

Precision Generalized Phase I-II Designs

Saijun Zhao

Department of Biostatistics and Health Data Science,
School of Medicine, Indiana University

October 24, 2024

Joint work with Drs. Peter Thall, Ying Yuan, and Pavlos Msaouel from MD Anderson Cancer Center, Juhee Lee from UC Santa Cruz, and Yong Zang from Indiana University.

- 1 Background and Motivation
- 2 Proposed design: PGen I-II design
- 3 Simulation Study

1 Background and Motivation

2 Proposed design: PGen I-II design

3 Simulation Study

Early phase dose-finding designs typically rely on **short-term outcomes**.

- **BOIN design** (phase I): finding the maximum tolerated dose (MTD) based on toxicity.
- **BOIN12 design** (phase I-II): identifies the optimal biological dose (OBD) by considering both toxicity and short-term efficacy.

These designs rely on the assumption that selecting a dose based on short-term outcomes ensures **long-term therapeutic success**.

However, the selected dose may result in **suboptimal** survival due to high relapse rates.

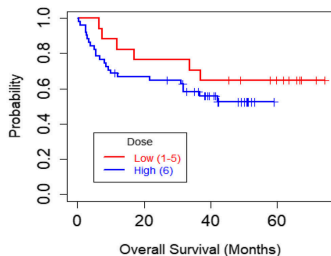


Figure 1: A dose-finding trial that failed.

Thall et al. (2023) proposed a generalized Bayesian phase I-II (**Gen I-II**) design to optimize the **long-term** therapeutic success.

- First, a conventional phase I-II design based on short-term outcomes identifies a **set of candidate doses**.
- Additional patients are then randomized among these candidates, and the final dose is selected based on long-term therapeutic success.

- The trial considered in the Gen I-II design aims to optimize the dose of chimeric antigen receptor (CAR) natural killer (NK) cells for advanced hematologic malignancies.

Table 1: Three Disease Subgroups in the CAR NK Cell Trial

Disease Group	Description
1	AML, ALL, CMML, MDS, and blastic CML
2	Non-T-cell Hodgkin's and non-Hodgkin's lymphoma
3	T-cell non-Hodgkin's lymphoma

AML = acute myelogenous leukemia, ALL = acute lymphocytic leukemia, CMML = chronic myelomonocytic leukemia, MDS = myelodysplastic syndrome, and CML = chronic myelogenous leukemia.

- The Gen I-II design assumed homogeneous outcome distributions across these subgroups and ignored patient **heterogeneity**.

- 1 Background and Motivation
- 2 Proposed design: PGen I-II design
- 3 Simulation Study

To solve this issue, we propose PGen I-II designs to refine Gen I-II design by accounting for patient heterogeneity across subgroups.

Notations:

- $Y_j \in \{0, \dots, L_j - 1\}$ denotes **ordinal** efficacy ($j = E$) and toxicity ($j = T$) outcomes over $[0, t_1]$.
- Y_S denotes the time to failure over the longer follow-up period $[0, t_2]$.
- Dose values d_j for levels $j = 1, \dots, J$.
- $g \in \{1, \dots, G\}$ represents pre-specified subgroups.
- **Latent cluster membership variables** $\mathbf{z} = (z_1, \dots, z_G)$: $z_1 = 1$ and for $g \geq 2$, if subgroup $j \in \{1, \dots, g-1\}$ is in the same cluster as g , set $z_g = z_j$. Otherwise, $z_g = \max_{i \in \{1, \dots, g-1\}} z_i + 1$.
- Thus, if $z_g = z_{g'}$ for $g \neq g'$, subgroups g and g' are combined into the same cluster.

To construct a parametric model for $p(Y_E, Y_T | d, g)$, we adopt the **multivariate probit latent variable approach**.

- (X_E, X_T) denote **real-valued latent variables** following the **bivariate normal distribution**

$$(X_E, X_T) \Big| d, g \sim N_2((\mu_E(d, g), \mu_T(d, g)), \begin{pmatrix} \sigma_{11} & \sigma_{12} \\ \sigma_{12} & \sigma_{22} \end{pmatrix}),$$

where $\sigma_{11} = \sigma_{22} = 1$ and $-1 \leq \sigma_{12} \leq 1$.

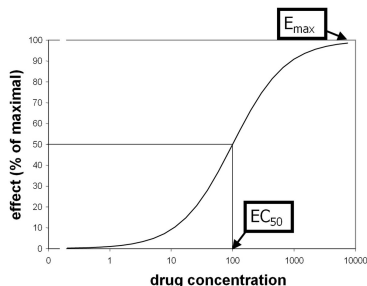
- Given real-valued **cutpoints**, we obtain

$$Y_E = \begin{cases} 0 & \text{if } \eta_0 < X_E < \eta_1 \\ 1 & \text{if } \eta_1 \leq X_E < \eta_2 \\ \dots & \\ L_E - 1 & \text{if } \eta_{L_E-1} \leq X_E < \eta_{L_E} \end{cases}, Y_T = \begin{cases} 0 & \text{if } \zeta_0 < X_T < \zeta_1 \\ 1 & \text{if } \zeta_1 \leq X_T < \zeta_2 \\ \dots & \\ L_T - 1 & \text{if } \zeta_{L_T-1} \leq X_T < \zeta_{L_T} \end{cases},$$

where $\eta_1 = \zeta_1 = 0$ to ensure identifiability.

- For efficacy, we assume the flexible four-parameter Emax model

$$\mu E(d, g) = \alpha_{0,g} + \frac{\alpha_{1,g} d^{\alpha_{3,g}}}{\alpha_{2,g} + d^{\alpha_{3,g}}}$$



- For toxicity, we model more simply as $\mu_T(d, g) = \beta_{0,g} + \beta_{1,g}d$.

- Given an elicited utility function $U(y_E, y_T)$, **the early outcome optimality criterion** for dose d in subgroup g is the mean utility

$$\phi_{ET}(d, g) = \sum_{y_E, y_T} \pi(y_E, y_T | d, g, \theta_{ET}) U(y_E, y_T).$$

- In the CAR NK Cell Trial, Y_S is progression-free survival (PFS) time and $[Y_E = 0]$ means progressive disease (PD).
- Define $Y'_S = (Y_S - t_1) I(Y_E > 0)$ as the treatment failure time starting from t_1 . If $Y_E = 0$, then $Y'_S = 0$.
- The long-term dose optimality criterion** is defined as the long-term success probability

$$\begin{aligned} \phi_S(d, g) &= \Pr(Y_S > t_2 | d, g) = \Pr(Y'_S > t_2 - t_1 | d, g) \\ &= \sum_{y=1}^{L_E-1} \Pr(Y_E = y | d, g) \Pr(Y'_S > t_2 - t_1 | d, g, Y_E = y). \end{aligned}$$

- For robustness, we assume that Y_S follows a **piecewise exponential** (PE) distribution, with the following PE hazard function

$$h_S(t | Y_E^+, Y_T, d, g, \theta_S) = \left\{ \sum_{k=1}^K \lambda_k I(c_{k-1} \leq t < c_k) \right\} \\ \times \exp \left(\gamma_{E, Y_E^+, g} + \gamma_{T, Y_T, g} + \sum_{j=1}^J \gamma_{D, j, g} I(d = d_j) \right), \quad t > 0,$$

where Y_E^+ represents values of $Y_E > 0$, λ_k represents the baseline PE hazard on the k^{th} subinterval and $\gamma_{E, 1, g} = \gamma_{T, 0, g} = \gamma_{D, 1, g} = 0$.

- Let \mathcal{D}_{TE}^n represent the data on (Y_T, Y_E) , \mathcal{D}_S^n represent the data on Y_S , and \mathcal{D}^n represent the complete dataset for the first n patients.

We use the latent cluster membership vector \mathbf{z} to adaptively cluster similar subgroups:

- Each possible configuration of \mathbf{z} represents a model, indexed by $\nu = 1, \dots, M$. If $G = 3$, there are $M = 5$ possible models

$$\nu = 1 \quad \text{if} \quad \mathbf{z} = (1, 1, 1),$$

$$\nu = 2 \quad \text{if} \quad \mathbf{z} = (1, 1, 2),$$

$$\nu = 3 \quad \text{if} \quad \mathbf{z} = (1, 2, 1),$$

$$\nu = 4 \quad \text{if} \quad \mathbf{z} = (1, 2, 2),$$

$$\nu = 5 \quad \text{if} \quad \mathbf{z} = (1, 2, 3).$$

- Assume a discrete uniform prior: $\Pr(\nu = k) = 1/5$, $k = 1, \dots, 5$.
- Due to the adaptive model dimension change, we use **Reversible-Jump Markov Chain Monte Carlo (RJMCMC)** to obtain posterior samples on spaces of varying dimensions.

The following criteria will be used by the PGen I-II design:

- **The early-term dose optimality criteria:**

$$\hat{\phi}_{ET}(d, g | \mathcal{D}_{ET}^n) = \int \phi_{ET}(d, g | \theta_{ET}^*, z_g) p(\theta_{ET}^*, z_g | \mathcal{D}_{ET}^n) d\theta_{ET}^* dz_g.$$

- **The long-term dose optimality criteria:**

$$\hat{\phi}_S(d, g | \mathcal{D}^n) = \int \phi_S(d, g | \theta) p(\theta | \mathcal{D}^n) d\theta.$$

- The posterior predictive probability that the long-term optimality criterion for dose d in subgroup g is greater than $\underline{\phi}_S$ is given by:

$$\hat{\Pr} \left\{ \phi_S(d, g | \mathcal{D}^n) > \underline{\phi}_S | \mathcal{D}^n \right\} = \int \Pr \left\{ \phi_S(d, g) > \underline{\phi}_S | \theta \right\} p(\theta | \mathcal{D}^n) d\theta.$$

- For $j = E, T$ and each (d, g) , denote the vector of L_j marginal early outcome probabilities by:

$$\boldsymbol{\pi}_j(d, g, \boldsymbol{\theta}_j) = (\pi_j(0 | d, g, \boldsymbol{\theta}_j), \pi_j(1 | d, g, \boldsymbol{\theta}_j), \dots, \pi_j(L_j - 1 | d, g, \boldsymbol{\theta}_j))'$$

- We say that a dose d is acceptable for subgroup g if:

$$\begin{aligned} \Pr \{ \mathbf{b}'_E \boldsymbol{\pi}_E(d, g, \boldsymbol{\theta}_E) > \underline{\pi}_E \mid \mathcal{D}_{ET}^n \} &> p_E, \\ \Pr \{ \mathbf{b}'_T \boldsymbol{\pi}_T(d, g, \boldsymbol{\theta}_T) < \bar{\pi}_T \mid \mathcal{D}_{ET}^n \} &> p_T, \end{aligned} \quad (1)$$

where \mathbf{b}_E and \mathbf{b}_T are design parameter vectors with all entries 0 or 1, and $\underline{\pi}_E$ and $\bar{\pi}_T$ are prespecified fixed limits.

- In our illustration, $\mathbf{b}_E = (0, \dots, 0, 1)'_{L_E}$ and $\mathbf{b}_T = (0, \dots, 0, 1)'_{L_T}$.

The PGen I-II design has **three stages**:

- the sample size for each stage: n_s ;
- the overall sample size is $N = n_1 + n_2 + n_3$;
- the sample size for subgroup g in stage s : $n_{s,g}$;
- the sample size for subgroup g with dose d_j : $n_s(d_j)$.

Stage 1:

- Due to limited within-subgroup sample sizes, the PGen I-II design temporarily **ignores subgroups**, using the **BOIN12 design** to assign doses and determine **an acceptable dose set** $\mathcal{A}_{ET}^{n_1}$ for all subgroups combined.
- Other phase I-II designs can also be used in stage 1.

Stage 2:

- For each subgroup g , $n_{2,g}$ additional patients are randomized fairly among the acceptable doses in $\mathcal{A}_{ET}^{n_1}$.
- At the end of Stage 2, for each g , a subgroup-specific acceptable dose set $\mathcal{A}_{ET,g}^{n_{12}}$ is determined using the criteria in (1). The acceptable dose set in subgroup g is defined as

$$\mathcal{A}_{S,g}^{n_{12}} = \left\{ d_j \in \mathcal{A}_{ET,g}^{n_{12}} : \Pr \{ \phi_S(d_j, g) > \underline{\phi}_S \mid \mathcal{D}^{n_{12}} \} > p_{S,L} \right\}.$$

- The maximum over the set of acceptable doses for subgroup g is

$$\hat{\phi}_{ET,g}^{max} = \max_{d_j \in \mathcal{A}_{ET,g}^{n_{12}}} \hat{\phi}_{ET}(d_j, g \mid \mathcal{D}^{n_{12}}).$$

A **candidate dose set for subgroup g** then is defined as

$$\mathcal{C}_g^{n_{12}} = \left\{ d_j \in \mathcal{A}_{S,g}^{n_{12}} : \hat{\phi}_{ET}(d_j, g \mid \mathcal{D}^{n_{12}}) \geq \rho \hat{\phi}_{ET,g}^{max} \right\}.$$

Stage 3:

- For each subgroup g , $n_{3,g}$ additional patients are randomized among the doses in $C_g^{n_{12}}$ and followed to time t_2 .
- Denote $n_{123,g} = n_{1,g} + n_{2,g} + n_{3,g}$.
- The updated candidate dose set C_g^N must satisfy both the toxicity requirement in (1) and the minimal long-term success probability requirement, formally

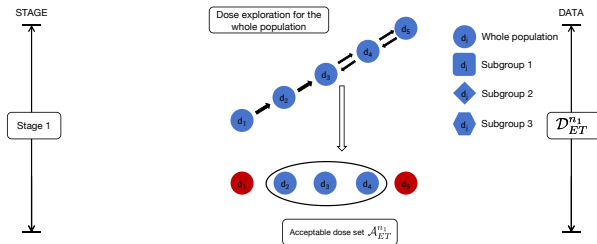
$$C_g^N = \left\{ d_j \in C_g^{n_{12}} : \Pr \left\{ \mathbf{b}'_T \pi_T(d_j, g, \boldsymbol{\theta}_T) < \bar{\pi}_T \mid \mathcal{D}^N \right\} > p_{C,T,L} \right. \\ \left. \text{and } \Pr \left\{ \phi_S(d_j, g) > \underline{\phi}_S \mid \mathcal{D}^N \right\} > p_{C,S,L} \right\}.$$

In our simulated design, we used $p_{C,T,L} = .10$ and $p_{C,S,L} = .50$.

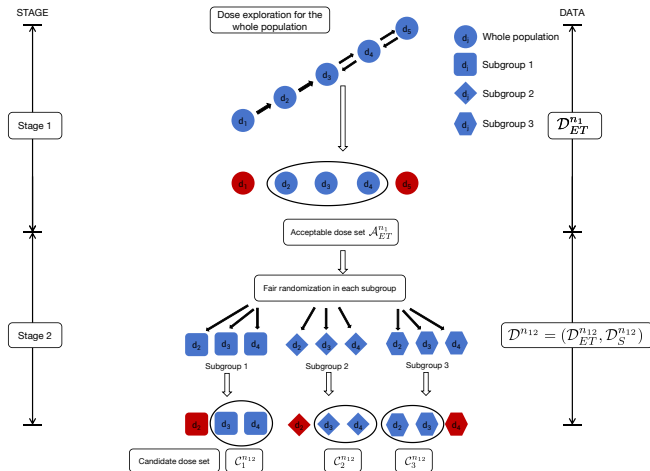
- **The optimal dose** for that subgroup is the acceptable dose that maximizes the estimated long-term benefit criterion,

$$d_j^{\text{opt},g} = \underset{d_j \in C_g^N}{\operatorname{argmax}} \hat{\phi}_S(d_j, g \mid \mathcal{D}^N). \quad (2)$$

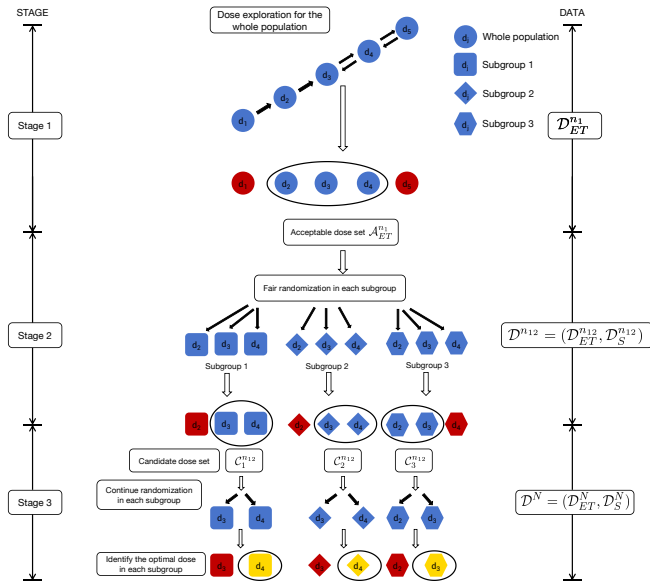
Schematic for the PGen I-II design



Schematic for the PGen I-II design



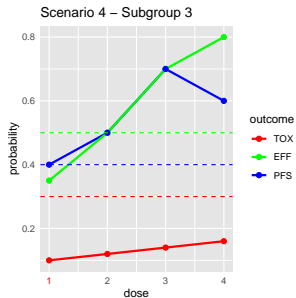
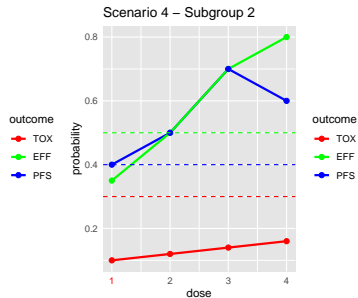
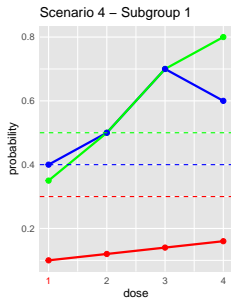
Schematic for the PGen I-II design



1 Background and Motivation

2 Proposed design: PGen I-II design

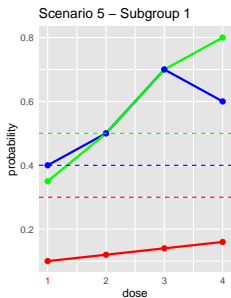
3 Simulation Study



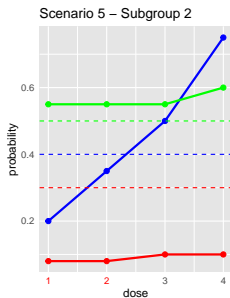
Scenario 4

	Subgroup z_g	g=1					g=2					g=3				
		1					1					1				
		d_0	d_1	d_2	d_3	d_4	d_0	d_1	d_2	d_3	d_4	d_0	d_1	d_2	d_3	d_4
	Dose (d)															
	$\pi_T^{\text{true}}(1 d, g)$	0.10	0.12	0.14	0.16		0.10	0.12	0.14	0.16		0.10	0.12	0.14	0.16	
	$\pi_E^{\text{true}}(1 d, g)$	0.30	0.30	0.20	0.15		0.30	0.30	0.20	0.15		0.30	0.30	0.20	0.15	
	$\pi_E^{\text{true}}(2 d, g)$	0.35	0.50	0.70	0.80		0.35	0.50	0.70	0.80		0.35	0.50	0.70	0.80	
	$\phi_{ET}^{\text{true}}(d, g)$	54.0	65.1	76.9	82.5		54.0	65.1	76.9	82.5		54.0	65.1	76.9	82.5	
	$\phi_S^{\text{true}}(d, g)$	0.40	0.50	0.70	0.60		0.40	0.50	0.70	0.60		0.40	0.50	0.70	0.60	
PGen I-II	Selection %	2.3	4.1	6.4	78.8	8.4	2.1	3.6	4.7	81.1	8.5	2.2	4.5	4.4	81.4	7.5
	Patients		8.1	11.1	13.6	11.5		8.5	11.1	13.5	11.5		11.0	14.5	18.1	15.2
PGen I-II-Comb	Selection %	2.2	2.2	5.5	81.2	8.9	2.2	2.2	5.5	81.2	8.9	2.2	2.2	5.5	81.2	8.9
	Patients		7.7	11.3	13.7	11.6		8.0	11.2	13.8	11.5		10.6	14.7	18.3	15.3
PGen I-II-ET	Selection %	2.0	1.2	4.5	14.1	78.2	2.0	1.0	4.5	14.2	78.3	1.9	1.6	4.2	13.7	78.6
	Patients		8.1	11.2	13.5	11.6		8.4	11.1	13.5	11.5		11.0	14.4	17.9	15.4
PGen I-II-Sep	Selection %	2.3	11.7	17.3	58.8	9.9	2.5	14.4	17.6	56.3	9.2	2.1	11.4	15.3	61.9	9.3
	Patients		8.3	11.1	13.4	11.5		8.6	11.0	13.5	11.3		11.2	14.6	17.9	15.2

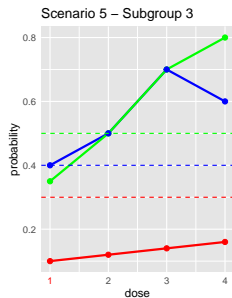
Simulation Settings



outcome
TOX
EFF
PFS



outcome
TOX
EFF
PFS



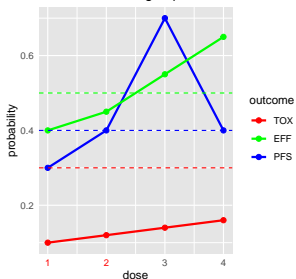
outcome
TOX
EFF
PFS

Scenario 5

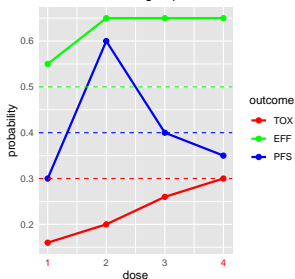
	Subgroup z_g	g=1					g=2					g=3				
		1					2					1				
		d_0	d_1	d_2	d_3	d_4	d_0	d_1	d_2	d_3	d_4	d_0	d_1	d_2	d_3	d_4
	Dose (d)															
	$\pi_T^{\text{true}}(1 d, g)$	0.10	0.12	0.14	0.16		0.08	0.08	0.10	0.10		0.10	0.12	0.14	0.16	
	$\pi_E^{\text{true}}(1 d, g)$	0.30	0.30	0.20	0.15		0.40	0.40	0.40	0.35		0.30	0.30	0.20	0.15	
	$\pi_E^{\text{true}}(2 d, g)$	0.35	0.50	0.70	0.80		0.55	0.55	0.55	0.60		0.35	0.50	0.70	0.80	
	$\phi_{ET}^{\text{true}}(d, g)$	54.0	65.1	76.9	82.5		73.3	73.3	72.6	75.0		54.0	65.1	76.9	82.5	
	$\phi_S^{\text{true}}(d, g)$	0.40	0.50	0.70	0.60		0.20	0.35	0.50	0.75		0.40	0.50	0.70	0.60	
PGen I-II	Selection %	2.4	7.7	10.3	67.3	12.3	5.3	2.2	4.7	24.0	63.8	2.3	7.7	10.1	67.1	12.8
	Patients		9.5	11.6	12.8	10.6		9.5	11.4	13.1	10.8		12.7	15.2	17.3	14.1
PGen I-II-Comb	Selection %	4.0	2.0	4.0	64.0	26.0	4.0	2.0	4.0	64.0	26.0	4.0	2.0	4.0	64.0	26.0
	Patients		9.1	12.4	12.2	10.6		9.5	12.1	12.8	10.9		11.6	15.5	16.7	14.6
PGen I-II-ET	Selection %	1.7	4.3	5.4	18.8	69.8	1.0	10.9	8.6	18.4	61.1	1.5	3.8	5.8	18.5	70.4
	Patients		9.5	11.6	12.8	10.6		10.3	11.5	12.6	10.6		12.7	15.1	17.3	14.1
PGen I-II-Sep	Selection %	2.1	13.5	14.9	59.5	10.0	6.1	3.0	4.1	18.7	68.1	2.4	11.6	15.9	61.8	8.3
	Patients		9.6	11.6	12.7	10.6		9.4	11.4	13.1	10.9		12.8	15.1	17.4	14.1

Simulation Settings

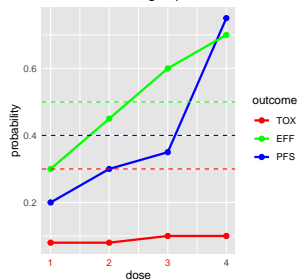
Scenario 8 – Subgroup 1



Scenario 8 – Subgroup 2



Scenario 8 – Subgroup 3



Scenario 8

	Subgroup z_g	g=1					g=2					g=3				
		1					2					3				
		d_0	d_1	d_2	d_3	d_4	d_0	d_1	d_2	d_3	d_4	d_0	d_1	d_2	d_3	d_4
	Dose(d)															
	$\pi_T^{\text{true}}(1 d, g)$	0.10	0.12	0.14	0.16		0.16	0.20	0.26	0.30		0.08	0.08	0.10	0.10	
	$\pi_E^{\text{true}}(1 d, g)$	0.20	0.20	0.20	0.15		0.35	0.30	0.30	0.30		0.30	0.30	0.25	0.20	
	$\pi_E^{\text{true}}(2 d, g)$	0.40	0.45	0.55	0.65		0.55	0.65	0.65	0.65		0.30	0.45	0.60	0.70	
	$\phi_{ET}^{\text{true}}(d, g)$	54.9	58.2	65.3	70.9		69.2	74.0	72.0	70.6		50.7	62.4	72.0	78.4	
	$\phi_S^{\text{true}}(d, g)$	0.30	0.40	0.70	0.40		0.30	0.60	0.40	0.35		0.20	0.30	0.35	0.75	
PGen I-II	Selection %	12.0	2.5	9.6	70.0	5.9	14.4	5.0	64.2	11.1	5.3	18.3	0.8	2.7	5.7	72.5
	Patients		9.4	11.3	12.3	10.7		9.6	11.8	12.0	10.8		11.5	14.2	16.3	16.6
PGen I-II-Comb	Selection %	8.7	0.8	14.8	41.8	33.9	8.7	0.8	14.8	41.8	33.9	8.7	0.8	14.8	41.8	33.9
	Patients		8.5	11.3	12.6	11.1		8.6	11.1	12.8	11.3		11.7	15.1	16.8	15.2
PGen I-II-ET	Selection %	2.8	10.1	11.8	14.3	61.0	3.1	21.4	15.9	16.4	43.2	4.3	5.3	14.0	13.8	62.6
	Patients		9.7	11.1	11.9	11.0		10.0	11.1	12.0	10.9		12.2	14.9	16.4	15.3
PGen I-II-Sep	Selection %	10.5	3.2	8.8	76.1	1.4	12.3	4.8	69.5	11.1	2.3	18.4	0.9	1.5	3.7	75.5
	Patients		9.6	11.3	12.3	10.7		9.7	11.9	12.1	10.6		11.5	14.0	16.2	16.7

- Liu, S., & Yuan, Y. (2015). Bayesian optimal interval designs for phase I clinical trials. *Journal of the Royal Statistical Society: Series C: Applied Statistics*, 64(3), 507-523.
- Lin, R., Zhou, Y., Yan, F., Li, D., and Yuan, Y. (2020). BOIN12: Bayesian optimal interval phase I/II trial design for utility-based dose finding in immunotherapy and targeted therapies. *JCO Precision Oncology*, 4, 1393-1402.
- Thall, P.F., Zang, Y., and Yuan, Y. (2023). Generalized phase I-II designs to increase long term therapeutic success rate. *Pharmaceutical Statistics*, 22, 692-706.
- Green, P. J. (1995). Reversible jump Markov chain Monte Carlo computation and Bayesian model determination. *Biometrika*, 82(4), 711-732.

Thank you for your attention.

Please feel free to ask any questions.