

# **BOP2: A Bayesian Optimal Design for Phase 2 Clinical Trials with simple and complex endpoints**

---

Ying Yuan

Professor, Chief of Section of Adaptive Clinical Trials  
MD Anderson Cancer Center

Joint work with Heng Zhou and Jack Lee

# Phase II Trial Design

---

- The key question of phase II trial design is how to make accurate go/no-go decision based on interim data.
- Simon's optimal two-stage design (1989) is the most commonly used phase II design.
  - Pros: simple to implement & controls type I error
  - Cons: restrictive in the number and timing of interims, and assumes a simple binary endpoint

# New Challenges

---

- Recent developments in immunotherapy and molecularly targeted agents have made the endpoint of phase II trials more complicated:
  - Nested efficacy endpoint
  - Co-primary efficacy endpoint
  - Considering toxicity and efficacy jointly
- In some applications, it is beneficial to perform more than 1 interim to improve trial efficiency
  - Platform and basket trials

# Example 1: Binary efficacy endpoint

---

- Pembrolizumab in treating patients with small bowel adenocarcinoma.
- Endpoint: the objective response rate , defined by RECIST version 1.1.
- The treatment is regarded as futile if  $ORR \leq 0.2$ , and promising if  $ORR \geq 0.4$ .

## Example 2: Nested Endpoint

---

- Refractory / relapsed acute myeloid leukemia (AML) treated with combination of nivolumab and 5-azacytidine.
- Response (CR/PR/SD/DP) is scored using the response criteria modified by the International Working Group.
- Target:  $CR + PR \geq 30\%$  or  $CR \geq 15\%$

Reference: Cheson, B. D., *et al* (2003). Revised recommendations of the international working group for diagnosis, standardization of response criteria, treatment outcomes, and reporting standards for therapeutic trials in acute myeloid leukemia. *Journal of Clinical Oncology*, 21(24), 4642-4649.

## Example 3: Co-primary Endpoint

---

- A phase II trial of trebananib in patients with persistent / recurrent carcinoma of the endometrium.
- Endpoints: objective response rate (ORR) and progression free survival at 6 months (PFS6).
- Null Hypothesis:  $ORR \leq 10\%$  and  $PFS6 \leq 20\%$ .
- Alternative Hypothesis:  $ORR \geq 30\%$  or  $PFS6 \geq 35\%$ .

Reference: Moore, K. N., *et al* (2015). A phase II trial of trebananib (AMG 386; IND \# 111071), a selective angiopoietin 1/2 neutralizing peptibody, in patients with persistent / recurrent carcinoma of the endometrium: An NRG / Gynecologic Oncology Group trial. *Gynecologic oncology*. 138(3), 513-518.

## Example 4: Efficacy and Toxicity

---

- Safety and efficacy of lenalidomide in combination with rituximab in recurrent indolent non-follicular lymphoma.
- Primary endpoints: ORR and toxicity rate
- Target: ORR  $\geq 45\%$  and toxicity rate  $\leq 30\%$ .

Sacchi, S., *et al* (2016). Safety and efficacy of lenalidomide in combination with rituximab in recurrent indolent non-follicular lymphoma: final results of a phase II study conducted by the Fondazione Italiana Linfomi. *Haematologica*, haematol-2015.

# BOP2: Bayesian Optimal Phase 2 Design

---

- BOP2 design provides a unified framework to handle all aforementioned trials
- Explicitly controls the type I error rate, thereby bridging the gap between Bayesian designs and frequentist designs
- Optimal by (i) maximizing power, given a fixed  $N$  and type I error; or (ii) minimizing the  $E(N|H_0)$ , given fixed type I and II error rates



# Notation

---

- Let  $Y$  denote the primary endpoint of a phase II trial, which is a multinomial variable with  $K$  categories.
  - Example 1: 1 = response, 2 = no response.
  - Example 2: 1 = CR, 2 = PR, 3 = SD and 4 = PD.
  - Example 3: 1 = (OR, PFS6), 2 = (OR, no PFS6), 3 = (no OR, PFS6) and 4 = (no OR, no PFS6).
  - Example 4: 1 = (toxicity, OR), 2 = (no toxicity, OR), 3 = (toxicity, no OR) and 4 = (no toxicity, no OR).
- Let  $\theta = (\theta_1, \dots, \theta_K)^T$  denote the probability that  $Y$  belongs to each category.

# Model

---

- Dirichlet-multinomial model

$$Y|\boldsymbol{\theta} \sim \text{Multinom}(\theta_1, \dots, \theta_K)$$

$$(\theta_1, \dots, \theta_K) \sim \text{Dir}(a_1, \dots, a_K)$$

where  $a_1, \dots, a_K$  are hyperparameters.

- Let  $D_n = (x_1, \dots, x_K)$  denote the interim data from  $n$  enrolled patients, where  $x_k$  denote the number of patients with  $Y = k$ .

- Posterior distribution of  $\boldsymbol{\theta}$  is given by

$$\boldsymbol{\theta}|D_n \sim \text{Dir}(a_1 + x_1, \dots, a_K + x_K)$$

where  $\sum_{k=1}^K a_k = 1$  such that the prior is vague and equivalent to a prior sample size of 1.

# BOP2 Design

---

- The BOP2 design consists of  $R$  interim looks, which occur when the number of enrolled patients reaches  $n_1, \dots, n_R$ , and a final look when all  $N$  patients are enrolled.
- At each of these looks, the go/no-go decision is made based on the posterior probabilities of one or more linear combinations of the model parameters  $\theta$  in the form of

$$\Pr(\mathbf{b}\theta \leq \phi | D_n) > C(n),$$

where  $\mathbf{b}$  is a design vector with elements of 0 and 1,  $\phi$  is a prespecified limit, and the cutoff  $C(n)$  is a function of the interim sample size  $n$ .

# Interim Go/No-Go Rule

---

- Example 2: Nested endpoint

If  $\Pr(\theta_1 \leq 0.15 | D_n) > C(n)$  and

$\Pr(\theta_1 + \theta_2 \leq 0.3 | D_n) > C(n)$ , then stop;

otherwise go.

- ◆ Recall  $(\theta_1, \theta_2, \theta_3, \theta_4) = (\Pr(\text{CR}), \Pr(\text{PR}), \Pr(\text{SD}), \Pr(\text{DP}))$ .

# Interim Go/No-Go Rule

---

- Example 2: Nested endpoint

If  $\Pr(\theta_1 \leq 0.15 | D_n) > C(n)$  and

$\Pr(\theta_1 + \theta_2 \leq 0.3 | D_n) > C(n)$ , then stop;  
otherwise go.

- ◆ Recall  $(\theta_1, \theta_2, \theta_3, \theta_4) = (\Pr(\text{CR}), \Pr(\text{PR}), \Pr(\text{SD}), \Pr(\text{DP}))$ .

# Interim Go/No-Go Rule

---

- Example 2: Nested endpoint

If  $\Pr(\theta_1 \leq 0.15 | D_n) > C(n)$  and

$\Pr(\theta_1 + \theta_2 \leq 0.3 | D_n) > C(n)$ , then stop;  
otherwise go.

- ◆ Recall  $(\theta_1, \theta_2, \theta_3, \theta_4) = (\Pr(\text{CR}), \Pr(\text{PR}), \Pr(\text{SD}), \Pr(\text{DP}))$ .

# Interim Go/No-Go Rule

---

## ■ Example 2: Nested endpoint

If  $\Pr(\theta_1 \leq 0.15 | D_n) > C(n)$  and  
 $\Pr(\theta_1 + \theta_2 \leq 0.3 | D_n) > C(n)$ , then stop;  
otherwise go.

◆ Recall  $(\theta_1, \theta_2, \theta_3, \theta_4) = (\Pr(\text{CR}), \Pr(\text{PR}), \Pr(\text{SD}), \Pr(\text{DP}))$ .

## ■ Example 3: Co-primary endpoint

If  $\Pr(\theta_1 + \theta_2 \leq 0.1 | D_n) > C(n)$  and  
 $\Pr(\theta_1 + \theta_3 \leq 0.2 | D_n) > C(n)$ , then stop;  
otherwise go.

● Recall  $(\theta_1, \theta_2, \theta_3, \theta_4) = (\Pr(\text{OR, PFS6}), \Pr(\text{OR, no PFS6}), \Pr(\text{no OR, PFS6}), \Pr(\text{no OR, no PFS6}))$ .

# Interim Go/No-Go Rule

---

## ■ Example 2: Nested endpoint

If  $\Pr(\theta_1 \leq 0.15 | D_n) > C(n)$  and  
 $\Pr(\theta_1 + \theta_2 \leq 0.3 | D_n) > C(n)$ , then stop;  
otherwise go.

◆ Recall  $(\theta_1, \theta_2, \theta_3, \theta_4) = (\Pr(\text{CR}), \Pr(\text{PR}), \Pr(\text{SD}), \Pr(\text{DP}))$ .

## ■ Example 3: Co-primary endpoint

If  $\Pr(\theta_1 + \theta_2 \leq 0.1 | D_n) > C(n)$  and  
 $\Pr(\theta_1 + \theta_3 \leq 0.2 | D_n) > C(n)$ , then stop;  
otherwise go.

- Recall  $(\theta_1, \theta_2, \theta_3, \theta_4) = (\Pr(\text{OR, PFS6}), \Pr(\text{OR, no PFS6}), \Pr(\text{no OR, PFS6}), \Pr(\text{no OR, no PFS6}))$ .



# Optimizing Design Parameters

---

- The posterior probability cutoff  $C(n)$  used for stopping adaptively changes with the interim sample size  $n$

$$C(n) = 1 - \lambda \left(\frac{n}{N}\right)^\gamma$$

- Cutoff parameters  $\lambda$  and  $\gamma$  are optimized such that the power is maximized, given prespecified total sample size  $N$  and type I error rate
- Optimization is done using grid search

# Statistical Properties

- **Property 1.** Given  $\boldsymbol{\theta} \sim \text{Dir}(a_1 + x_1, \dots, a_K + x_K)$  and a design vector  $\mathbf{b} = (b_1, \dots, b_K)$  with elements of 0 and 1,  $\mathbf{b}\boldsymbol{\theta}$  follows a Beta distribution  $\text{Beta}(\sum_{k=1}^K b_k(a_k + x_k), \sum_{k=1}^K (1 - b_k)(a_k + x_k))$ .
  - $\Pr(\mathbf{b}\boldsymbol{\theta} \leq \phi | D_n)$  can be easily evaluated.
- **Lemma 1.**  $\Pr(\mathbf{b}\boldsymbol{\theta} \leq \phi | D_n)$  is a monotonic function of  $\sum_{k=1}^K b_k x_k$ .
  - The stopping boundary can be enumerated prior to the onset of the trial !!

# Stopping Boundaries of BOP2 Design

Trial	Stop the trial if	Number of patients treated						
		10	15	20	25	30	35	40
Example 1	# of OR $\leq$	1	2	4	5	7	9	10
Example 2	and # of CR $\leq$ # of CR/PR $\leq$	0	1	3	4	5	7	9
		2	3	5	8	10	13	16
Example 3	and # of OR $\leq$ # of PFS6 $\leq$	0	1	2	3	4	5	7
		1	2	4	5	7	9	12
Example 4	or # of OR $\leq$ # of Toxicities $\geq$	2	5	7	10	13	16	19
		5	6	8	9	10	11	12

OR: objective response

1.  $H_0: \Pr(OR) = 0.2; H_1: \Pr(OR) = 0.4.$
2.  $H_0: \Pr(CR) = 0.15, \Pr(CR/PR) = 0.3; H_1: \Pr(CR) = 0.25, \Pr(CR/PR) = 0.5.$
3.  $H_0: \Pr(OR) = 0.1, \Pr(PFS6m) = 0.2; H_1: \Pr(OR) = 0.3, \Pr(PFS6m) = 0.35.$
4.  $H_0: \Pr(OR) = 0.45, \Pr(Toxicity) = 0.3; H_1: \Pr(OR) = 0.6, \Pr(Toxicity) = 0.2.$

# Stopping Boundaries of BOP2 Design

Trial	Stop the trial if	Number of patients treated						
		10	15	20	25	30	35	40
Example 2	and # of CR $\leq$ # of CR/PR $\leq$	0 2	1 3	3 5	4 8	5 10	7 13	9 16

OR: objective response

# Stopping Boundaries of BOP2 Design

Trial	Stop the trial if	Number of patients treated							
		10	15	20	25	30	35	40	
Example 3	and	# of OR $\leq$ # of PFS6 $\leq$	0 1	1 2	2 4	3 5	4 7	5 9	7 12

OR: objective response

# Simulation Results: Binary Efficacy Endpoint

**Table:** Operating characteristics under the BOP2 design and TS design (Thall and Simon, 1994) with binary efficacy endpoint. The interims occur when  $n = 10, 15, 20, 25, 30, 35, 40$ .

Response rate (ORR)	Claim promising(%)		Early termination(%)		Sample size	
	BOP2	TS	BOP2	TS	BOP2	TS
0.20 <sup>§</sup>	9.6	9.4	88.8	89.8	20.2	15.3
0.30	55.2	42.6	46.2	56.7	31.0	24.9
0.40 <sup>†</sup>	<b>88.3</b>	<b>76.4</b>	11.4	23.5	37.6	33.6
0.50	98.2	93.3	1.8	6.7	39.5	38.1

<sup>§</sup>: null hypothesis; <sup>†</sup>: alternative hypothesis.

# Simulation Results: Nested Endpoints

**Table:** Operating characteristics under the BOP2 design and TS design (Thall and Simon, 1994) with nested efficacy endpoints. The interims occur when  $n = 10, 15, 20, 25, 30, 35, 40$ .

(CR, CR/PR)	Claim promising(%)		Early termination(%)		Sample size	
	BOP2	TS	BOP2	TS	BOP2	TS
(0.15, 0.30) <sup>§</sup>	8.7	9.9	82.1	89.6	25.4	15.7
(0.20, 0.30)	24.2	9.6	63.8	89.9	29.0	15.6
(0.25, 0.45)	72.3	59.0	19.3	40.9	37.1	29.1
(0.25, 0.50) <sup>†</sup>	<b>85.5</b>	<b>74.2</b>	9.9	25.7	38.5	33.0
(0.30, 0.55)	95.7	85.2	3.0	14.8	39.5	35.9

<sup>§</sup>: null hypothesis; <sup>†</sup>: alternative hypothesis.

# Simulation Results: Co-primary Endpoints

**Table:** Operating characteristics under the BOP2 design and TSE design (Thall, Simon and Estey, 1995) with two co-primary efficacy endpoints. The interims occur when  $n = 10, 15, 20, 25, 30, 35, 40$ .

(ORR, PFS 6m)	Claim promising(%)		Early termination(%)		Sample size	
	BOP2	TSE	BOP2	TSE	BOP2	TSE
(0.10, 0.20) <sup>§</sup>	7.2	7.3	80.9	92.3	24.5	13.7
(0.15, 0.20)	23.9	17.4	58.9	82.3	29.7	16.5
(0.25, 0.30)	85.9	60.7	7.6	39.3	38.7	28.5
(0.30, 0.35) <sup>†</sup>	<b>96.1</b>	<b>75.5</b>	2.4	24.5	39.5	32.8
(0.30, 0.40)	98.5	82.6	0.8	17.4	39.8	34.8

<sup>§</sup>: null hypothesis; <sup>†</sup>: alternative hypothesis.



[www.trialdesign.org](http://www.trialdesign.org)

stage.trialdesign.org

**INTEGRATED PLATFORM  
FOR DESIGNING  
CLINICAL TRIALS**

RESEARCH · EDUCATION · INNOVATION

**PHASE I-II**

$s^2 = \frac{1}{n} \sum (x_i - \bar{x})^2$

$s^2 = \frac{pq}{(n-1)}$

$\bar{x} = \frac{\sum x_i}{N}$

$b_1 = r \cdot (s_y / s_x)$

# www.trialdesign.org

The screenshot shows the trialdesign.org website interface. At the top, there is a navigation bar with links for HOME, SOFTWARE, OUR TEAM, PUBLICATIONS, and CONTACT. Below this is a filter bar with tabs for ALL, PHASE I, PHASE II, PHASE I-II, BASKET & PLATFORM, SAMPLE SIZE CALCULATION, and EDUCATION. The main content area displays a grid of 15 trial design software options, each with a circular icon, a title, a brief description, and a 'more...' link. Some options also feature a download icon. The options are:

- BO** BOIN Suite: Bayesian optimal interval (BOIN) designs provide a novel platform to design phase *more...*
- CRM** CRM & BMA-CRM: The continual reassessment method (CRM) is a model-based dose-finding approach *more...*
- KB** Keyboard Design: The keyboard design provides an upgrade to the modified toxicity probability *more...*
- TTK** Time-to-Event Keyboard: The time-to-event keyboard design can handle toxicity data that are pending due *more...*
- S2S** Simon's Two Stage Design: The Simon's two stage design is a commonly used phase II design. It controls type 1 *more...*
- BOP** Bayesian Optimal Phase 2 (BOP2) Design: The Bayesian optimal phase II (BOP2) design is a flexible Bayesian design that allows *more...*
- TOP** Time-to-Event Bayesian Optimal Phase II Trial Design: The time-to-event Bayesian Optimal Phase II (TOP) design is a flexible and efficient design for phase II clinical trials. *more...*
- PP** Bayesian Efficacy Monitoring with Predictive Probability: Bayesian efficacy monitoring with options of early futility *more...*
- DL** Bayesian Phase 2 Design with Delayed Outcomes: One practical impediment in adaptive phase II trials is that outcomes must be observed soon enough *more...*
- TM** Bayesian Toxicity Monitoring: Bayesian toxicity monitoring for evaluating drug safety.
- PO** Bayesian Efficacy Monitoring with Posterior Probability: Bayesian efficacy monitoring with options of early futility and/or efficacy stopping using posterior probability.
- OBD** Find Optimal Biological Dose for Immunotherapy: **Down for maintenance. Sorry for the inconvenience.**
- CBH** Calibrated Bayesian Hierarchical Model Design: Bayesian hierarchical modeling has been proposed to adaptively borrow information across cancer *more...*
- BST** Bayesian Latent Subgroup Design for Basket Trials: The innovation of the BLAST design is that it adaptively clusters cancer types within a basket *more...*
- PAS** Bayesian Drug Combination Platform Trial Design with Adaptive Shrinkage: ComPAS provides a flexible Bayesian screen design to efficiently screen a large number of drug *more...*

# BOP2 App: Binary Endpoint

The screenshot shows a web browser window with the URL `ibl.mdanderson.org/BOP2/`. The page title is "BOP2: Bayesian Optimal Phase II Design with Simple and Complex Endpoints". Below the title, it states "Version: V1.0.0 ; Last Updated: 06/20/2017" and lists the authors "Heng Zhou, Ying Yuan and Jack J. Lee" from the "Department of Biostatistics, MD Anderson Cancer Center".

The interface has four tabs: "Trial Setting" (selected), "Simulation", "Protocol", and "Reference".

The "Trial Setting" tab is active and contains several sections:

- Endpoints:** A list of endpoint types with radio buttons:
  - Binary Efficacy
  - Binary Toxicity
  - Efficacy & Toxicity
  - Co-Primary Efficacy
  - Ordinal Efficacy
- Interims:** A section with a "Help" icon and a text input field containing "10 20 35 50". Below the input is the instruction: "Sample sizes when interim analyses to be performed, seperated by space. The last number must be the total sample size."
- Null Hypothesis:** A section with a "Help" icon and a text input field.
- Response Rate:** A section with a text input field.

To the right of the "Trial Setting" tab, there is a "Stopping boundaries" section with a text input field.

The browser's address bar shows the URL `ibl.mdanderson.org/BOP2/#tab-5385-2`.

# BOP2 App: Binary Endpoint

BOP2: Bayesian Optimal

ibbl.mdanderson.org/BOP2/

**Endpoints:**

- Binary Efficacy
- Binary Toxicity
- Efficacy & Toxicity
- Co-Primary Efficacy
- Ordinal Efficacy

**Interims:** [Help](#)

Sample sizes when interim analyses to be performed, separated by space. The last number must be the total sample size.

10 20 35 50

**Null Hypothesis:** [Help](#)

**Response Rate**

0.2

**Alternative Hypothesis:** [Help](#)

**Response Rate**

0.4

**Type I error rate:**

0.1

Calculate stopping boundaries

Stopping boundaries

Calculation in progress This may take a while...

# BOP2 App: Binary Endpoint

The screenshot shows the BOP2 App interface for a binary endpoint trial. The left sidebar contains input fields for trial parameters, and the main area displays the calculated stopping boundaries.

**Endpoints:**

- Binary Efficacy
- Binary Toxicity
- Efficacy & Toxicity
- Co-Primary Efficacy
- Ordinal Efficacy

**Interims:** [Help](#)

Sample sizes when interim analyses to be performed, separated by space. The last number must be the total sample size.

10 20 35 50

**Null Hypothesis:** [Help](#)

**Response Rate**

0.2

**Alternative Hypothesis:** [Help](#)

**Response Rate**

0.4

**Type I error rate:**

0.1

**Calculate stopping boundaries**

**Stopping boundaries**

Optimal stopping boundaries that maximize power

CSV Excel PDF Print Search:

# patients treated	Stop if # responses <=
10	1
20	3
35	7
50	13

Showing 1 to 4 of 4 entries Previous **1** Next

The power of this trial is: 0.9231

# BOP2 App: Binary Endpoint

The screenshot displays the BOP2 App interface. At the top, the authors are listed as Heng Zhou, Ying Yuan, and Jack J. Lee, from the Department of Biostatistics, MD Anderson Cancer Center. The interface has four tabs: Trial Setting, Simulation (selected), Protocol, and Reference.

**Simulation scenarios:**

- Add a Scenario** (button) **Remove a Scenario** (button)
- Scenario 1**  
**Response Rate**  
0.2
- Scenario 2**  
**Response Rate**  
0.3
- Scenario 3**  
**Response Rate**  
0.4
- Scenario 4**  
**Response Rate**  
0.5

**Number of scenarios (read only):**  
4

**Number of simulations**  
10000

**Simulate** (button)

**Operating characteristics**

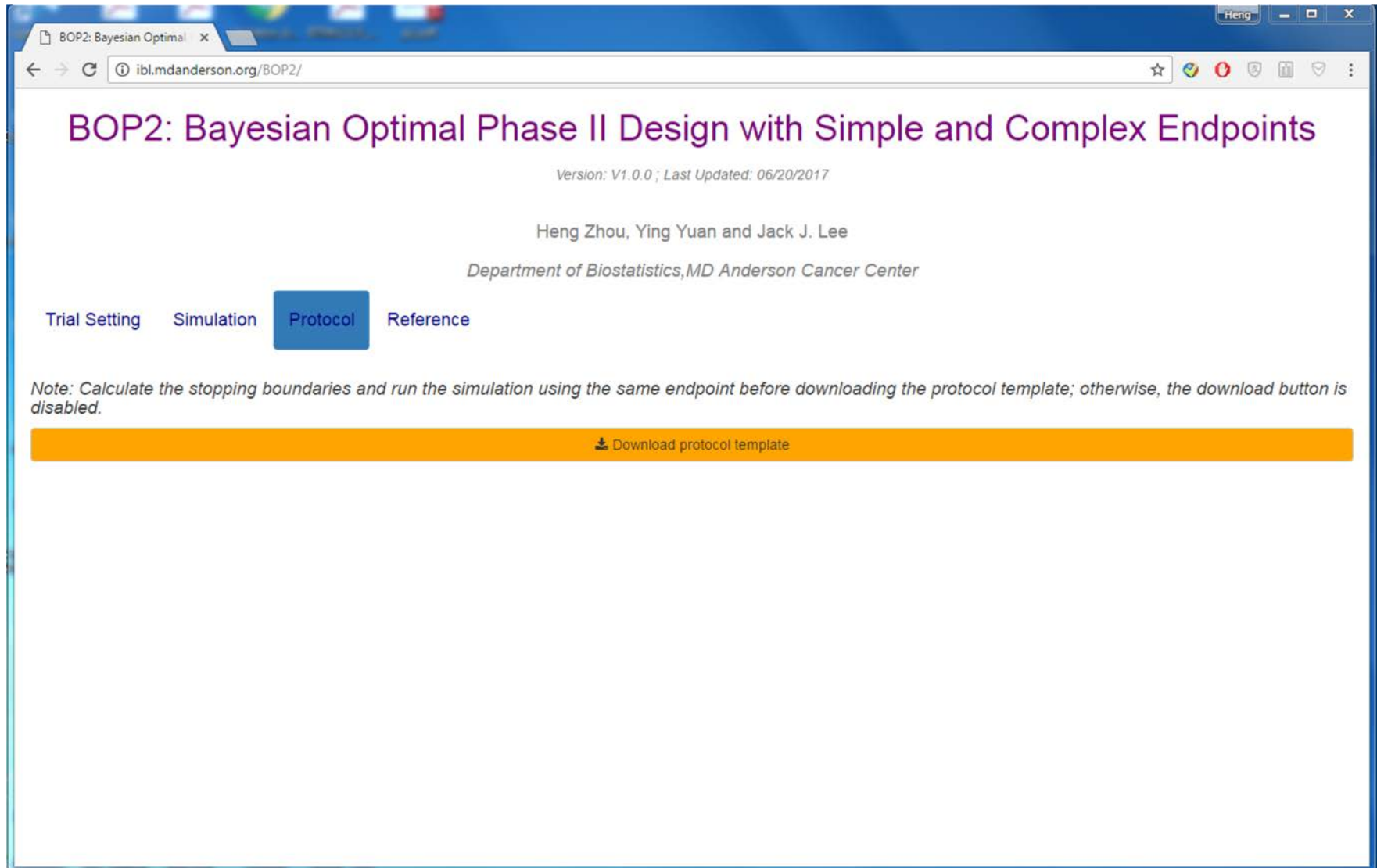
Operating characteristics

CSV Excel PDF Print Search:

Response rate	Early stopping (%)	Claim promising (%)	Sample size
0.2	69.12	8.92	28.2
0.3	24.88	59.36	41.8
0.4	5.84	92.31	47.8
0.5	1.24	98.73	49.5

Showing 1 to 4 of 4 entries Previous **1** Next

# BOP2 App: Binary Endpoint



The screenshot shows a web browser window with the address bar displaying `ibl.mdanderson.org/BOP2/`. The page title is "BOP2: Bayesian Optimal Phase II Design with Simple and Complex Endpoints". Below the title, it indicates "Version: V1.0.0 ; Last Updated: 06/20/2017" and lists the authors "Heng Zhou, Ying Yuan and Jack J. Lee" from the "Department of Biostatistics, MD Anderson Cancer Center". A navigation menu includes "Trial Setting", "Simulation", "Protocol" (which is highlighted in blue), and "Reference". A note states: "Note: Calculate the stopping boundaries and run the simulation using the same endpoint before downloading the protocol template; otherwise, the download button is disabled." At the bottom, there is a prominent orange button labeled "Download protocol template".

BOP2: Bayesian Optimal Phase II Design with Simple and Complex Endpoints

Version: V1.0.0 ; Last Updated: 06/20/2017

Heng Zhou, Ying Yuan and Jack J. Lee  
Department of Biostatistics, MD Anderson Cancer Center

Trial Setting   Simulation   **Protocol**   Reference

Note: Calculate the stopping boundaries and run the simulation using the same endpoint before downloading the protocol template; otherwise, the download button is disabled.

Download protocol template

# BOP2 App: Binary Endpoint

The screenshot shows a Microsoft Word document with the following content:

## Efficacy Monitoring

We monitor the efficacy endpoint using the Bayesian optimal phase 2 (BOP2) design (Zhou, Lee and Yuan, 2017). Specifically, let  $n$  denote the interim sample size and  $N$  denote the maximum sample size. Let  $p_{eff}$  denote the probability of efficacy (response rate) and define the null hypothesis  $H_0: p_{eff} \leq 0.2$ , representing that the treatment is inefficacious. We will stop enrolling patients and claim that the treatment is not promising if

$$Pr(p_{eff} > 0.2 | data) < \lambda \left(\frac{n}{N}\right)^\alpha,$$

where  $\lambda$  and  $\alpha$  are design parameters optimized to minimize the chance of incorrectly claiming that an efficacious treatment is not promising (i.e., type II error) under the alternative hypothesis  $H_1: p_{eff} = 0.4$ , while controlling the type I error rate at 0.1 (i.e., the chance of incorrectly claiming that an inefficacious treatment is promising is no more than 10%). Assuming a Beta(0.2,0.8) prior distribution for  $p_{eff}$ , the above decision rule corresponds to the following stopping boundaries and yields a statistical power of 0.9231 under  $H_1$ :

# patients treated	Stop if # responses <=
10	1
20	3
35	7
50	13

*Table 1: Optimized stopping boundaries*

Based on Table 1, we perform the interim analysis when the number of enrolled patients reaches 10, 20, 35. When the total number of patients reaches the maximum sample size of 50, we reject the null hypothesis and conclude that the treatment is promising if the number of responses are greater than 13; otherwise we conclude that the treatment is not promising.

Below are the operating characteristics for monitoring the response rate based on 10000 simulations using the BOP2 web application, which is available at <http://www.trialdesign.org>.

Response rate	Early stopping (%)	Claim promising (%)	Sample size
---------------	--------------------	---------------------	-------------

The document footer shows: PAGE 1 OF 2, 318 WORDS, ENGLISH (UNITED STATES), and a zoom level of 100%.



# BOP2 App: Co-primary Endpoints

The screenshot displays the BOP2 App interface for configuring a trial with co-primary endpoints. The browser address bar shows the URL `ibl.mdanderson.org/BOP2/`. The main configuration area is light blue and contains the following sections:

- Endpoints:** A list of radio buttons with **Co-Primary Efficacy** selected. Other options include Binary Efficacy, Binary Toxicity, Efficacy & Toxicity, and Ordinal Efficacy.
- Interims:** A text input field containing `10 20 35 50`. A **Help** button is present.
- Null Hypothesis:** Three input fields for  $\text{Pr}(\text{Eff1})$  (0.1),  $\text{Pr}(\text{Eff2})$  (0.2), and  $\text{Pr}(\text{Eff1 \& Eff2})$  (0.05). A **Help** button is present.
- Alternative Hypothesis:** Three input fields for  $\text{Pr}(\text{Eff1})$  (0.3),  $\text{Pr}(\text{Eff2})$  (0.35), and  $\text{Pr}(\text{Eff1 \& Eff2})$  (0.15). A **Help** button is present.
- Type I error rate:** An input field containing `0.1`.

A large blue button at the bottom left is labeled **Calculate stopping boundaries**. On the right side of the page, the text **Stopping boundaries** is displayed in orange. A small dialog box at the bottom right indicates **Calculation in progress** with a progress bar and the text **This may take a while...**

# BOP2 App: Co-primary Endpoints

The screenshot shows the BOP2 App interface with the following sections:

- Endpoints:** Radio buttons for Binary Efficacy, Binary Toxicity, Efficacy & Toxicity, Co-Primary Efficacy (selected), and Ordinal Efficacy.
- Interims:** A text input field containing "10 20 35 50" and a "Help" button.
- Null Hypothesis:** Input fields for Pr(Eff1) (0.1), Pr(Eff2) (0.2), and Pr(Eff1 & Eff2) (0.05), with a "Help" button.
- Alternative Hypothesis:** Input fields for Pr(Eff1) (0.3), Pr(Eff2) (0.35), and Pr(Eff1 & Eff2) (0.15), with a "Help" button.
- Type I error rate:** An input field containing "0.1".
- Calculate stopping boundaries:** A blue button at the bottom of the input section.
- Stopping boundaries:** A section titled "Optimal stopping boundaries that maximize power" containing a table and a "Search" input field.

The table displays the following data:

# patients treated	Stop if # Eff1 <=	AND # Eff2 <=
10	0	1
20	2	4
35	4	8
50	8	14

Below the table, it shows "Showing 1 to 4 of 4 entries" and navigation buttons for "Previous", "1", and "Next".

The power of this trial is: 0.983

# BOP2 App: Co-primary Endpoints

The screenshot displays the BOP2 App interface. At the top, it shows the browser address bar with the URL `ibl.mdanderson.org/BOP2/` and the version information: `Version: V1.0.0 ; Last Updated: 06/20/2017`. Below this, the authors' names are listed: `Heng Zhou, Ying Yuan and Jack J. Lee`, along with their affiliation: `Department of Biostatistics, MD Anderson Cancer Center`.

The interface has four main tabs: `Trial Setting`, `Simulation` (which is currently selected), `Protocol`, and `Reference`.

The `Simulation` tab contains two main sections:

- Simulation scenarios:** This section allows users to manage scenarios. It includes buttons for `Add a Scenario` and `Remove a Scenario`. There are two scenarios defined:
  - Scenario 1:** `Pr(Eff1)` is 0.1, `Pr(Eff2)` is 0.2, and `Pr(Eff1 & Eff2)` is 0.05.
  - Scenario 2:** `Pr(Eff1)` is 0.2, `Pr(Eff2)` is 0.3, and `Pr(Eff1 & Eff2)` is 0.1.
- Simulation parameters:** This section includes a `Number of scenarios (read only):` field set to 2, and a `Number of simulations` field set to 10000. A large blue `Simulate` button is located at the bottom of this section.

The **Operating characteristics** section is also visible. It features a search bar and buttons for `CSV`, `Excel`, `PDF`, and `Print`. Below these is a table with the following data:

<code>Pr(Eff1)</code>	<code>Pr(Eff2)</code>	<code>Pr(Eff1 &amp; Eff2)</code>	<code>Early stopping (%)</code>	<code>Claim promising (%)</code>	<code>Sample size</code>
0.1	0.2	0.05	66.38	9.51	31
0.2	0.3	0.1	10.92	78.68	46.8

At the bottom of the table, it indicates `Showing 1 to 2 of 2 entries` and includes navigation buttons for `Previous`, `1` (the current page), and `Next`.

# BOP2 App: Co-primary Endpoints

The screenshot displays a web browser window with the URL `ibl.mdanderson.org/BOP2/`. The page title is "BOP2: Bayesian Optimal Phase II Design with Simple and Complex Endpoints". Below the title, it indicates "Version: V1.0.0 ; Last Updated: 06/20/2017" and lists the authors: "Heng Zhou, Ying Yuan and Jack J. Lee" from the "Department of Biostatistics, MD Anderson Cancer Center". A navigation menu includes "Trial Setting", "Simulation", "Protocol" (which is the active tab), and "Reference". A note states: "Note: Calculate the stopping boundaries and run the simulation using the same endpoint before downloading the protocol template; otherwise, the download button is disabled." Below this note is a prominent orange button labeled "Download protocol template". At the bottom of the browser window, a file named "CoeffMonitoring.docx" is open, and a "Show all" button is visible in the bottom right corner.

BOP2: Bayesian Optimal Phase II Design with Simple and Complex Endpoints

Version: V1.0.0 ; Last Updated: 06/20/2017

Heng Zhou, Ying Yuan and Jack J. Lee  
Department of Biostatistics, MD Anderson Cancer Center

Trial Setting   Simulation   **Protocol**   Reference

*Note: Calculate the stopping boundaries and run the simulation using the same endpoint before downloading the protocol template; otherwise, the download button is disabled.*

Download protocol template

CoeffMonitoring.docx   Show all

# BOP2 App: Co-primary Endpoints

**Co-primary Efficacy Monitoring**

We simultaneously monitor two co-primary efficacy endpoints using the Bayesian optimal phase 2 (BOP2) design (Zhou, Lee and Yuan, 2017). Specifically, let  $n$  denote the interim sample size and  $N$  denote the maximum sample size. Let  $Y_1$  and  $Y_2$  denote the two co-primary endpoints, with  $Y_1 = 1$  and  $Y_2 = 1$  indicating that patients experienced favorable treatment responses in the two respective endpoints. We assume that the joint distribution of  $(Y_1, Y_2)$  follows a multinomial distribution with 4 elementary outcomes:  $(Y_1, Y_2) = (1, 1), (0, 1), (1, 0)$  and  $(0, 0)$ . Let  $p_1 = Pr(Y_1 = 1)$ ,  $p_2 = Pr(Y_2 = 1)$ , and define the null hypothesis  $H_0: p_1 \leq 0.1$  and  $p_2 \leq 0.2$ , representing that the treatment is inefficacious in both co-primary endpoints. We will stop enrolling patients and claim the treatment is not promising if

$$Pr(p_1 > 0.1 \mid data) < \lambda \left(\frac{n}{N}\right)^\alpha,$$

or

$$Pr(p_2 > 0.2 \mid data) < \lambda \left(\frac{n}{N}\right)^\alpha,$$

where  $\lambda$  and  $\alpha$  are design parameters optimized to minimize the chance of incorrectly claiming that an efficacious treatment is not promising (i.e., type II error) under the alternative hypothesis  $H_1: p_1 = 0.3$  and  $p_2 = 0.35$ , while controlling the type I error rate at 0.1 (i.e., the chance of incorrectly claiming that an inefficacious treatment is promising is no more than 10%). Assuming a Dirichlet prior distribution  $Dir(0.05, 0.05, 0.15, 0.75)$  for the treatment effect, the above decision rule corresponds to the following stopping boundaries and yields a statistical power of 0.983 under  $H_1$ :

# patients treated	Stop if # Eff1 <=	AND # Eff2 <=
10	0	1
20	2	4
35	4	8
50	8	14

*Table 1: Optimized stopping boundaries*

Based on Table 1, we perform the interim analysis when the number of enrolled patients reaches 10, 20, 35. When the total number of patients reaches the maximum sample size of

PAGE 1 OF 2   402 WORDS   ENGLISH (UNITED STATES)   100%

# Summary

---

- BOP2 design provides a unified framework for phase II clinical trials with simple and complex endpoints.
- Compared to existing posterior probability based Bayesian phase II design, BOP2 yields higher power to detect the efficacious treatment with well controlled type I error.
- Stopping boundaries of the BOP2 design can be tabulated before the onset of the trial, making the implementation of the design extremely simple.
- Easy-to-use software is freely available to generate stopping boundaries, operating characteristics and protocol for the BOP2 design.

# Reference

---

- Zhou, H., Lee, JJ. and Yuan, Y. (2017) BOP2: Bayesian Optimal Design for Phase II Clinical Trials with Simple and Complex Endpoints. *Statistics in Medicine* , 36, 3302-3314.

