Title: Extension of Bayesian Ordered Lattice Design to Backfills for Phase I Clinical Trials

Authors:

Gi-Ming Wang, Case Comprehensive Cancer Center, Case Western Reserve University. (Postal address: 2103 Cornell Rd, Cleveland, OH 44106; Email: gxw174@case.edu)

Curtis Tatsuoka (Corresponding author), Department of Epidemiology and Public Health, Division of Biostatistics and Bioinformatics, University of Maryland-Baltimore (Postal address: 660 W. Redwood Street, HH109, Baltimore, MD 21201; Email: ctatsuoka@som.umaryland.edu)

Abstract:

This research builds upon the Bayesian Ordered Lattice Design (BOLD), integrating its applications to address a challenge encountered in Phase I clinical trials. It is widely agreed that the conventional Phase I trial designs are not consistently effective in determining safe and active dose levels, and the US FDA launched the Project Optimus, aimed at reforming the paradigms of dose optimization and selection. We propose a backfilled BOLD design (BF-BOLD) that employs the algorithm of BOLD for activity evaluation alongside the evaluation of toxicity within the same timeframe. Backfill strategies are proposed within the BF-BOLD framework to offer clinicians greater flexibility during the backfilling process. Our method for determining the Optimal Biological Dose (OBD) by balancing the toxicity activity against the identified Maximum Tolerated Dose (MTD) is straightforward and does not require a complex statistical model. Simulation result shows that conducting toxicity trials with backfills enhances the accuracy of MTD identification and reduces the incidence of overdoses compared to trials without backfills. It also demonstrates that the performance of BF-BOLD is superior to that of Backfilled Bayesian Optimal Interval Design (BF-BOIN) in accurately identifying the MTD.

Key words: Early Phase I clinical trial, Bayesian design, order, backfill, optimal dose.

Acknowledgements: This work was supported in part by NCI grants P30 CA134274-10 (University of Maryland, Baltimore) and P30 CA043703-34 (Case Comprehensive Cancer Center).

1. Introduction

Phase I clinical trials aim to establish the maximum tolerable dose (MTD), defined as the dose that presents a likelihood of dose-limiting toxicity (DLT) closest to a specified target rate.[1] The design consists of two main components. The first component includes a series of criteria for deciding how to assign dose levels to each participant. The second component involves a procedure for identifying the MTD at the conclusion of the trial.[2] From both a practical and ethical standpoint, it is essential to limit decision-making errors and decrease the likelihood of subjecting patients to either insufficiently therapeutic or overly toxic dosages.[1]

Numerous Phase I trial designs have been developed to identify the MTD for single-agent studies, and one of which is the Bayesian Ordered Lattice Design (BOLD) established by Wang and Tatsuoka (2025).[3] The term "lattice" originates from the idea that the dose levels lie within a formal order structure, while the dose levels in single-drug trials can be perceived as a linearly ordered lattice.[4] Web Figure 1 of the Supplementary Material depicts the lattice diagram that encompasses 5 dose levels, which also features both top and bottom states. The bottom state indicates that the MTD is less than the minimum dose. Conversely, the top state indicates that the MTD exceeds the maximum dose. Classification is associated with a specific state within the lattice, while a designated dose level signifies that the dose corresponds to the MTD. Relevant research on Bayesian sequential classification within lattice models includes works by Tatsuoka and Ferguson (2003),[5] Tatsuoka (2014),[6] and Tatsuoka, Chen, and Lu (2022).[7]

The BOLD design combines prior knowledge and ordering constraints with empirical toxicity data through a Bayesian framework to facilitate adaptive dose selection, early trial termination, and the classification of the MTD in Phase I clinical trials.[3] The primary characteristic of BOLD lies in its integration of posterior information and the implementation of order constraints at each stage of decision-making, facilitating the sharing of information across different doses. It represents an innovative and rapid computational method that showcases distinct conceptual and performance benefits in Bayesian dose-finding trials compared to well-established designs like the Bayesian Optimal Interval Design (BOIN), Continual Reassessment Method (CRM), and the

conventional 3+3 design.[3] The performance of BOLD for Phase I clinical trials can be evaluated via simulations conducted with R (code available at https://github.com/hiddenmanna1996/BOLD).

The aim of this study is to expand the application of BOLD design, enabling its adaptation to tackle an existing challenge of dose optimization. The medical research community has recognized that traditional approaches for determining the MTD might not be sufficient, and it is widely acknowledged that standard Phase I trial designs fail to consistently determine safe and active doses.[8] When the conditions of a trial permit an evaluation of treatment activity within the same period necessary to assess toxicity, it is reasonable to take both toxicity and activity into account when determining the appropriate dose for subsequent investigation.[2] The US FDA also launched the Project Optimus, aimed at reforming the paradigms of dose optimization and selection, incorporating the idea of backfilling. To create backfill strategies that employ an algorithm akin to BOLD, with the goal of identifying the Optimal Biological Dose (OBD), we propose the BF-BOLD (Backfilled BOLD) design to fulfill this aim.

2. Methodology of BOLD

Notation and Terminology

Suppose we have $J > 1$ doses. Define DLT rate for dose j as π_j, $1 \leq j \leq J$. We assume the natural order constraints $\pi_j \leq \pi_k$ if $j < k$. We also define the *cover* dose for dose j to be dose $j+1, j < J$, and the *anti-cover* to be dose $j-1, j > 1$. Let ϕ be the target DLT rate. N_{max} is the maximum number of overall patients in the trial before stopping is invoked; n_j^* is the upper limit of patients of at dose j before stopping of the trial is considered; n_c is the number of patients per cohort at a given stage, where a *stage* is defined as the period of administration and observation for a DLT after a dose has been selected; n_j represents the total number of patients who have received dose level j, and x_j denotes the number of those patients with observed DLTs, with $0 \leq x_j \leq n_j$. Finally, γ_j represents the toxicity threshold for dose j, indicating the posterior probability of the DLT rate at this dose exceeds ϕ.

Prior and conjugate posterior distribution

Given that that a DLT is a binary outcome, we employ Beta distributions, which are repre-

sented as $\pi_j \sim Beta(\alpha_j, \beta_j)$ for dose level j. Importantly, Beta distributions have conjugate posteriors $\pi_j|x_j, n_j \sim Beta(\alpha_j + x_j, \beta_j + n_j - x_j)$. In the case of trials assuming non-informative priors, we define the parameters of the prior distribution α and β by setting the prior mean equal to the target rate, while the prior effective sample size (PESS), $\alpha + \beta$, is suggested to match the standard cohort size (e.g., $n_c = 3$) at a given stage of administration.

Order constraints on posterior probabilities

We define $CPAT$ to be the Conjugate Posterior Probability of the DLT rate being Above the Target, expressed as

$$CPAT_j = P(\pi_j > \phi \mid x_j, n_j) \tag{1}$$

Order constraints on $CPAT$ values can be implemented after each stage using the Pool-Adjacent-Violators Algorithm (PAVA),[9] an isotonic regression method that is appropriate for toxicity which is presumed to be non-decreasing with dose.[10,11] PAVA ensures that the resulting estimates exhibit monotonicity and it is applied in a "local" manner, in the sense that only the $CPATs$ of the current dose and its cover and anti-cover doses are included. A corresponding constrained $CPAT_j$ value following PAVA is referred to as the PAVA-adjusted Posterior Probability of the DLT rate being Above the Target ($PPAT_j$).

Dose selection

Dose selection is restricted to either escalation, de-escalation, or maintaining the current dose. The proposed selection criterion involves identifying the dose for which the PPAT is nearest to τ, the PPAT threshold parameter, among the current dose and its cover/anti-cover doses, where $0 < \tau \leq 0.5$. The default value of τ is 0.5, and a PPAT value of 0.5 signifies the highest level of uncertainty regarding the safety of that dose level (akin to a coin flip). If we denote this set of dose levels as J^*, then dose $j^* \in J^*$ is selected if it minimizes (2):

$$\mid PPAT_{j^*} - \tau \mid = min_{j \in J^*} \mid PPAT_j - \tau \mid \tag{2}$$

When there is a tie in minimizing (2), with $\mid PPAT_j - \tau \mid = \mid PPAT_k - \tau \mid, j \leq k$, the decision rule is as follows:

Case 1: If both PPAT values are below τ, choose dose k.

Case 2: If both PPAT values exceed τ, choose j.

Case 3: If $PPAT_j \leq \tau$ while $PPAT_k > \tau$, choose dose j.

Toxicity control

During experimentation, we will dynamically remove doses that are considered excessively toxic. Given a threshold $\gamma_j > 0$, a dose j is excessively toxic if $CPAT_j > \gamma_j$. Note that if deemed overly toxic, dose j and all doses k, $j \leq k$, are removed from further consideration. Furthermore, each dose can have its own threshold.

Stopping rule

Stopping of the trial is invoked when:

Case 1: The minimum dose (i.e., dose 1) is excessively toxic, as indicated by $CPAT_1 > \gamma_1$, suggesting that the MTD is lower than this minimum dose level.

Case 2: The number of patients administered treatment in the trial reaches N_{max}.

Case 3: The number of patients at a dose j has reached or exceeds n_j^* and dose j is the next stage selection.

MTD identification

The MTD will be determined upon the stopping of the trial. MTD identification relies on isotonically-regressed posterior means of DLT rates. As noted, if the lowest dose is found to be excessively toxic ($CPAT_1 > \gamma_1$), all doses are considered too toxic and no MTD is selected. Otherwise, we implement PAVA "locally" on the conjugate posterior means of the final selected dose, along with its cover and/or anti-cover doses (if applicable), and regard these as the potential candidates. The candidates for MTD are required to have trial data. Let K^* be the group of candidates, and μ_k^p be the PAVA-adjusted posterior mean for dose k. The dose $k \in K^*$ with smallest PAVA-adjusted posterior mean difference with target DLT rate, $\mid \mu_k^p - \phi \mid$, is selected as MTD. If a tie occurs among $j, k \in K^*, j < k$ so that $\mid \mu_j^p - \phi \mid = \mid \mu_k^p - \phi \mid$, identification is as follows:

Case 1: If $\mu_j^p \leq \phi$ and $\mu_k^p > \phi$, then dose j is selected as MTD.

Case 2: If $\mu_j^p = \mu_k^p \leq \phi$, then dose k is selected.

Case 3: If $\mu_j^p = \mu_k^p > \phi$, then dose j is selected.

3. Backfills in Phase I clinical trials

Introduction to backfills

In addition to the assessment of toxicity, the evaluation of activity is crucial in clinical trials. For example, the emergence of molecular targeted agents and immunotherapies, characterized by toxicities that are frequently independent of dosage, has resulted in the formulation of the optimal biological dose concept, which takes into account both therapeutic activity and associated toxicity.[12] Activity refers to the capacity of an intervention, such as a cancer treatment drug, to achieve the intended positive outcome. It determines whether a treatment successfully attains the intended beneficial effects under ideal circumstances. In the lack of proven activity, a treatment is unlikely to be considered a feasible choice for clinical application.

Dose optimization aims to determine a dosage that maintains clinical activity while ensuring the tolerability of adverse effects, representing a key objective in the field of drug development. Alongside determining the appropriate dose for the dose-finding process aimed at identifying the MTD, it is essential to evaluate the activity at dose levels to reach dose optimization. In conventional practice, Phase I trials primarily concentrate on assessing the safety of the treatment using a limited sample size, and Phase II trials aim to evaluate the activity by increasing the sample size. The highest dose that is deemed tolerable, determined through a limited Phase I study, forms the foundation for the subsequent Phase II trial, which assesses the clinical activity.

However, proposals have been made advocating for the random assignment of patients to lower dose levels at an initial phase to assess the activity within the drug development process.[13] The method to be utilized is known as “backfill”,[14] signifying that we are able to simultaneously backfill a dose level considered to be safe. The main objective of backfilling is to assess whether a plateau exists in the activity of a drug. If such a plateau is identified, it would be advantageous to lower the dose in relation to the DLT rate while still preserving a similar level of activity, thus

revealing any possible trade-offs. In situations where the optimal dose is less than the MTD, the backfill strategy could prove to be particularly beneficial.[15] In practical clinical trials, it will be essential to conduct a comprehensive evaluation of both the possible benefits and the potential risks of a treatment in order to achieve a balance between activity and safety.

Literature review of backfills

Numerous methodologies exist regarding backfilling strategies in the early phases of clinical trials. Dehbi (2021) introduced an algorithm that involves randomly backfilling patients to dose levels that are lower than the currently administered dose, while employing Bayesian CRM as the framework for the toxicity trial to determine the MTD.[14] Backfilling to the lower dose levels is stopped when they demonstrate inadequate activity in comparison to higher doses, beginning with the lowest dose. Although the toxicity data from backfilled patients is not incorporated into the toxicity trial, the activity data from all patients, including both backfilled and dose-finding participants, are utilized to establish the activity plateau for determining the optimal dose.

BF-BOIN (Backfilling incorporated into the BOIN design) has implemented a backfill strategy utilizing the BOIN algorithm in Phase I clinical trials.[16] The doses available for backfilling are those that are lower than the currently administered dose, provided that at least one response has been observed at or below that level. If the DLT rate for a specific dose, along with the pooled DLT rate for that dose and its corresponding cover dose, exceeds the de-escalation threshold established by the BOIN design, that particular dose will be removed from consideration as a backfill candidate. A cohort of patients will be backfilled at the highest available dose. At the end of the trial, toxicity data from all doses, including those from both the toxicity trial and backfills, will be gathered to determine the MTD, while activity data will be collected to inform future dose optimization efforts.

The backfill i3+3 design represents an advancement of the i3+3 framework specifically for backfills.[17] The primary innovation of the i3+3 model is the incorporation of an additional criterion into the 3+3 design, allowing for the consideration of data variability, with the letter “i”

denoting "interval". The doses permitted for backfilling are those that are lower than the currently administered dose, unless they demonstrate insufficient activity compared to higher doses. Each patient is randomly backfilled with an equal opportunity. At the end of the trial, activity data from backfill patients and toxicity data from the i3+3 trial are collected, and the OBD is established by identifying the change point of the activity plateau.

4. BF-BOLD for determining MTD and OBD

The BOLD algorithm for identifying the MTD can be utilized for backfilling purposes to achieve the optimal dose during the ongoing trial for MTD. This implies that the ideal dose may be lower than the identified MTD while still preserving the similar level of activity as that of the MTD. Consequently, this approach (BF-BOLD) is expected to enhance the safety of the dose without undermining the treatment's effectiveness.

Prior and conjugate distributions of activity

The initial phase of BF-BOLD involves specifying the prior distribution of activity. The activity outcome, represented by x_j^a for dose level j, is presumed to be binary and to follow a Binomial distribution characterized by the activity rate π_j^a. It is further presumed that π_j^a follows a uniform distribution of Beta(1,1), whereas the mean activity rate is 0.5, PESS = 2, and the activity rate spans from 0 to 1. As a result, the conjugate distribution of π_j^a follows a Beta distribution: $\pi_j^a | x_j^a, n_j \sim Beta(1 + x_j^a, 1 + n_j - x_j^a)$ once the activity outcome x_j^a and sample size n_j have been observed with $x_j^a \leq n_j$.

Backfill eligibility

The main factor in determining an appropriate backfilled dose is safety. It is crucial to avoid assigning patients a dose that lacks adequate evidence to verify its safety. Consequently, the dose for backfilling must be lower than the dose currently being administered in the trial for MTD. In addition, backfill designs necessitate meticulous planning to reduce the likelihood of administering potentially inactive agents and/or sub-therapeutic doses to additional patients.[16] We define $CPAT^a$ as the Conjugate Probability of the Activity Rate being Above the Target, expressed in Equation 3, where "Target" means the minimum acceptable activity rate represented

by ϕ^a. If the value of $CPAT^a$ for dose j is below γ^a (with default value of 0.2), then all doses below j will be considered insufficiently active and not eligible for backfilling. Such an activity control is performed whenever backfill is employed at each stage of the trial. A dose that satisfies all of these criteria is considered an appropriate candidate for backfilling.

$$CPAT_j^a = P(\pi_j^a > \phi^a \mid x_j^a, n_j) \tag{3}$$

Backfill assignment

Two potential strategies for backfill assignment are suggested:

Strategy 1: We allocate patients to the eligible doses of backfill uniformly, commencing with the anti-cover dose of the dose presently being administered in the MTD trial, referred to as dose *c*, provided it meets eligibility criteria. The number of backfilled patients in the cohort may be set at 3 or randomly observed from the options of 1, 2, or 3. For example, in the event that the cohort size is 3 and the dose level of *c* is 4 or greater, then provided that all 3 doses immediately below *c* qualify for backfill, one patient will receive treatment with one of the 3 doses designated for backfill without replacement. If only 2 doses are eligible when the dose level of c is 3 or above, each dose will be randomly assigned to either one or 2 patients from a cohort of 3 patients for backfill. If only a single dose is eligible for backfill when the dose level of *c* is 2 or higher, all 3 patients will receive it for backfill. The same principle applies in cases where the cohort size is variable.

Strategy 2: We allocate all the patients in the cohort to the anti-cover dose of *c* for backfill, provided it meets eligibility criteria. The cohort size may be established at 3, randomly observed from the choices of 1, 2, or 3, or fixed at only one.

Observation of toxicity and activity

In the BF-BOLD design, the activity counts are monitored and documented in the backfill trials. Additionally, the DLTs of backfilled patients are also recorded alongside activity counts, and this information is vital for the dose escalation in the MTD trial. If the selected dose corresponds to the anti-cover dose of the currently administered dose in the MTD trial, its toxicity data

will be integrated into the dose selection process. This is because the selection of doses based on toxicity in BOLD involves evaluating doses that are adjacent to the current dose by comparing their PPATs. Thus, the backfill serves a dual purpose: evaluating both activity and toxicity. Furthermore, the activity data for the dose currently being administered in the MTD trial is observed and integrated in addition to the observed DLT. Consequently, the MTD trial aids in the backfill process, while backfill subsequently supports the dose selection for the MTD trial.

Stopping rule

The trial utilizing BF-BOLD is halted once one of the following conditions is met: 1) The overall count of patients who receive treatment in the MTD trial (not including the patients for backfilling) has reached a designated figure (e.g., $N_{max} = 21$ or 30); 2) The overall number of patients at the currently administered dose, including those treated in the MTD trial and participants in the backfill process, has reached a specified maximum limit n_j^*. The value of n_j^* is set to be 12 for $j > 1$, and n_1^* is set to be 15 in order to enhance classification to the bottom state, similar to the BOLD design.

Optimal dose identification

The optimal backfill dose (OBD) is identified following the completion of the MTD trial. The OBD is determined by the isotonically-adjusted (PAVA) posterior means of activity rates for doses that do not exceed the identified MTD (M). Local PAVA (for the dose of interest and its anti-cover and cover dose) will be applied. There are two thresholds for OBD. First, a reduction (trade-off) from the identified MTD for the PAVA posterior mean activity rate is accepted for OBD, where the default trade-off is set to be 10% (user choice). Second, the PAVA posterior mean activity rate of the OBD must be at least ϕ^a, the minimum acceptable toxicity rate. Additionally, OBD cannot exceed M. The lowest dose which does not exceed M and whose PAVA posterior mean activity rate is both at least ϕ^a and within the trade-off from that of M will be identified as OBD. OBD is not determined if the minimum dose is considered to be excessively toxic.

Flowchart of backfills

The flowchart illustrating the backfills integrated with BOLD is displayed in Figure 1. A summarized description of the flowchart is as follows: 1) Specify the target DLT rate ϕ and the minimum acceptable activity rate ϕ^a. 2) Specify the prior information at dose level, such as prior mean DLT rates (default values $= \phi$), PESS (default values $= 3$), and prior distribution of activity rate (default is Beta(1,1)). 3) Determine the dose decision in MTD trial, while commencing with the minimal dose if the trial is at the initial stage. Note that the specifics of BOLD for toxicity trials were previously illustrated and are not included in this flowchart. 4) Monitor if the overall sample size ceiling for MTD trial (e.g., $N_{max} = 21$ or 30, not including the patients for backfill), is reached. 5) Monitor if the total number of treated patients (including MTD trial and backfill) at the current dose reaches maximum (e.g., 12 or 15) when the decision is maintaining the current dose. 6) Administer the selected dose to a cohort of patients (default $n_c = 3$) in the MTD trial to gather data on both toxicity and activity. 7) Update $CPAT$ values which represent the posterior probability that a dose is overly toxic. 8) Update $CPAT^a$ values, which indicate the posterior probability of a dose being active, and assess whether any dose that is lower than the currently administered dose in the MTD trial is less than γ^a (with a default value of 0.2) to determine its ineligibility for backfill. 9) Choose a backfill strategy, determine the backfill dose(s), and assign patient(s) for backfills. 10) Observe both DLTs and dose activity of backfilled patients. 11) Update PPATs of toxicity at dose level by incorporating backfill data. 12) Repeat Steps 3 to 11 until the trial stops. 13) Once stopping occurs, identify the MTD by considering PAVA-adjusted posterior mean DLT rate and OBD by PAVA-adjusted posterior mean activity rate, except when the lowest dose exhibits excessive toxicity, leading to the conclusion that there is no MTD or OBD.

Demonstration of backfills

A demonstration of backfills for single-drug trials can be presented by simulation (using R 4.4.3). We assume that the trial comprises 5 distinct dose levels and each cohort consists of 3 participants. The target toxicity rate ϕ and minimum acceptable activity rate ϕ^a are set at 0.25 and 0.3, respectively. The MTD trial is limited to a maximum of 30 patients ($N_{max} = 30$). No more than 12 patients are assigned to each specific dose, including those in the MTD trial and

backfills, and 15 patients for the lowest dose, if the decision of MTD trial is to maintain the current dose. The true MTD and true OBD are assumed to be dose level 4 and 2, respectively. The prior mean DLT rate is assumed to be equivalent to ϕ for all dose levels, and the prior distribution of activity rate is assumed to be Beta(1,1). The PESS is established at 3. The $CPAT$ criterion for identifying excessively toxic doses is established at $\gamma_j = 0.95$ for $j > 1$ and $\gamma_1 = 0.9$ for the lowest dose, while the criterion for identifying inactive doses is uniformly set at 0.2 across all doses. The true DLT rates in the simulation are presumed to be 0.1, 0.11, 0.12, 0.25, and 0.5 for doses 1 to 5, respectively. Consequently, the true activity rates are 0.35, 0.5, 0.5, 0.5, and 0.5, indicating that the activity stabilizes from dose level 2 onward, and the activity rate of its anti-cover dose is 0.15 lower than 0.5.

The utilized backfill strategy involves administering the anti-cover dose of the current dose in the MTD trial to a randomly observed cohort of up to 3 backfilled patients. This strategy may prove to be more practical in real-world clinical trials, for backfills are typically conducted after a cohort of patients in the MTD trial has received the currently administered dose and prior to the observation of their toxicity responses, which will provide toxicity data for dose decision. The length of this interval differs among patients, leading to a variable number of patients available for backfills, which refers to the staggering of accrual. The longer this interval extends, the greater the number of patients that can be backfilled. This variability directly corresponds to our random selection of backfilled patients (for instance, ranging from 1 to 3) in the simulation. Additionally, a 10% trade-off of activity rate is utilized for the OBD identification.

The process of dose escalations and backfills is presented in Figure 2. The trial comes to an end at the tenth stage when the total sample size of MTD trial reaches 30. Dose 4, which is the current dose at the tenth stage, is accurately recognized as the MTD due to its isotonically-adjusted posterior mean toxicity rate being nearest to ϕ when compared among doses 3, 4, and 5 (specifically, 0.162, 0.183, and 0.625) at the end of the trial. Dose 2 is correctly identified as the OBD, as it signifies the minimum dose at which the isotonically-adjusted posterior mean activity rates are within a 10% margin of the identified MTD (dose 4) and greater than ϕ^a. This outcome

indicates that the backfill strategy has proven to be effective, as dose level 2, recognized as the OBD, is situated below the MTD (dose 4) and marks the onset of the activity plateau.

Performance analysis

The performance of backfills can be assessed through simulation. In a clinical design consisting of 5 doses, for example, the true MTD levels with target DLT rates encompass "<1", 1, 2, 3, 4, 5, and ">5", where "<1" indicates that all doses are too toxic and ">5" signifies that all doses exhibit DLT rates that fall below the target. The simulations are conducted using random scenarios that incorporate both true DLT rates and activity rates, whereas the lower and upper limit are set at 0.01 and 0.7, respectively. We assert that the disparity between the true DLT rate for the cover dose of MTD and that of MTD, i.e., upper δ, is either 0.1, 0.15, 0.2 or 0.25, whereas the true DLT rate for the anti-cover dose of MTD is at least 0.05 lower than that of the MTD, assuming this is achievable. For a specified upper δ, 50 random scenarios are created for each MTD level, with a total of 350 random scenarios. For each of the 350 random scenarios, 100 simulations of response sequences are generated.

It is presumed that the true OBD is zero, one or two levels lower than the true MTD. Assume the true OBD is dose k, then the true activity rate of dose k is established at 0.5, which is the mean activity rate of the prior uniform distribution. It is also assumed that a plateau of activity rate begins from dose k, and the true activity rate of the anti-cover of dose k (i.e., dose k - 1 if it is present) is determined to be 0.5 minus the value of lower δ, with the default value of lower δ set at 0.15. In order to ensure that the activity rate of dose $k-1$ is distinctly different from that of dose k, we establish the value of lower δ to be no less than 10% of 0.5 (which is 0.05). This is for the activity rate of dose $k-1$ not to overlap with the trade-off range associated with dose k. The true activity rates of the doses situated below $k-1$, if applicable, are randomly generated within a lower limit of 0.01 and an upper limit corresponding to the true activity rate of $k-1$ minus 0.01.

Table 1 presents the simulation results of BF-BOLD design regarding the identification accuracies, efficiencies, numbers of observed DLTs, overdose rates, underestimation rates, and

overestimation rates with a 5-dose lattice at $\phi = 0.25$, $\phi^a = 0.3$, and N_{max} (for MTD trial) $= 21$, with upper δ for random scenarios of true DLT rates set to be 0.15. The accuracy includes the average number of accuracy rate of MTD and OBD identification, as well as the standard deviation derived across various scenarios. The efficiency includes the average number of treated patients and the standard deviation across various scenarios. The overdose rate is the proportion of the patients treated with doses higher than the true MTD, the underestimation rate is the proportion of trials with identified MTD/OBD being lower than the true MTD/OBD for a given scenario, and overestimation rate is the proportion of trials with identified MTD/OBD being higher than the true MTD/OBD for a given scenario. Again, OBD is assumed to be zero, one or two levels below the true MTD. An activity plateau is assumed to be present, and a random number of backfilled patients (up to 3) receive the anti-cover dose corresponding to the current dose in the MTD trial. A 10% activity rate trade-off is utilized for the identification of OBD. Table 2 illustrates the result for N_{max} (for MTD trial) $= 30$.

Comparative analysis with BF-BOIN

Comparative analyses are also conducted. Given that BOIN was chosen as the comparator for the BOLD design,[3] we similarly utilize BF-BOIN as the reference for BF-BOLD. Recall that the BF-BOLD has been employing a strategy of backfilling a random number of patients with the anti-cover dose of the current dose in MTD trial. This approach makes the two designs more comparable, for BF-BOIN assumes a time frame for the toxicity data from the MTD trial to be available before any dose decisions can be made, leading to variability in the cohort size of backfilled patients.

Figures 3 and 4 present the bar charts that compare the accuracy and efficiency, respectively, of the MTD trial between BF-BOLD and BF-BOIN, stratified by N_{max} and upper δ value for MTD. This comparison employs random scenarios of true DLT rates and activity rates for simulations, while the OBD is presumed to be one dose level lower than the MTD and a plateau of activity is present. Note that γ_1 is established at 0.9, while γ_j is designated as 0.95 for $j > 1$ in both designs. The values of ϕ and ϕ^a are established at 0.25 and 0.3, respectively. The bar charts

illustrating the accuracy with OBD at zero and two levels beneath MTD are displayed in Web Figures 2 and 3, respectively. The figures show that BF-BOLD surpasses BF-BOIN in accurately identifying the MTD across all dose levels, upper δ values, and gaps between MTD and OBD for both N_{max}, with the exception when all the doses are too toxic.

Comparison of backfill strategies in BF-BOLD

We also examine the effects of various proposed backfill approaches in clinical trials employing the BF-BOLD design, including the BOLD design without backfill implementation. Web Table 1 represents the summary of average accuracy (MTD and OBD) and efficiency (overall and backfilled) for a 5-dose lattice at $\phi = 0.25$ and $\phi^a = 0.3$ with $N_{max} = 21$ and 30 for MTD trial, utilizing random scenarios of true DLT rates and activity rates with upper δ for MTD trials set at 0.15, while OBD being zero, one or two levels below the true MTD with an activity plateau. Two backfill strategies are employed, utilizing 2 and 3 approaches respectively. For the first strategy, patients are evenly backfilled with the doses commencing from the anti-cover dose of the current dose being administered in the MTD trial, which has the cohort size of 3 or ≤ 3. For the second strategy, patients are backfilled with the anti-cover dose of the current dose, which has the cohort size of 3, ≤ 3 or 1. Web Table 2 presents the rates of overdose (overall and backfilled), underestimation (MTD and OBD), and overestimation (MTD and OBD) based on identical assumptions.

6. Discussion

The BOLD algorithm can facilitate the identification of the optimal dose through the implementation of the backfill method. The benefits of backfills are clear, as it allows for identifying a dose that is below the MTD while still maintaining the drug's effectiveness. Such a dose may be more favorable than the MTD, as it presents a lower risk of DLTs and may offer enhanced practical sustainability of the treatment. Our method for determining the OBD by balancing activity against the identified MTD is straightforward and does not require a complex statistical model. Simulations indicate that our proposed BF-BOLD method generally surpasses BF-BOIN in performance, except when the true MTD is below the lowest dose (Figure 3, Web Figures 2-3). Furthermore, it shows that conducting toxicity trials with backfills enhances the accuracy of

MTD identification and reduces the incidence of overdoses compared to trials without backfills (Web Tables 1-2).

The two backfilling strategies we propose within the BF-BOLD framework also offer clinicians greater flexibility during the backfilling process. Web Tables 1 and 2 indicate that the approach of backfilling a single patient with the anti-cover dose corresponding to the current MTD trial dose may be the most advantageous strategy. This approach requires a reduced number of backfilled patients and leads to a decreased occurrence of DLTs; yet it does not significantly compromise other operating characteristics, such as the accuracy of MTD and OBD, along with the rates of overdose, activity underestimation, and activity overestimation. Nonetheless, it is crucial to recognize that incorporating backfills requires treating a larger patient cohort, which in turn increases overall costs. Further investigation is needed to find a balance between the costs associated with backfills and the optimal dose determination.

7. Conclusion

The BOLD design framework can be utilized and modified for backfill designs aimed at determining the optimal dose. This adaptability underscores the extensive capabilities of the BOLD design for Phase I clinical trials.

8. References

1. Liu S, Yuan Y. *Bayesian optimal interval designs for Phase I clinical trials.* Journal of the Royal Statistical Society, 2015. 64: p. 507-523.

2. Braun T. *The current design of oncology Phase I clinical trials: progressing from algorithms to statistical models.* Chin Clin Oncol. 2014 Mar; 3(1):2.

3. Wang GM and Tatsuoka C. *Bayesian ordered lattice design for Phase I clinical trials*. MedRxiv, doi: https://doi.org/10.1101/2025.03.23.25324471.

4. Conaway MR, Dunbar S, Peddada SD. *Designs for single- or multiple-agent Phase I trials.* Biometrics, 2004. 60(3): p. 661-9.

5. Tatsuoka C, Ferguson T. *Sequential classification on partially ordered sets*. Journal of the Royal Statistical Society Series B. 2003. 65. p. 143-157.

6. Tatsuoka C. *Sequential Classification on Lattices with Experiment-Specific Response Distributions*, Sequential Analysis. 2014. 33(3), p. 400-420.

7. Tatsuoka C, Chen W, Lu X. *Bayesian Group Testing with Dilution Effects*. Biostatistics. 2022. doi: 10.1093/biostatistics/kxac004.

8. Thall PF, Garrett-Mayer E, Wages NA, et. al. *Current issues in dose-finding designs: A response to the US Food and Drug Administration's Oncology Center of Excellence Project Optimus.* Clin Trials. 2024 Jun; 21(3): 267-272.

9. Barlow RE, Bartholomew DJ, Bremner JM, et al. *Statistical Inference under Order Restrictions.* 1972: London, New York: John Wiley and Sons.

10. Leung DH, Wang Y. Isotonic designs for Phase I trials. *Control Clin Trials.* 2001 Apr; 22(2):126-38.

11. Ivanova A, Flournoy N. *Comparison of Isotonic Designs for Dose-Finding.* Stat Biopharm, Res. 2009 Feb 1; 1(1):101.

12. Fraisse J, Dinart D, Tosi D, et al. *Optimal biological dose: a systematic review in cancer phase I clinical trials.* BMC Cancer 21, 60 (2021).

13. Korn EL, Moscow JA, Freidlin B. *Dose optimization during drug development: whether and when to optimize.* J Natl Cancer Inst. 2023 May 8; 115(5):492-497.

14. Dehbi HM, O'Quigley J, Iasonos A. *Controlled Backfill in Oncology Dose-finding Trials, Contemporary Clinical Trials.* 2021.

15. Chen X, Zhang, J, Li B, et al. *Determining doses for backfill cohorts based on patient-reported outcome.* BMC Med Res Metho; dol 24, 27 (2024).

16. Zhao Y, Yuan Y, Korn EL, et. al. *Backfilling Patients in Phase I Dose-Escalation Trials Using Bayesian Optimal Interval Design (BOIN).* Clin Cancer Res. 2024 Feb. 16;30(4):673-679.

17. Liu J, Yuan S, Bekele BN, et al. *The Backfill i3+3 Design for Dose-Finding Trials in Oncology.* N Engl J Stat Data Sci 2(2024), no. 3, 271-283.