	41.0	37.3	36.2	35.9	45.9	30.6	27.4	25.5	29.5	28.8	29.6	5.8	18.0		
	ATD	26.9	25.4	23.5	26.6	43.5	39.0	37.6	37.3	37.7	48.9	34.9	32.7	32.8	28.0	27.7	30.4	5.1	16.0		
	BCD	42.4	33.0	30.1	32.3	45.6	59.4	54.3	50.1	49.8	60.4	43.3	38.6	36.4	41.1	40.8	41.9	3.2	5.6		
	KIR	44.6	34.4	31.0	31.3	42.9	58.1	55.4	48.8	50.9	59.6	43.3	37.6	35.1	44.2	39.9	44.5	3.3	5.4		
	GUD	49.4	38.7	34.4	29.3	46.4	64.9	60.9	55.1	51.3	68.4	47.5	39.9	33.6	50.1	47.0	44.0	1.9	1.7		
	cGUD	54.9	36.8	33.6	29.0	48.4	68.1	60.6	57.4	53.7	71.2	45.0	40.4	32.5	50.0	48.1	45.8	1.7	0.8		
	Optimal	57.0	34.5	33.5	33.5	71.5	80.6	69.0	69.0	69.0	88.0	53.8	52.9	52.6	47.4	47.4	47.4				
	3+3	51.0	30.9	22.5	18.8	19.0	58.0	36.8	29.2	24.7	24.4	33.6	24.1	19.0	33.7	27.6	23.9	3.4	8.7		
	ATD	31.0	26.2	24.0	25.2	34.4	39.6	31.9	30.2	31.4	40.3	31.2	27.6	26.9	27.0	26.3	29.1	3.4	8.4		
	BCD	52.9	29.7	21.6	18.5	19.9	64.1	34.5	27.5	24.2	25.2	33.5	23.6	19.1	30.5	25.5	23.0	4.3	9.0		
% of patients at MTD	KIR	63.8	30.7	20.2	15.8	15.0	71.9	35.9	28.2	23.5	22.0	32.9	21.5	15.6	33.4	26.5	22.8	4.8	8.6		
	GUD	60.9	31.8	21.7	15.7	16.8	67.3	37.8	32.0	23.8	25.0	33.6	22.9	15.4	35.7	30.4	23.3	3.6	7.7		
	cGUD	65.2	33.3	23.5	18.7	22.3	72.6	47.6	40.2	32.1	34.0	36.4	26.0	18.8	43.5	37.2	30.3	1.6	2.2		
	3+3	49.0	29.2	19.6	11.6	0.0	42.0	26.6	20.6	13.9	0.0	24.5	17.0	10.8	31.5	23.6	14.9	4.8	8.0		
	ATD	69.0	50.4	37.4	24.0	0.0	60.4	44.9	38.1	25.6	0.0	43.4	33.3	21.5	51.7	42.9	28.1	5.7	22.7		
	BCD	47.1	28.3	18.8	11.4	0.0	35.9	22.4	17.7	12.2	0.0	21.8	15.1	9.8	29.0	22.2	14.5	3.8	6.2		
	KIR	36.2	18.3	10.8	6.6	0.0	28.1	15.5	12.0	8.4	0.0	14.2	8.9	5.5	20.0	15.0	9.8	1.4	0.1		
	GUD	39.1	20.2	11.9	7.8	0.0	32.7	19.4	14.2	10.9	0.0	17.2	10.5	6.7	22.7	16.4	11.8	2.4	2.1		
	cGUD	34.8	19.2	12.9	9.5	0.0	27.4	19.8	17.1	13.2	0.0	15.9	11.4	7.7	23.9	20.2	15.0	2.9	2.5		
	Selection % of doses above MTD	3+3	23.7	17.4	15.7	14.7	0.0	13.9	12.5	12.6	12.2	0.0	10.4	9.4	8.7	21.0	21.0	19.4	1.3	0.0	
ATD		39.3	33.0	29.9	24.2	0.0	21.9	19.7	19.9	17.4	0.0	17.9	16.8	14.2	34.8	34.6	28.2	4.3	8.7		
BCD		37.9	33.1	29.2	22.2	0.0	19.4	21.0	21.7	18.2	0.0	19.3	18.8	15.5	35.7	33.2	27.5	4.9	8.8		
KIR		33.4	28.3	23.9	18.8	0.0	19.0	20.0	20.9	18.1	0.0	17.9	16.8	13.1	31.8	30.4	24.8	3.3	6.5		
GUD		38.6	29.7	22.6	19.6	0.0	22.7	21.4	20.0	20.0	0.0	19.6	17.0	13.4	32.8	28.4	27.4	4.1	7.5		
cGUD		32.3	24.5	20.4	19.8	0.0	18.9	19.7	20.4	20.7	0.0	16.0	14.8	13.8	30.7	29.9	28.8	3.1	6.1		
Optimal		43.0	38.1	37.9	37.9	0.0	19.4	19.0	19.0	19.0	0.0	18.8	18.8	18.8	40.6	40.6	40.6				

Table 3

Simulation results with a target toxicity rate of $\phi = 0.2$.

Summary statistics	Design	Scenario															Average rank	Average loss	
		1	2	3	4	5	6	7	8	9	10	11	12	13	14	15			16
Selection % of MTD	3+3	32.6	27.0	26.1	27.7	53.7	44.2	43.6	43.2	43.4	62.9	37.7	36.0	36.2	31.5	31.0	32.8	4.8	11.2
	ATD	26.0	24.7	25.2	29.5	59.2	41.0	40.9	41.9	44.3	64.7	38.9	38.6	39.6	26.2	28.1	33.0	4.4	11.7
	BCD	41.7	30.7	26.7	26.6	35.1	53.5	49.9	47.5	45.6	51.4	38.0	33.8	28.8	41.0	39.8	39.2	4.1	10.0
	KIR	41.1	29.4	26.2	28.3	40.7	53.4	49.4	47.4	49.5	58.6	37.9	34.3	34.1	38.6	38.3	41.2	4.1	8.8
	GUD	45.5	33.5	33.1	29.7	55.5	61.9	55.6	54.6	53.3	73.4	44.5	42.5	36.7	42.1	43.3	42.8	1.7	2.5
% of patients at MTD	cGUD	53.4	33.7	30.2	26.9	45.2	64.7	59.6	56.5	54.2	69.3	41.4	36.8	31.0	49.9	48.1	46.4	2.0	2.6
	Optimal	61.5	37.8	37.2	37.2	70.7	83.9	74.2	74.2	74.2	90.0	55.9	55.4	55.0	53.6	53.6	53.6		
	3+3	43.0	29.6	23.9	21.6	24.2	50.7	35.6	29.4	26.6	28.8	33.4	26.1	22.5	31.6	27.1	25.3	3.3	9.5
	ATD	25.6	23.3	23.2	26.5	41.9	34.5	29.5	29.4	32.7	47.5	29.0	27.6	29.6	23.6	24.8	29.1	3.5	9.6
	BCD	63.3	28.6	18.3	14.5	13.9	73.5	32.5	25.3	20.8	19.2	31.6	20.2	14.7	29.6	23.1	20.0	5.4	11.4
% of patients above MTD	KIR	64.2	29.5	20.2	16.9	17.9	74.0	35.1	28.5	25.2	25.6	32.6	22.4	17.4	31.7	26.1	23.8	4.3	8.8
	GUD	58.0	31.9	24.0	19.9	23.2	67.6	38.0	33.3	28.5	31.8	35.1	26.4	20.3	34.3	30.6	27.1	2.4	6.3
	cGUD	69.0	31.0	22.1	18.0	21.8	76.9	45.6	37.8	31.8	33.9	34.6	25.1	18.9	41.0	34.5	29.5	2.2	3.7
	3+3	57.0	38.3	27.2	16.8	0.0	49.3	34.5	27.5	18.7	0.0	32.3	24.2	15.8	40.3	31.1	19.9	4.8	14.9
	ATD	74.4	58.2	45.6	31.7	0.0	65.5	52.8	45.1	32.2	0.0	51.3	41.4	28.5	60.2	50.2	35.5	5.7	29.9
Selection % of doses above MTD	BCD	36.7	19.1	11.8	7.0	0.0	26.5	14.3	10.8	7.7	0.0	13.7	8.7	5.4	19.5	14.1	9.7	1.8	0.7
	KIR	35.8	19.0	12.8	8.5	0.0	26.0	14.8	11.8	9.4	0.0	13.9	9.7	6.6	20.2	15.7	11.8	2.3	1.4
	GUD	42.0	24.9	17.1	11.9	0.0	32.4	21.7	17.4	14.2	0.0	19.9	14.2	10.3	27.1	21.3	16.2	3.9	6.0
	cGUD	31.0	17.6	12.3	9.5	0.0	23.1	16.5	13.7	11.8	0.0	13.6	9.9	7.7	21.3	17.2	14.3	2.5	1.6
	3+3	35.9	30.8	29.4	27.3	0.0	23.1	22.1	22.6	22.1	0.0	19.3	19.2	17.9	34.7	35.2	34.2	4.5	8.7
Selection % of doses above MTD	ATD	52.1	47.4	44.4	38.2	0.0	31.6	30.6	29.7	26.2	0.0	28.6	27.5	24.8	50.3	47.8	41.1	5.7	17.8
	BCD	27.7	22.3	19.0	15.3	0.0	14.4	15.2	14.3	13.3	0.0	12.5	11.7	10.2	25.6	23.1	20.9	1.9	0.7
	KIR	26.9	23.2	21.8	18.0	0.0	13.9	15.5	16.0	13.7	0.0	13.5	14.0	11.1	26.9	25.7	22.4	2.8	1.7
	GUD	34.4	32.1	28.8	27.0	0.0	17.5	21.9	20.9	21.4	0.0	19.6	19.8	18.9	36.1	32.8	32.2	4.3	8.0
	cGUD	24.8	19.9	17.6	17.9	0.0	13.1	14.6	14.6	16.3	0.0	11.8	11.5	12.4	24.6	23.4	24.4	1.9	0.7
Optimal	38.5	33.6	33.4	33.4	0.0	16.1	15.8	15.8	15.8	0.0	15.4	15.4	15.4	36.5	36.5	36.5	36.5		

Table 4

Simulation results with a target toxicity rate of $\phi = 0.3$.

Summary statistics	Design	Scenario																Average rank	Average loss
		1	2	3	4	5	6	7	8	9	10	11	12	13	14	15	16		
Selection % of MTD	3+3	30.3	19.1	14.4	13.7	18.6	34.7	28.6	26.8	25.0	26.6	22.6	17.7	15.5	23.9	22.3	21.4	5.8	25.1
	ATD	26.8	23.1	19.9	21.0	28.5	34.4	31.9	31.5	29.0	34.4	29.3	25.4	24.3	25.4	24.8	25.1	5.1	20.5
	BCD	42.0	31.3	27.7	30.0	45.5	57.5	51.1	47.6	47.2	60.4	40.6	37.1	34.6	39.5	38.8	39.8	3.2	5.7
	KIR	41.6	30.7	27.8	29.2	48.5	57.8	50.9	48.6	48.2	63.1	40.8	36.7	34.9	38.2	38.0	39.9	3.2	5.5
	GUD	43.2	33.3	33.1	33.2	51.8	60.1	53.9	53.2	51.3	68.9	45.1	42.8	39.5	40.1	41.6	42.5	1.4	1.8
	cGUD	53.5	34.0	30.2	26.9	39.8	65.0	55.3	54.3	48.7	59.7	41.4	36.8	30.1	46.2	46.0	41.8	2.1	3.3
% of patients at MTD	Optimal	54.0	32.2	31.0	31.0	71.3	78.0	65.1	65.1	65.1	86.8	51.4	50.2	49.8	43.8	43.8	43.8		
	3+3	59.4	31.0	20.0	15.0	13.4	65.0	36.6	27.0	21.0	18.6	32.8	20.8	14.9	34.4	25.8	20.5	4.8	7.7
	ATD	36.5	28.2	23.7	22.7	26.1	45.0	33.3	29.7	28.6	31.8	32.3	26.5	23.7	29.1	26.7	27.0	3.4	6.7
	BCD	52.0	29.0	21.4	18.8	20.4	63.1	34.0	27.1	24.2	26.1	32.7	23.7	19.2	30.0	25.0	23.3	4.8	6.8
	KIR	54.9	30.0	22.0	19.3	21.7	65.5	35.3	29.1	26.1	28.3	33.4	23.9	19.8	31.5	26.6	24.8	3.3	5.4
	GUD	53.5	33.9	25.4	20.5	21.7	62.7	41.3	35.0	28.9	30.1	37.8	27.8	21.0	36.9	32.0	27.3	2.0	2.7
% of patients above MTD	cGUD	67.4	31.3	21.5	16.6	18.4	74.3	43.5	35.6	27.1	27.7	34.5	23.8	16.6	39.6	32.9	25.5	2.7	2.6
	3+3	40.6	21.3	12.6	7.2	0.0	35.0	19.7	14.1	9.4	0.0	17.7	11.0	6.7	23.6	16.5	10.2	1.9	1.8
	ATD	63.5	42.9	29.7	18.0	0.0	55.0	37.9	30.2	19.4	0.0	36.7	26.1	16.3	44.2	34.4	21.4	5.7	16.2
	BCD	48.0	29.4	20.3	12.4	0.0	36.9	23.3	18.7	12.8	0.0	22.7	16.1	10.3	29.9	23.1	15.2	4.3	6.4
	KIR	45.1	27.2	18.8	12.0	0.0	34.5	21.9	17.9	13.2	0.0	21.0	15.1	10.0	28.4	22.4	15.4	3.4	5.4
	GUD	46.5	27.0	17.5	11.1	0.0	37.3	25.1	20.0	13.8	0.0	22.2	15.1	9.4	30.5	23.5	15.7	4.1	6.1
Selection % of doses above MTD	cGUD	32.6	17.8	11.8	8.5	0.0	25.7	18.2	15.0	11.1	0.0	14.4	10.0	7.0	22.3	17.8	12.7	1.7	0.5
	3+3	14.1	9.3	7.2	6.2	0.0	8.0	6.5	6.4	5.3	0.0	5.2	4.2	3.2	11.9	11.3	9.7	1.3	0.0
	ATD	27.7	21.2	18.2	13.6	0.0	13.6	11.7	11.0	9.7	0.0	10.7	9.2	7.8	23.2	22.3	16.9	2.2	6.8
	BCD	37.2	32.3	30.2	23.6	0.0	20.5	21.9	22.1	18.8	0.0	19.4	18.9	16.3	35.0	33.6	28.0	4.1	15.6
	KIR	37.2	32.8	29.9	24.9	0.0	19.8	22.1	22.4	19.8	0.0	20.1	19.9	16.8	35.9	34.5	29.2	4.7	16.1
	GUD	40.4	35.2	31.0	25.5	0.0	22.9	24.8	24.7	23.1	0.0	22.2	21.2	17.3	39.3	36.8	32.2	5.7	18.0
Optimal	cGUD	28.9	21.6	18.3	16.5	0.0	17.1	18.4	18.0	17.0	0.0	13.8	12.4	11.2	27.8	26.5	23.9	3.1	10.2
	Optimal	46.0	40.6	40.3	40.3	0.0	22.0	21.6	21.6	21.6	0.0	21.4	21.3	21.3	42.9	42.9	42.9		