A Seamless Two-stage Phase I/II Trial Design with Backfill and Joint Monitoring for Dose Optimization

Maximize learning per patient via backfilling and randomization in a seamless design that accelerates efficacy claims, shortens trials, and improves reliable dose optimization

Belay B Yimer, PhD (Astellas)

Joint work with Kentaro Takeda (Astellas), Jing Zhu (Astellas), Akihiro Hirakawa (Institute of Science Tokyo)

October 22, 2026

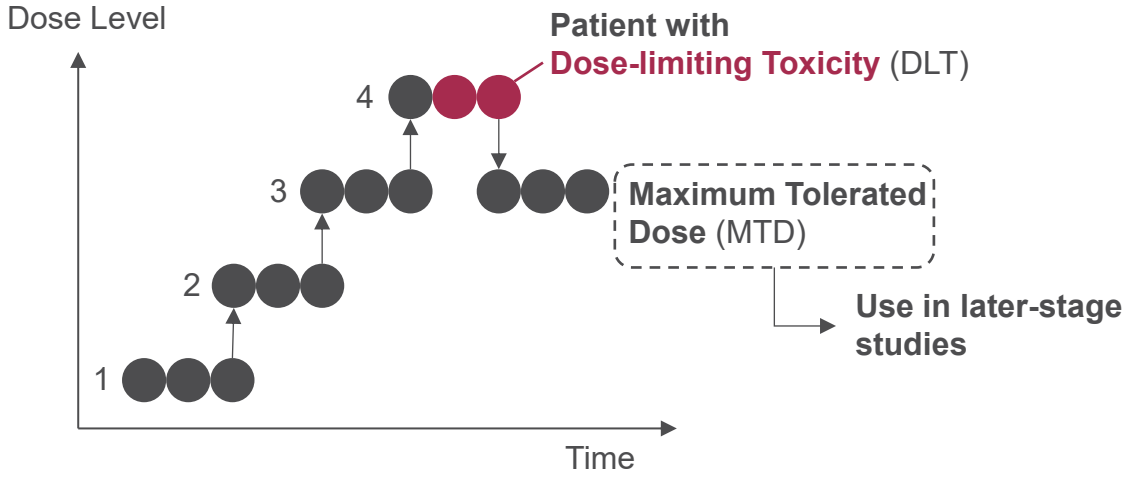
Disclaimer

The authors are paid employees of Astellas. This presentation is intended for informational purposes only and does not replace independent professional judgment. This presentation is not intended to be legal advice. Statements of fact, positions taken and opinions expressed are those of the speaker individually and, unless expressly stated to the contrary, do not necessarily reflect the opinion or position of the speaker’s employer, Astellas, or any of its subsidiaries and/or related entities.

Dose Optimization Requires More Than Finding the MTD

Traditional Dose Selection

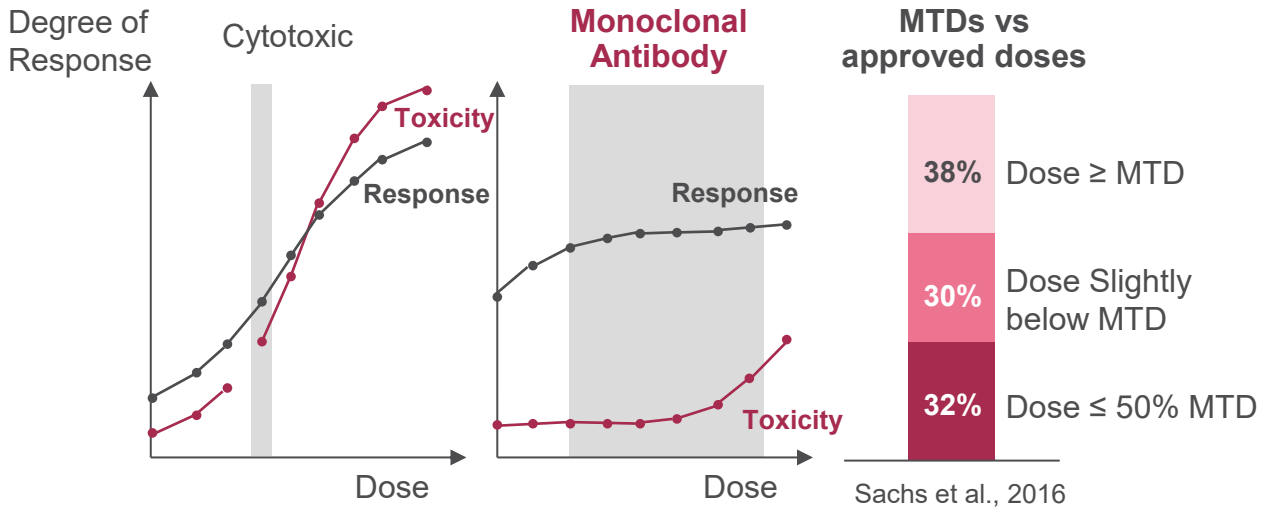
"More is Better" Approach for Cytotoxic Chemotherapy



Recent Trend of Oncology

Prior Approach Not Appropriate for Targeted Therapies

(Shah et al., 2021)



For targeted and immune therapies, efficacy may plateau while toxicity continues to increase.

The MTD may not provide the most favorable benefit–risk profile.

Objective: identify an efficacious, tolerable dose with the best efficacy–toxicity trade-off.

FDA Project Optimus emphasizes robust dose optimization based on efficacy, safety, and tolerability rather than reliance on the MTD alone.

Current Dose-Development Pathways Leave Three Gaps

Limited information across doses

Conventional escalation may provide limited efficacy and tolerability information across dose levels

No definitive randomized comparison

Selecting a candidate from non-randomized early dose-finding data may leave substantial uncertainty about the optimal dose

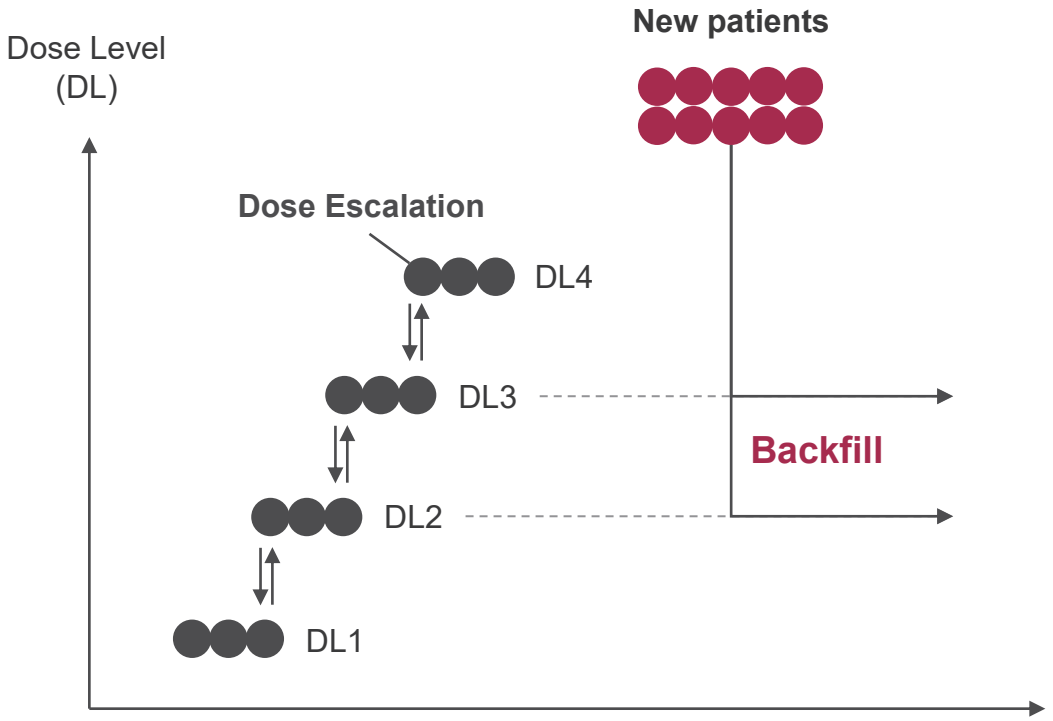
Separate development stages

Dose finding, dose optimization, and efficacy evaluation may require separate studies, increasing patient numbers and development time.

A design is needed that learns efficiently across doses, preserves randomization for candidate-dose comparison, and supports a statistically controlled efficacy conclusion.

Proposing the BF-BOIN-ET Based on Efficacy and Toxicity Outcomes

Backfill Diagram



What backfill cannot do alone

Provide a definitive randomized dose comparison.
Resolve all uncertainty between promising doses.
Replace formal efficacy and toxicity monitoring.

Benefits of Backfill

1. Concurrently assign patients to doses previously cleared for safety while prioritizing dose escalation
2. Simultaneously collect more safety & efficacy at multiple dose levels

▶ **Shortens the timeline of dose optimization**

Dose-finding Design Comparison

BF-BOIN-ET (Takeda et al., 2025) ▶ Identifies optimal dose based on efficacy and toxicity outcomes

BF-CRM (Dehbi et al., 2021)
BF-BOIN (Zhao et al., 2024)
BF-BLRM (Barnett et al., 2024)
Bi3+3 (Liu et al., 2024)

▶ Detect the MTD and use efficacy response as a trigger to open backfill cohorts

Proposed Solution: One Trial, Three Objectives

1 Explore

Explore doses using escalation and backfill.

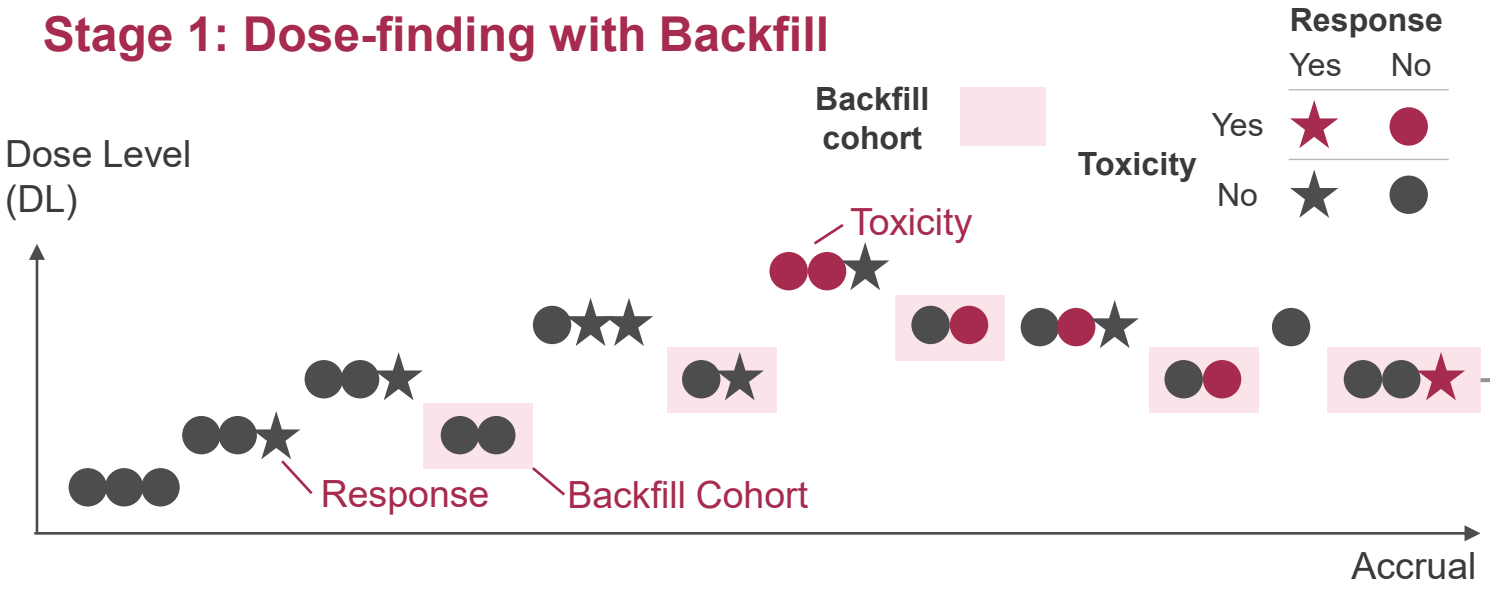
2 Compare

Compare two candidate doses through randomization.

3 Select and evaluate

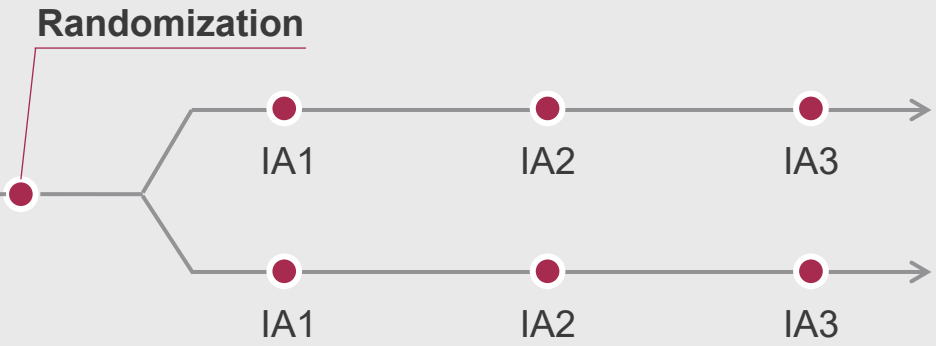
Select and evaluate the OD using joint efficacy and toxicity monitoring.

Stage 1: Dose-finding with Backfill



Stage 1 identifies two promising doses using efficacy and toxicity information.

Stage 2: Dose Optimization & Efficacy Declaration



Stage 2 compares candidate doses and supports a statistically controlled efficacy conclusion.

Stage 1 data at the selected doses contribute to Stage 2 monitoring

Stage 1:
Bayesian Optimal Boundaries For Toxicity and Efficacy

Notation for Toxicity

The same boundaries as the original BOIN design (Liu and Yuan, 2015)

- ϕ_{T0}: Target toxicity probability specified by the investigator.
- ϕ_{T1}: Highest toxicity probability, warranting dose escalation.
- ϕ_{T2}: Lowest toxicity probability considered overly toxic, requiring dose de-escalation.
- λ_{T1} and λ_{T2}: Lower and upper cutoffs, satisfying

$$\phi_{T1} < \lambda_{T1} < \phi_{T0} < \lambda_{T2} < \phi_{T2}.$$

- $\hat{p}_{Tj} = y_{Tj}/n_j$ denote the observed Bernoulli toxicity probability based on all cumulative toxicity data at dose level j, D_{Tj}

Bayesian Optimal Boundaries for Toxicity

$$\lambda_{T1} = \log\left(\frac{1-\phi_{T1}}{1-\phi_{T0}}\right) / \log\left(\frac{\phi_{T0}(1-\phi_{T1})}{(1-\phi_{T0})\phi_{T1}}\right),$$

$$\lambda_{T2} = \log\left(\frac{1-\phi_{T0}}{1-\phi_{T2}}\right) / \log\left(\frac{\phi_{T2}(1-\phi_{T0})}{(1-\phi_{T2})\phi_{T0}}\right).$$

Notation for Efficacy

Two simple hypotheses are considered to derive the optimal value for a binary efficacy outcome

- ϕ_{E0}: Minimum required efficacy probability specified by the investigator.
- ϕ_{E1}: Efficacy probability considered inefficacious.
- λ_E denotes the efficacy cutoff, satisfying $\phi_{E1} < \lambda_E < \phi_{E0}$
- $\hat{p}_{Ej} = y_{Ej}/n_j$ denote the observed Bernoulli efficacy probability based on all cumulative efficacy data at dose level j, D_{Ej}

Bayesian Optimal Boundaries for Efficacy

$$\lambda_E = \log\left(\frac{1-\phi_{E1}}{1-\phi_{E0}}\right) / \log\left(\frac{\phi_{E0}(1-\phi_{E1})}{(1-\phi_{E0})\phi_{E1}}\right).$$

Toxicity boundaries guide escalation, while the **efficacy boundary** determines backfill eligibility.

Stage 1:
Likelihood with Pending Data

The Problem

When a backfill-approach is considered:

- Usually, the dose-escalation cohort exploring a higher dose enrolls patients first, followed by the backfill cohort
- Therefore, the observed $\hat{p}_{Tj}$ and $\hat{p}_{Ej}$ may not be available because the backfill cohort has not completed the toxicity and efficacy evaluation period when the dose-escalation cohort has sufficient information

The Solution - Likelihood with pending efficacy and toxicity data *(Lin and Yuan (2020), Takeda et al. (2020))*

$$L(p_{Ej}|D_{Ej}) \propto p_{Ej}^{\tilde{x}_{Ej}}\left(1-p_{Ej}\right)^{\tilde{m}_{Ej}}, L(p_{Tj}|D_{Tj}) \propto p_{Tj}^{\tilde{x}_{Tj}}\left(1-p_{Tj}\right)^{\tilde{m}_{Tj}}$$

- $\tilde{x}_{Ej}$ and $\tilde{x}_{Tj}$ are the number of patients who experienced efficacy and toxicity by the interim time k, respectively
- $\tilde{m}_{Ej}$ and $\tilde{m}_{Tj}$ are the effective number of patients who have not experienced efficacy and toxicity
- Effective binomial efficacy and toxicity data $\tilde{D}_{Ej} = (\tilde{n}_{Ej}, \tilde{x}_{Ej})$ and $\tilde{D}_{Tj} = (\tilde{n}_{Tj}, \tilde{x}_{Tj})$, where $\tilde{n}_{Ej} = \tilde{x}_{Ej} + \tilde{m}_{Ej}$ and $\tilde{n}_{Tj} = \tilde{x}_{Tj} + \tilde{m}_{Tj}$ are the effective sample size for efficacy and toxicity at dose level j

→ $\hat{p}_{Ej}$ and $\hat{p}_{Tj}$ can be replaced by $\tilde{p}_{Ej} = \tilde{x}_{Ej}/\tilde{n}_{Ej}$ and $\tilde{p}_{Tj} = \tilde{x}_{Tj}/\tilde{n}_{Tj}$ as the maximum likelihood estimate of p_{Ej} and p_{Tj}

Stage 1:
Dose Allocation Algorithm

Dose Allocation Algorithm j_{min} is the minimum effective dose defined by the lowest dose that satisfies $\tilde{p}_{Ej} \geq \lambda_E$.



① If $\tilde{p}_{Tj} \leq \lambda_{T1}$,

Dose-escalation: the next dose is $j + 1$

Backfill: the admissible dose set is $A_j = \{j_{min}, \ldots, j\}$;

② If $\lambda_{T1} < \tilde{p}_{Tj} \leq \lambda_{T2}$,

Dose-escalation: the next dose is j

Backfill: the admissible dose set is $A_j = \{j_{min}, \ldots, j-1\}$

③ If $\tilde{p}_{Tj} > \lambda_{T2}$,

Dose-escalation: the next dose is $j - 1$

Backfill: the admissible dose set is $A_j = \{j_{min}, \ldots, j-2\}$

Enrolment Procedure

Dose-escalation cohort prioritizes enrollment to explore a higher dose

Backfill cohort enrolls patients at the dose level that maximizes efficacy-toxicity trade-off utility within the admissible dose set

Suspension Rule

Suspends dose allocation if more than a predefined percentage of patients (e.g., 33%) have not completed the toxicity evaluation period at the current dose level for the dose-escalation cohort

Enrollment Stopping Rule

The total number of patients in the dose-escalation cohort reaches the maximum total number, N_{esc}

The number of all patients at the current dose level reaches the maximum number of patients at a dose level, n_{stop}

Selection of candidate doses for Stage 2

Among the subset of dose levels that satisfy the tolerability, the doses that **provide high utilities** are selected as the candidate doses for Stage 2 at the end of dose-finding trial.

Stage 2:
Joint Monitoring for Efficacy and Toxicity for Each Dose

Dirichlet-multinomial model

Outcome Model:

$Y \sim Multinomial(\theta_1, \dots, \theta_k)$, where $\theta_k = Pr(Y = k)$ is the probability that Y belongs to the k^{th} category, $k = 1, \dots, K$.

Prior distribution of parameter:

$\boldsymbol{\theta} = (\theta_1, \dots, \theta_K)^T \sim Dir(a_1, \dots, a_K)$, where $a_1, \dots, a_K$ are hyperparameters.

Posterior distribution:

$\boldsymbol{\theta}|D_n \sim Dir(a_1 + x_1, \dots, a_K + x_K)$, with $\sum_{k=1}^{K} a_K = 1$ to have a vague prior, where $D_n = (x_1, \dots, x_K)$ denote interim data and x_k denote the number of patients with $Y = k$.

The Dirichlet-multinomial model becomes a beta-binomial model when Y is a binary outcome.

Joint efficacy and toxicity

1=(response, toxicity), 2=(response, no toxicity), 3=(no response, toxicity), and 4=(no response, no toxicity).

When the objective is to evaluate if both endpoints achieve a pre-specified threshold, the hypothesis will be:

$$H_0 : \theta_1 + \theta_2 \leq \phi_{01} \text{ or } \theta_1 + \theta_3 \geq \phi_{02}$$
$$\text{versus}$$
$$H_1 : \theta_1 + \theta_2 > \phi_{01} \text{ and } \theta_1 + \theta_3 < \phi_{02}$$

Stage 2: Stopping Boundaries

Optimizing Design Parameters

- The probability cutoffs for futility stopping $C_F(n)$ and efficacy stopping $C_E(n)$ are

$$C_F(n) = \lambda \left(\frac{n}{N}\right)^{\gamma},$$

$$C_E(n) = 1 - (1-\lambda)\left(\frac{n}{N}\right)^{\eta},$$

  where λ, γ, and η are tuning parameters.

- The tuning parameters (λ, γ, η) are optimized through a grid search such that the statistical power is maximized while controlling the type I error rate at a certain pre-specified level.

An example of futility and efficacy stopping boundaries. Type I error rate is 10%.

Type		Stop the trial if	N				
			24	30	36	42	48
Futility	Or	# of response ≤	5	6	8	10	13
		# of toxicities ≥	11	13	15	17	19
Efficacy	and	# of response ≥	10	11	12	13	14
		# of toxicities ≤	6	8	11	14	18

Stage 2:
Trial Design

Enrollment

Enroll n patients in the j^{th} dose levels.

Interim stopping

- Futility stopping: stop enrolling patients and claim failure of the treatment at the dose level if $\Pr(\theta_1 + \theta_2 > \phi_{01}|D_{N1}, D_{n2}) < C_F(n)$ or $\Pr(\theta_1 + \theta_3 < \phi_{02}|D_{N1}, D_{n2}) < C_F(n)$
- Efficacy stopping: stop enrolling patients at all dose levels, reject H0, and claim that the treatment is promising at the dose level and the dose level is the optimal dose if $\Pr(\theta_1 + \theta_2 > \phi_{01}|D_{N1}, D_{n2}) \geq C_E(n)$ and $\Pr(\theta_1 + \theta_3 < \phi_{02}|D_{N1}, D_{n2}) \geq C_E(n)$ and $U_j > U_{-j}$
- Otherwise: continue enrolling patients until the next interim analysis.

Treatment Evaluation

Once maximum sample size reached or accrual stopped for all doses due to futility, evaluate based on the final data D:

- Claim that the treatment is promising at the dose level if $\Pr(\theta_1 + \theta_2 > \phi_{01}|D) \geq C_E(n)$ ***and*** $\Pr(\theta_1 + \theta_3 < \phi_{02}|D) \geq C_E(n)$
- Otherwise, declare that the treatment is not efficacious at the dose level.
- Declare the dose level that maximizes U_j as the OD.

Simulation Studies

Target of Comparison

Proposed-E	Proposed design with the equal randomization backfill approach
Proposed-U	Proposed design with the utility-based backfill approach
BARD-BOIN	BARD-BOIN design with equal randomization
U-BOIN-E	U-BOIN design with the equal randomization approach
U-BOIN-U	U-BOIN design with the utility-based pick-the-winner approach

Design Parameters in Stage 1

- Target toxicity probability $\phi_{T0} = 0.30$; minimum required efficacy probability $\phi_{E0} = 0.25$
- The optimal boundaries: $\lambda_{T1} = 0.236, \lambda_{T2} = 0.359, \lambda_E = 0.197$
- Maximum sample size: 36

Design Parameters in Stage 2

- **Sample size**
  The first interim look was performed when the first 24 patients in Stage 1 and Stage 2 for each dose level were treated; the following interim looks were performed every 6 additional patients (24, 30, 36, 42, and 48)
- $H_0 = (0.10, 0.15, 0.30, 0.45)$ and $H_1 = (0.05, 0.45, 0.15, 0.35)$

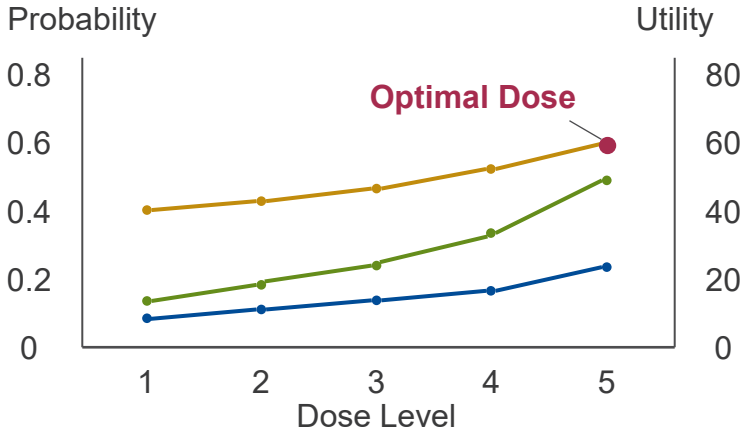
Type		Stop the trial if	N				
			24	30	36	42	48
Futility	Or	# of response ≤	5	6	8	10	13
		# of toxicities ≥	11	13	15	17	19
Efficacy	and	# of response ≥	10	11	12	13	14
		# of toxicities ≤	6	8	11	14	18

Simulation Scenarios

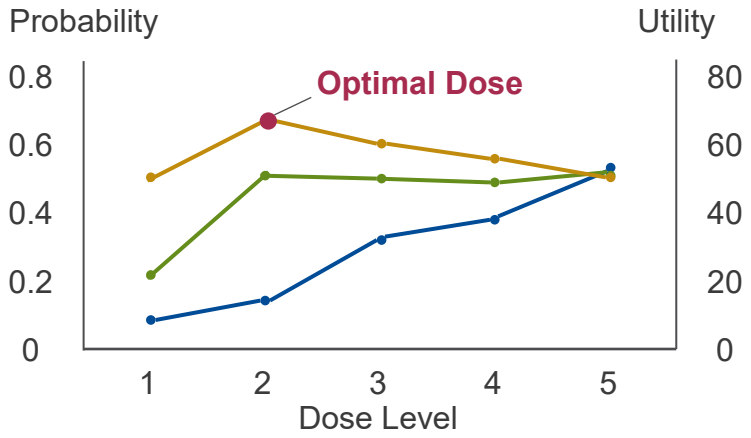
Simulation Across 12 Fixed Scenarios

Using four scenarios as examples

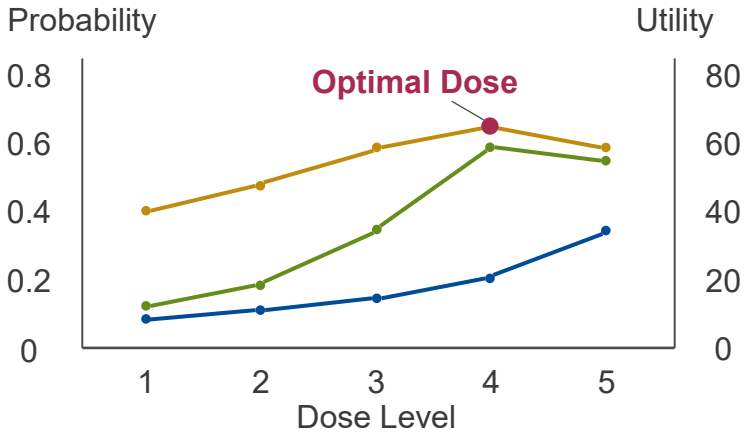
Scenario 1 (Increase)



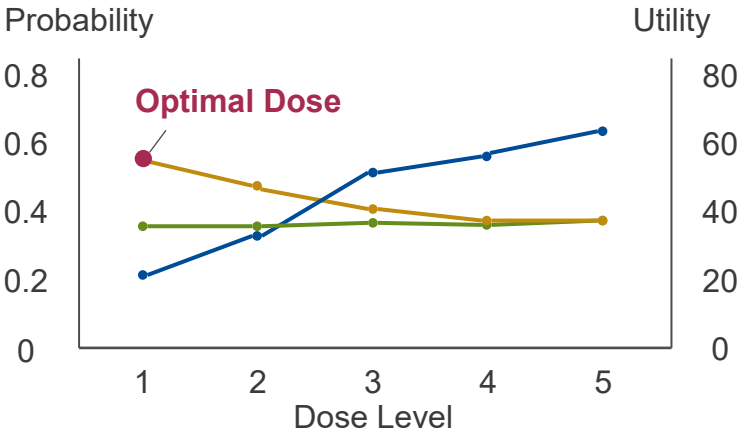
Scenario 8 (Bell-shaped)



Scenario 3 (Plateau)



Scenario 10



The **OD** is specified

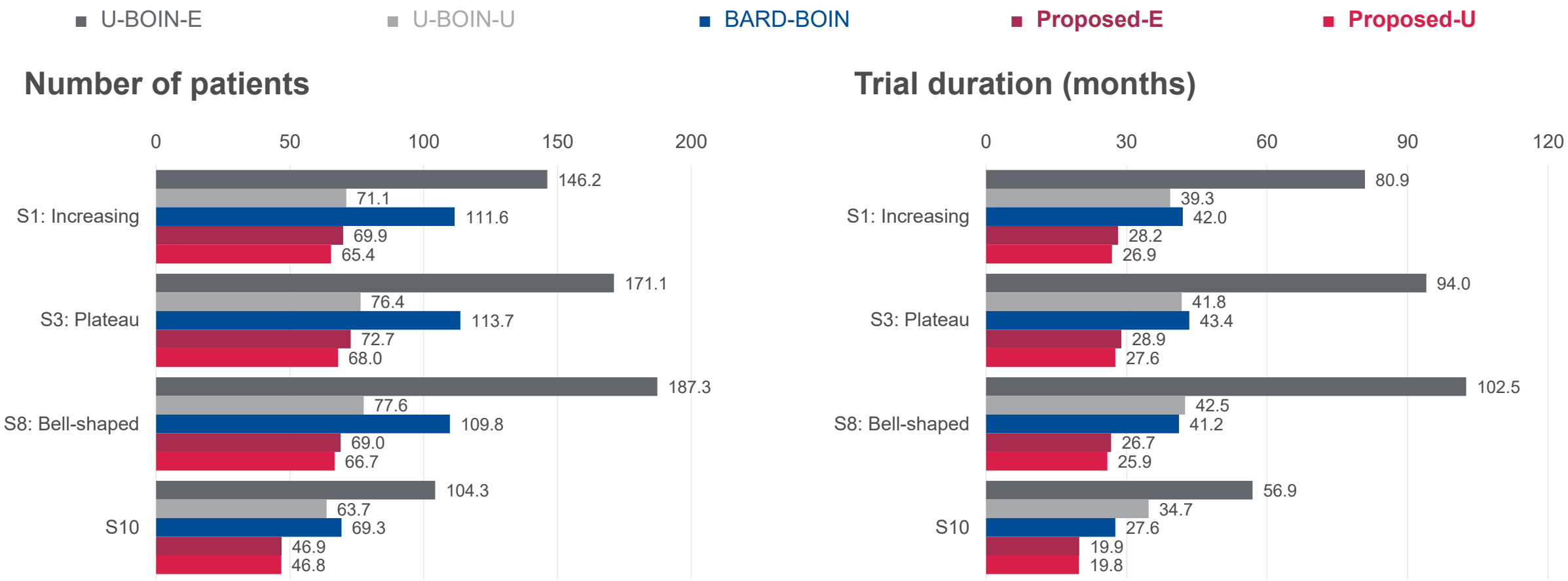
Orange line: the true utility

Green line: the true probability of efficacy → monotonically increases, plateaus, and is bell-shaped

Blue line: the true probability of toxicity → monotonically increases

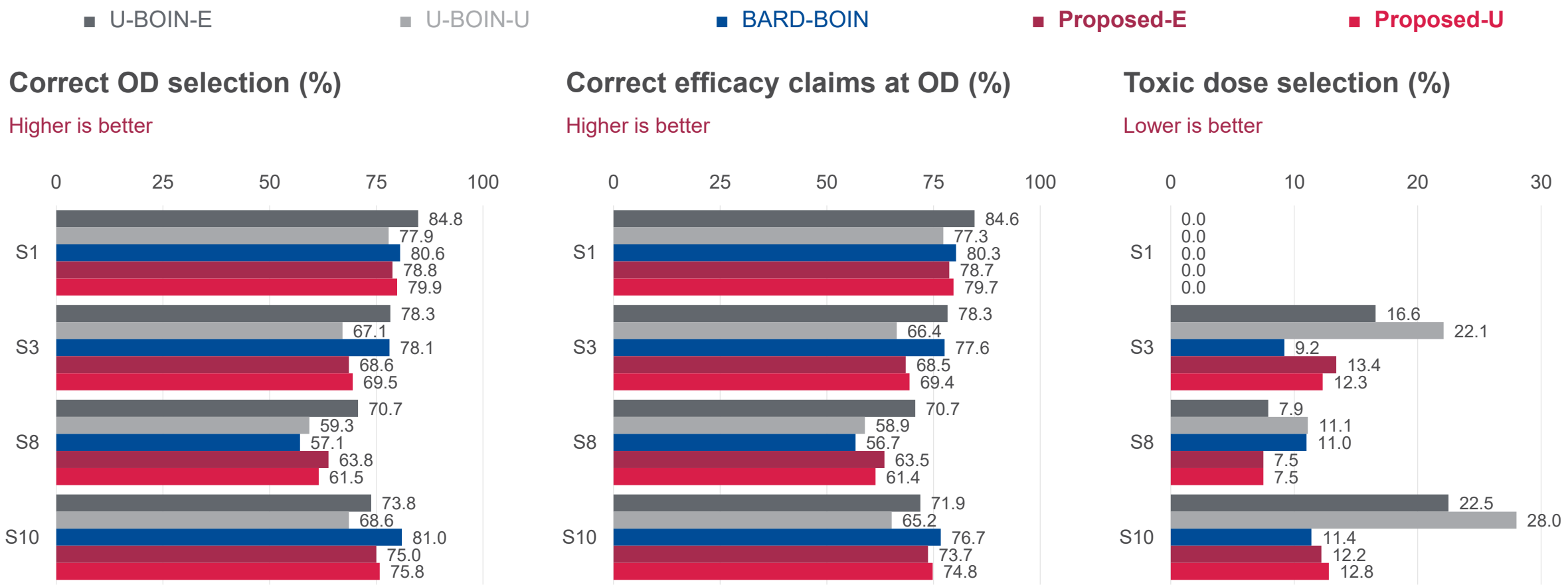
Proposed designs use fewer patients and shorten trials

Both proposed designs outperform all three comparators on patient numbers and duration in every scenario shown.



Proposed-U has the lowest patient numbers and shortest duration, followed by Proposed-E.

OD selection, efficacy claims and toxic dose selection



The proposed designs use fewer patients and shorten trials, but achieve slightly lower correct OD-selection and efficacy-claim rates than the best-performing comparator

Conclusion

- **Seamless phase I/II dose optimization:**
  Our two-stage design jointly monitors efficacy and toxicity, enabling backfilling at lower doses while maintaining dose-escalation priority.

- **Fewer patients and shorter trials, with a trade-off:**
  Across the scenarios shown, the proposed designs required fewer patients and shortened trial duration, with modest reductions in correct OD-selection and efficacy-claim rates relative to the best-performing comparator.

- **Low toxic dose selection:**
  The proposed designs performed well in achieving low toxic dose-selection rates across the scenarios shown.

- **Alignment with modern dose optimization:** The design integrates evidence on efficacy, toxicity, and benefit–risk across multiple dose levels rather than relying on the MTD alone.

Reference

- Takeda K, Zhu J, Yimer BB, Hirakawa A. A seamless two-stage phase I/II trial design with backfill and joint monitoring for dose optimization. Pharm Stat. 2026;25(3):e70092.

- Takeda K, Zhu J, Hirakawa A. BF-BOIN-ET: A backfill Bayesian optimal interval design using efficacy and toxicity outcomes for dose optimization. Pharm Stat. 2025;24(2):e2470.

- Xu X, Hashimoto A, Yimer BB, Takeda K. BOP2-FE: Bayesian optimal phase II design with futility and efficacy-stopping boundaries. J Biopharm Stat. published online: 17 Sep 2025.

- Zhou Y, Lee JJ, Yuan Y. A utility-based Bayesian optimal interval (U-BOIN) phase I/II design to identify the optimal biological dose for targeted and immune therapies. Stat Med. 2019;38(28):5299-5316.