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.

ClinicalTrials.gov Identifier: NCT02949219

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

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

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 $\boldsymbol{\theta} = (\theta_1, \ldots, \theta_K)^T$ denote the probability that *Y* belongs to each category.

Model

- Dirichlet-multinomial model

$$Y|\boldsymbol{\theta} \sim Multinom(\theta_1, \ldots, \theta_K)$$
$$(\theta_1, \ldots, \theta_K) \sim Dir(a_1, \ldots, a_K)$$

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

- Let $D_n = (x_1, \ldots, x_K)$ 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

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

  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 $n_1, \ldots, n_R$, 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 $\boldsymbol{\theta}$ in the form of

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

where $\boldsymbol{b}$ is a design vector with elements of 0 and 1, ϕ 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 $(\theta_1, \theta_2, \theta_3, \theta_4) = (\Pr(\mathrm{CR}), \Pr(\mathrm{PR}), \Pr(\mathrm{SD}), \Pr(\mathrm{DP}))$.

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 λ and γ 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 $\boldsymbol{\theta} \sim Dir(a_1 + x_1, \ldots, a_K + x_K)$ and a design vector $\boldsymbol{b} = (b_1, \ldots, b_K)$ with elements of 0 and 1, $\boldsymbol{b\theta}$ follows a Beta distribution $Beta(\sum_{k=1}^{K} b_k(a_k + x_k), \sum_{k=1}^{K}(1 - b_k)(a_k + x_k))$.

  $\longrightarrow$ $\Pr(\boldsymbol{b\theta} \leq \phi | D_n)$ can be easily evaluated.

- **Lemma 1**. $\Pr(\boldsymbol{b\theta} \leq \phi | D_n)$ is a monotonic function of $\sum_{k=1}^{K} b_k x_k$.

  $\longrightarrow$ 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 $\leq$	0	1	3	4	5	7	9
		# of CR/PR $\leq$	2	3	5	8	10	13	16
Example 3	and	# of OR $\leq$	0	1	2	3	4	5	7
		# of PFS6 $\leq$	1	2	4	5	7	9	12
Example 4	or	# of OR $\leq$	2	5	7	10	13	16	19
		# of Toxicities $\geq$	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 $\leq$	0	1	3	4	5	7	9
		# of CR/PR $\leq$	2	3	5	8	10	13	16

OR: objective response

Stopping Boundaries of BOP2 Design

			Number of patients treated						
Trial	Stop the trial if		10	15	20	25	30	35	40
Example 3	and	# of OR $\leq$	0	1	2	3	4	5	7
		# of PFS6 $\leq$	1	2	4	5	7	9	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	Claim promising(%)		Early termination(%)		Sample size	
(ORR)	BOP2	TS	BOP2	TS	BOP2	TS
0.20§	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†	88.3	76.4	11.4	23.5	37.6	33.6
0.50	98.2	93.3	1.8	6.7	39.5	38.1

§: null hypothesis; †: 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 $n = 10, 15, 20, 25, 30, 35, 40$.

	Claim promising(%)		Early termination(%)		Sample size	
(CR, CR/PR)	BOP2	TS	BOP2	TS	BOP2	TS
(0.15, 0.30)[§]	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)[†]	85.5	74.2	9.9	25.7	38.5	33.0
(0.30, 0.55)	95.7	85.2	3.0	14.8	39.5	35.9

[§]: null hypothesis; [†]: alternative hypothesis.

Simulation Results: Co-primary Endpoints

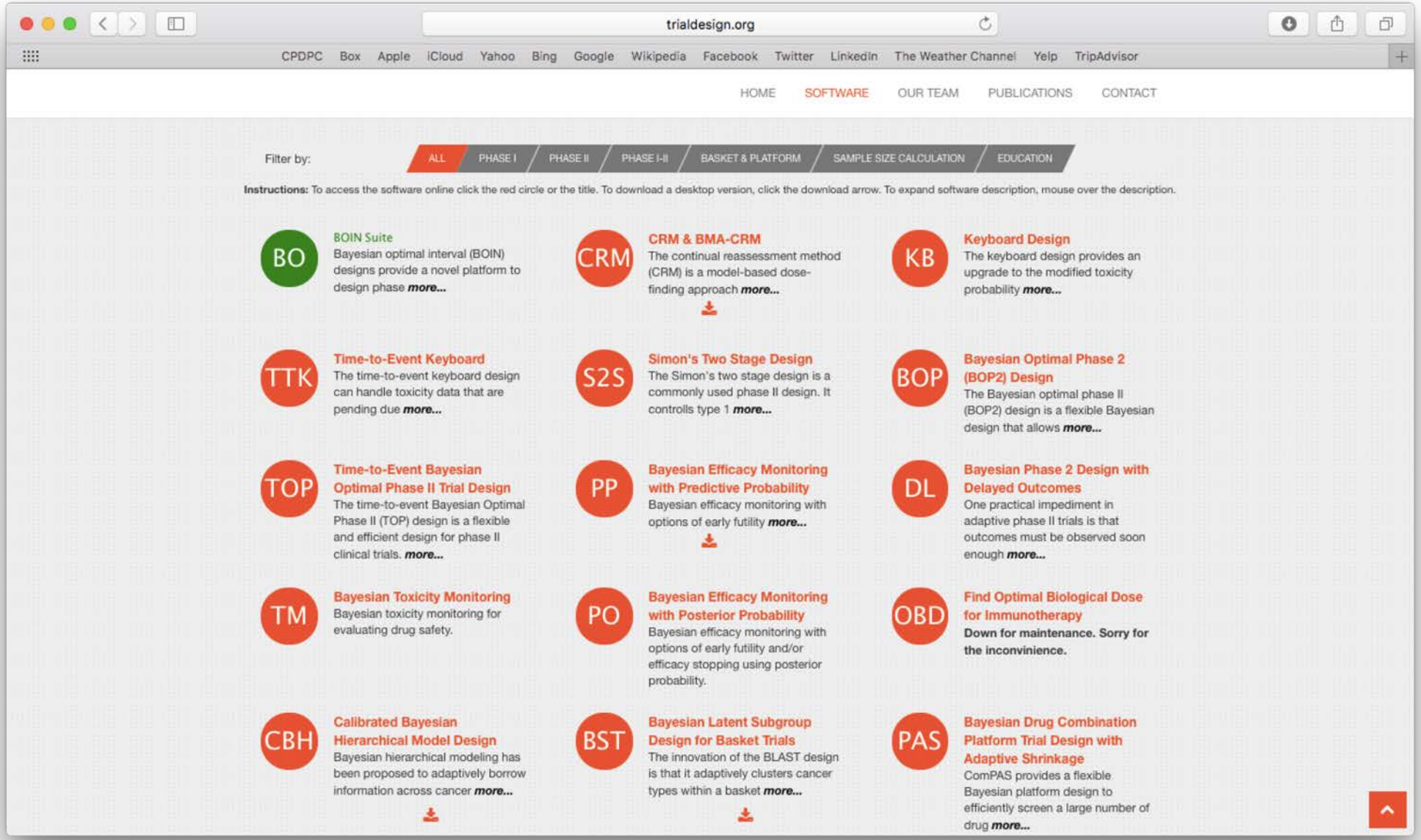
Table: Operating characteristics under the BOP2 design and TSE design (Thall, Simon and Estey, 1995) with two co-primary efficacy endpoints. The interims occur when $n = 10, 15, 20, 25, 30, 35, 40$.

	Claim promising(%)		Early termination(%)		Sample size	
(ORR, PFS 6m)	BOP2	TSE	BOP2	TSE	BOP2	TSE
(0.10, 0.20)[§]	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)[†]	96.1	75.5	2.4	24.5	39.5	32.8
(0.30, 0.40)	98.5	82.6	0.8	17.4	39.8	34.8

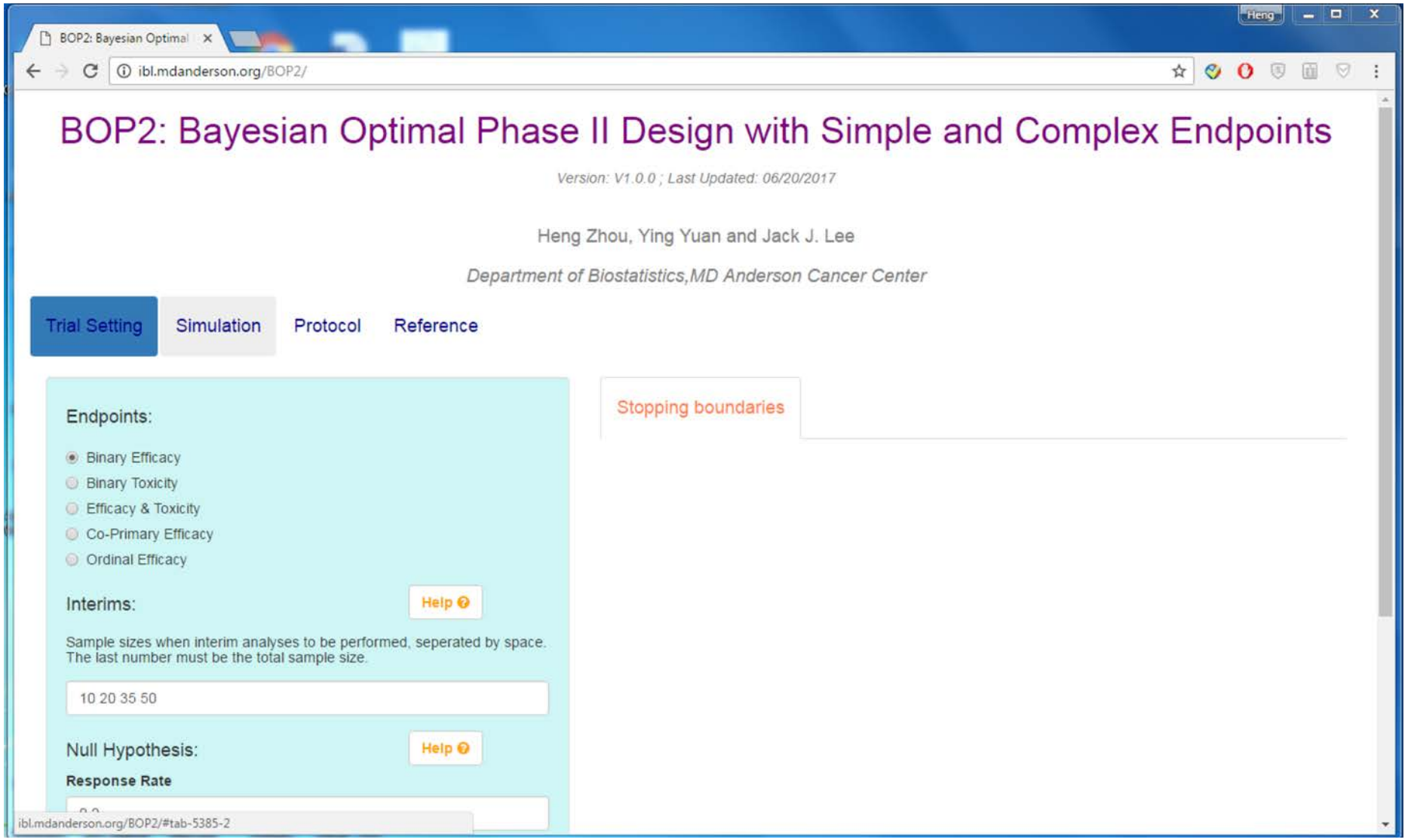
[§]: null hypothesis; [†]: alternative hypothesis.

www.trialdesign.org

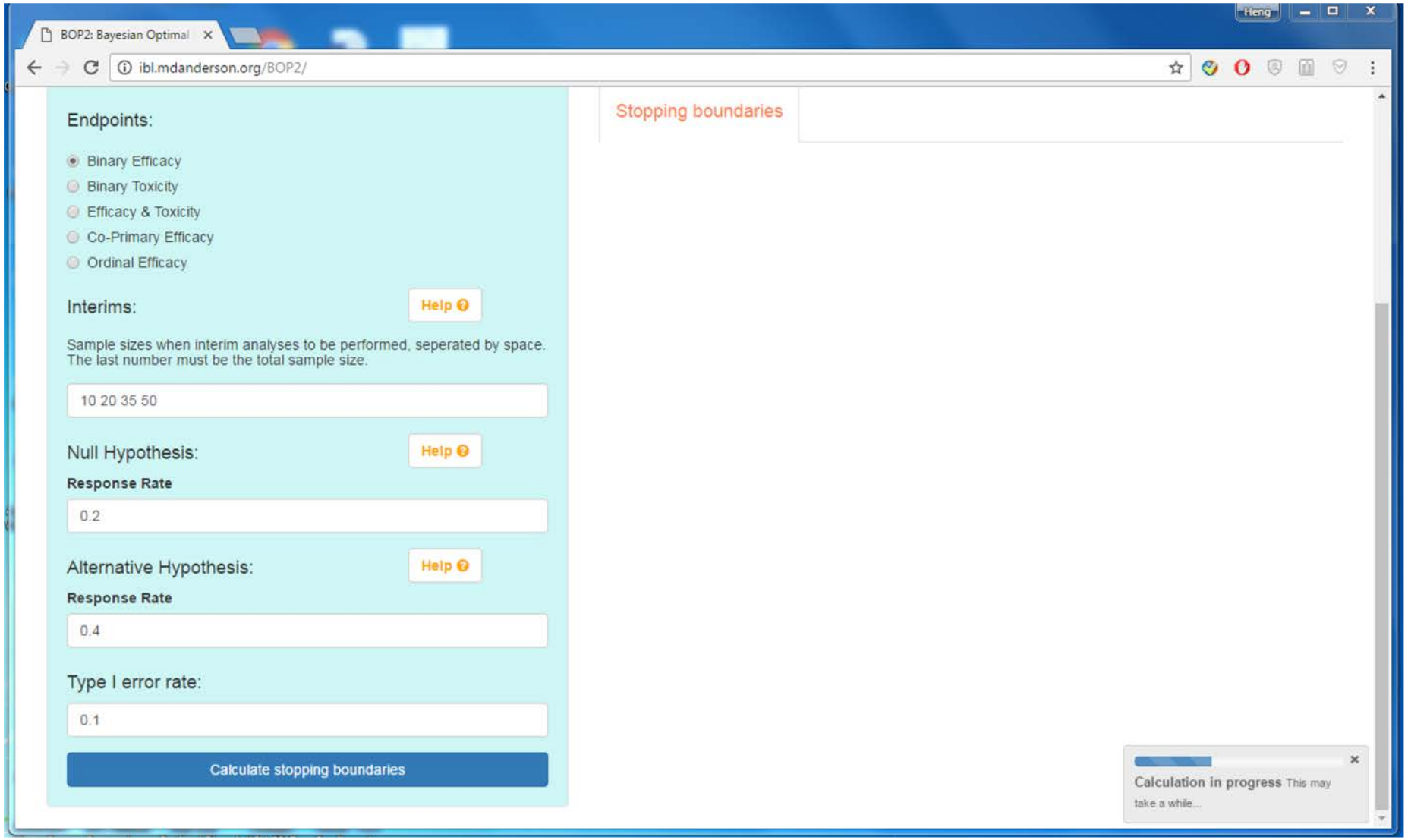
www.trialdesign.org



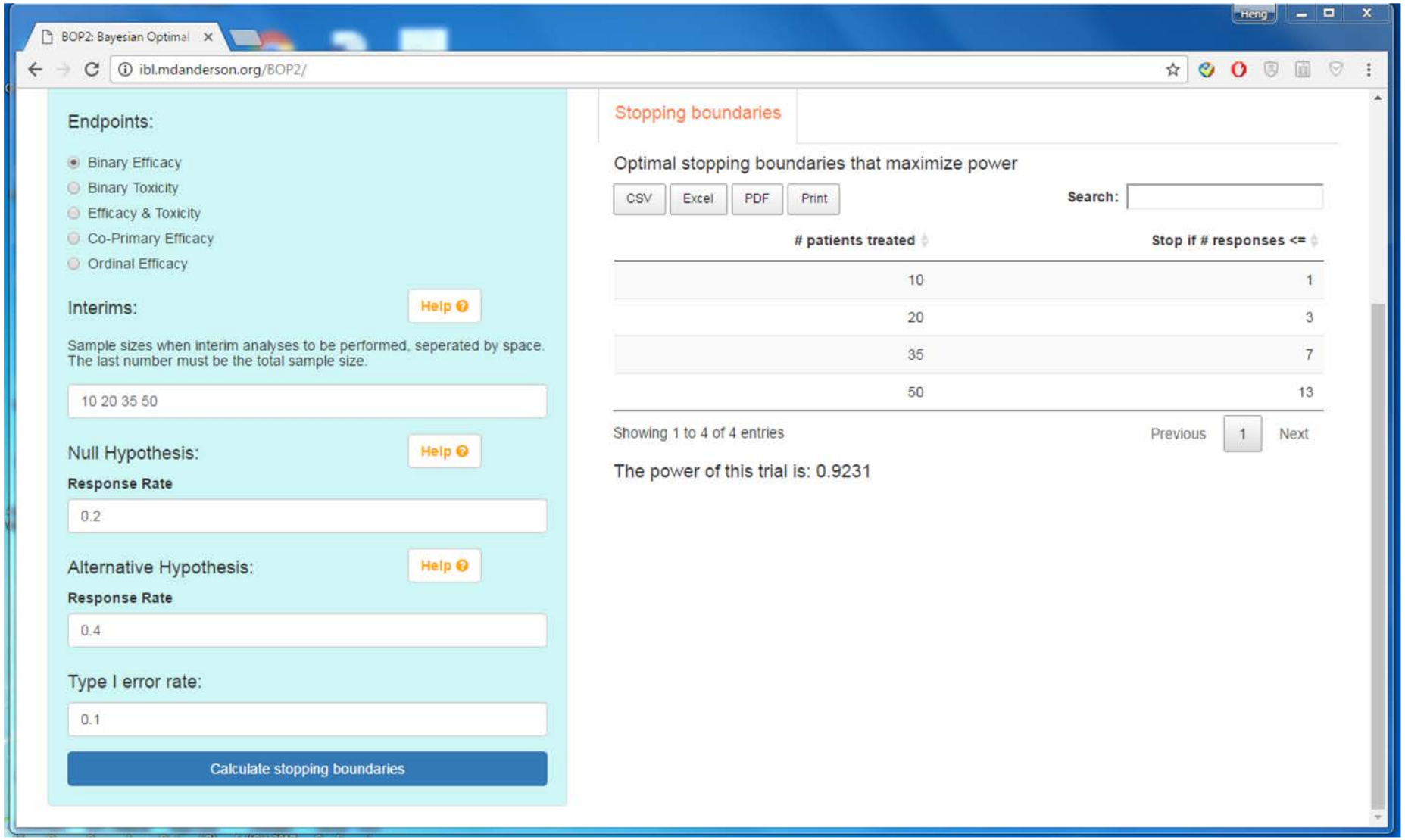
BOP2 App: Binary Endpoint



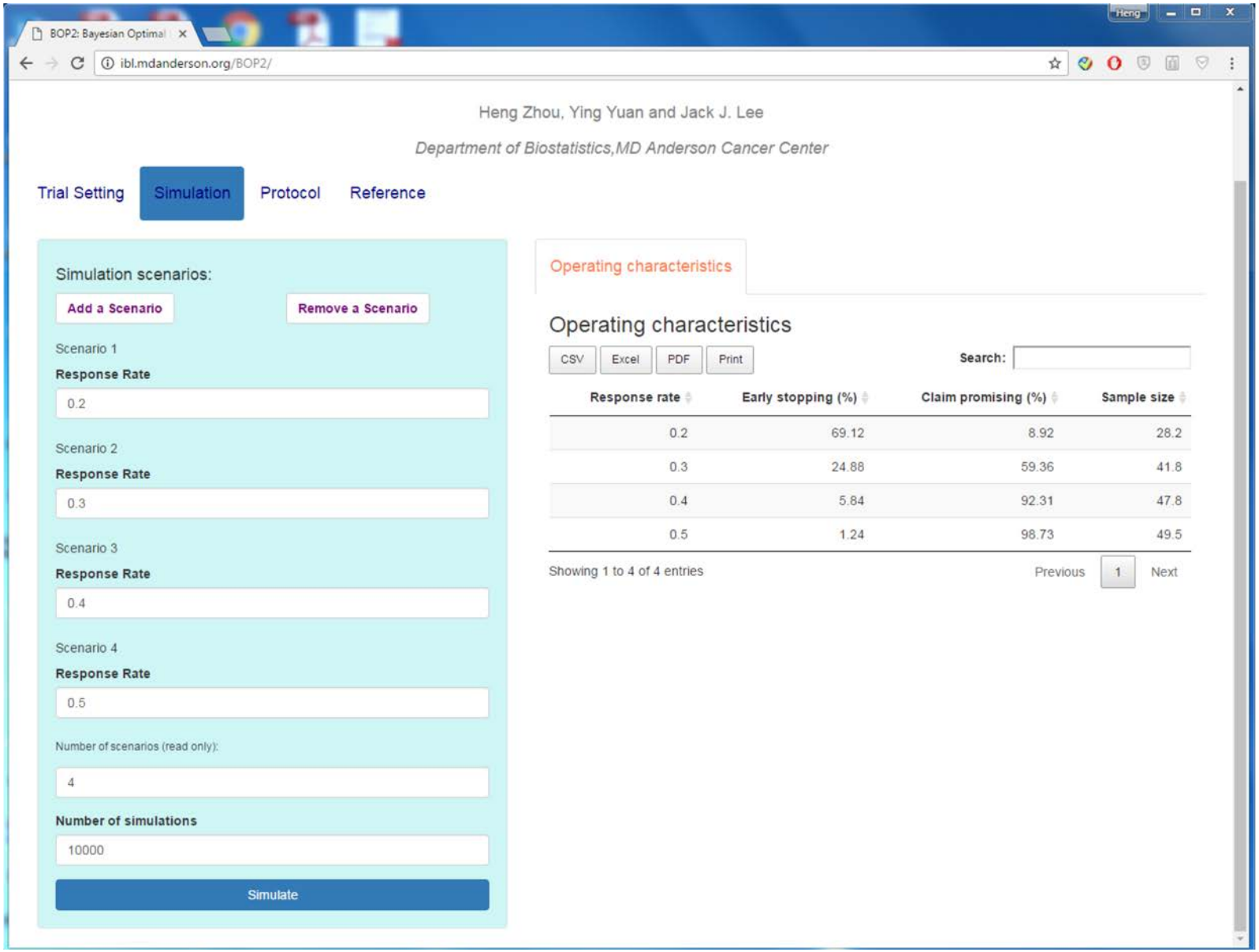
BOP2 App: Binary Endpoint



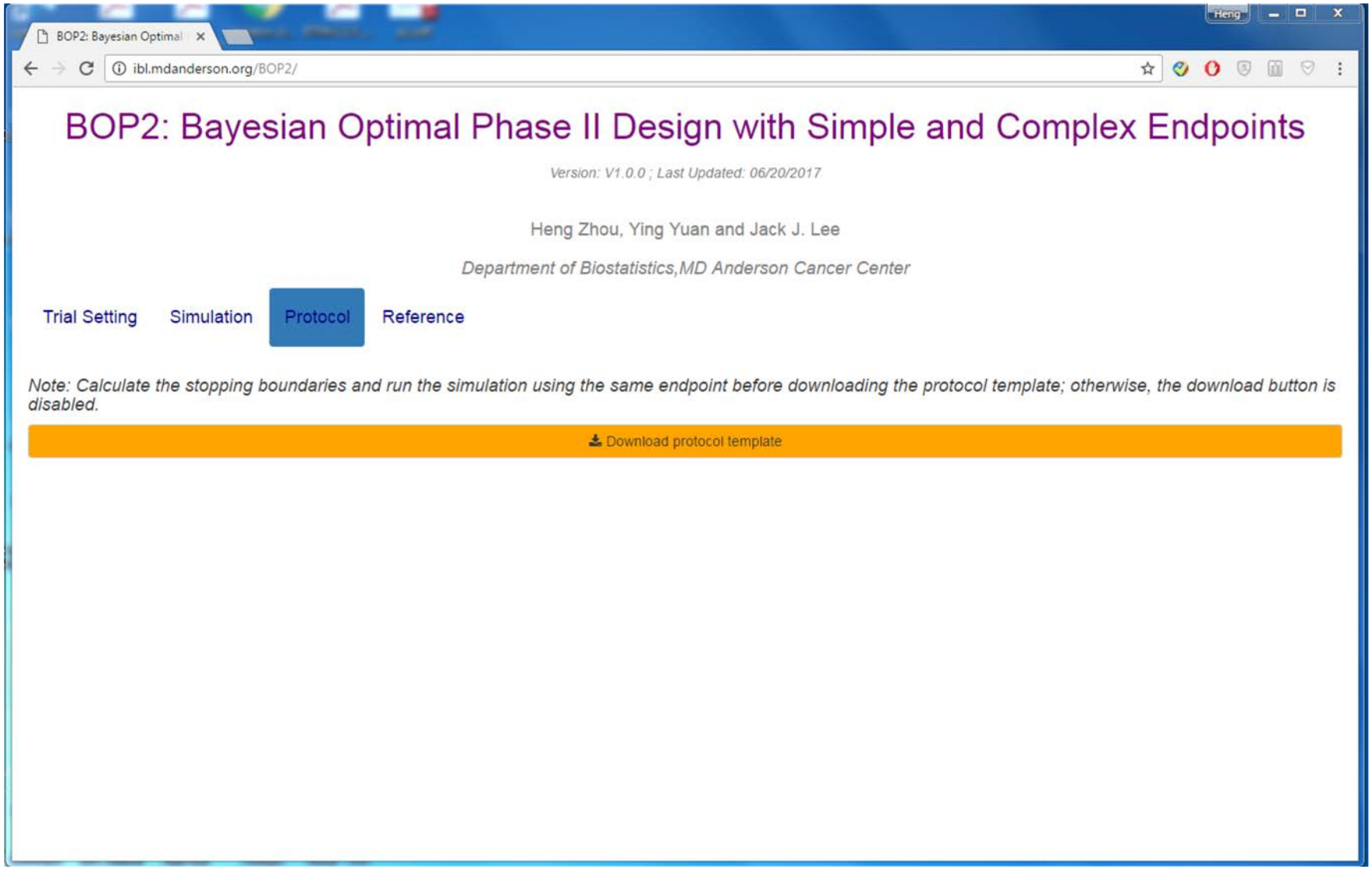
BOP2 App: Binary Endpoint



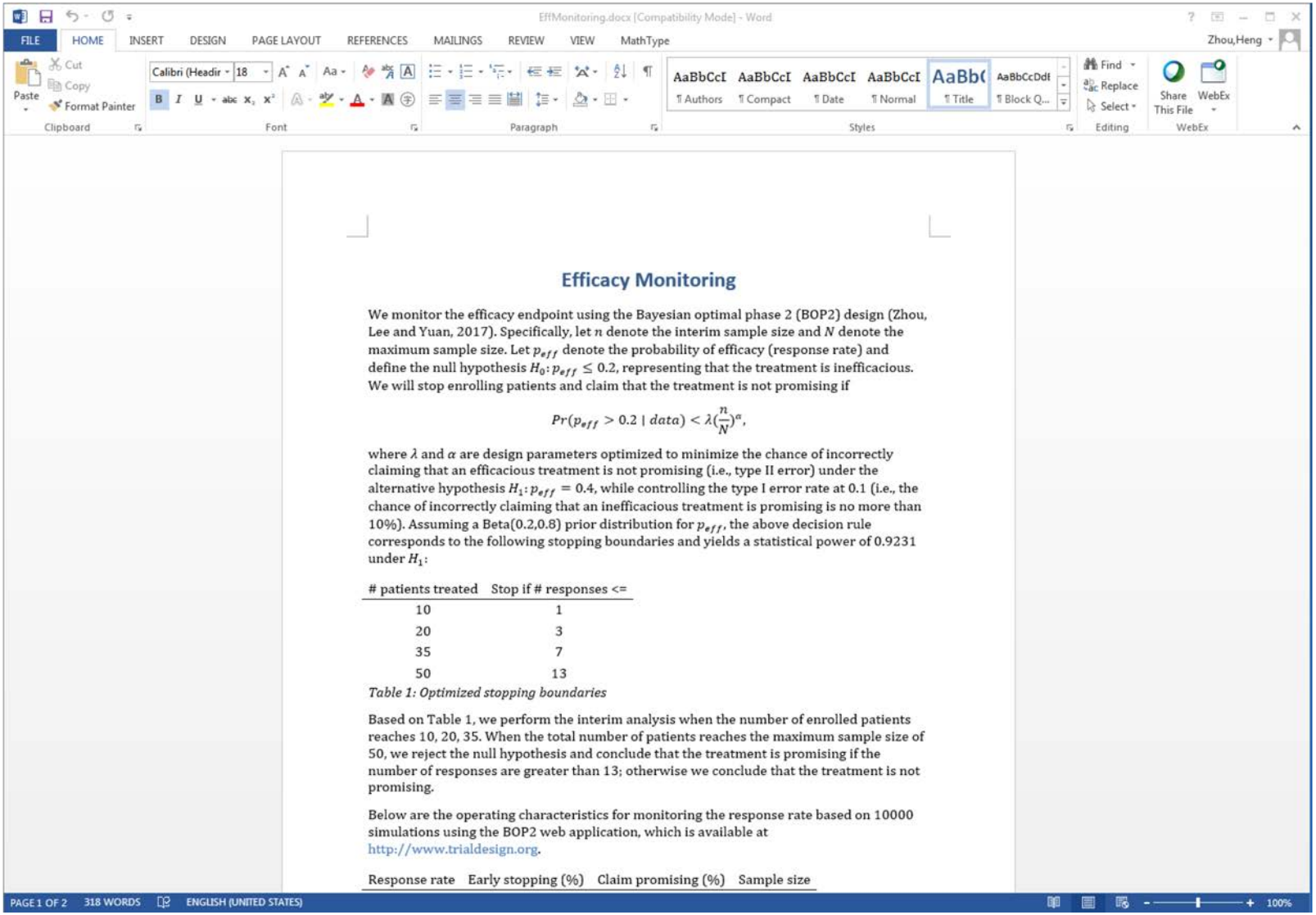
BOP2 App: Binary Endpoint



BOP2 App: Binary Endpoint



BOP2 App: Binary Endpoint



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

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

where λ and α 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

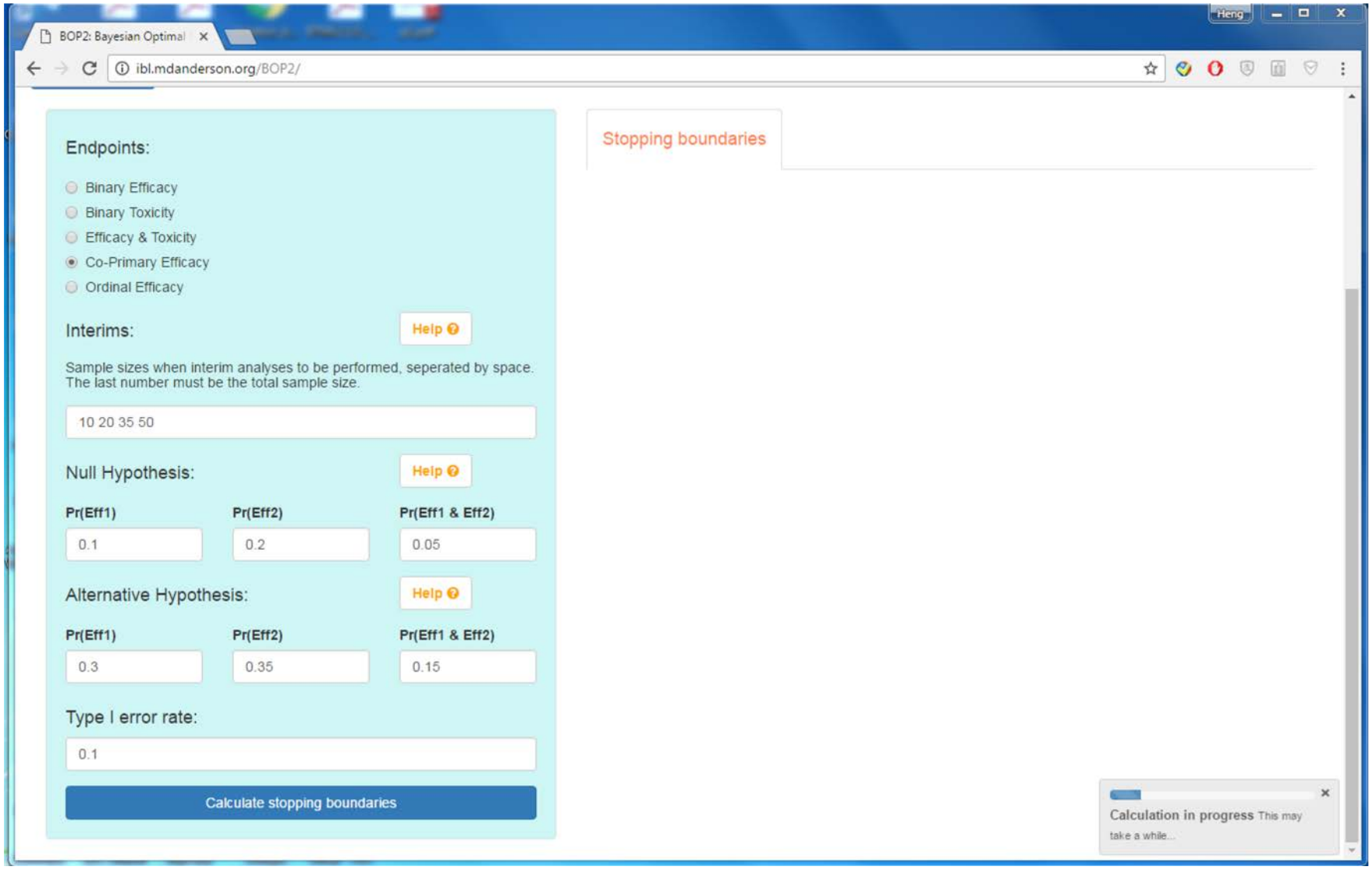
Table 1: Optimized stopping boundaries

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.

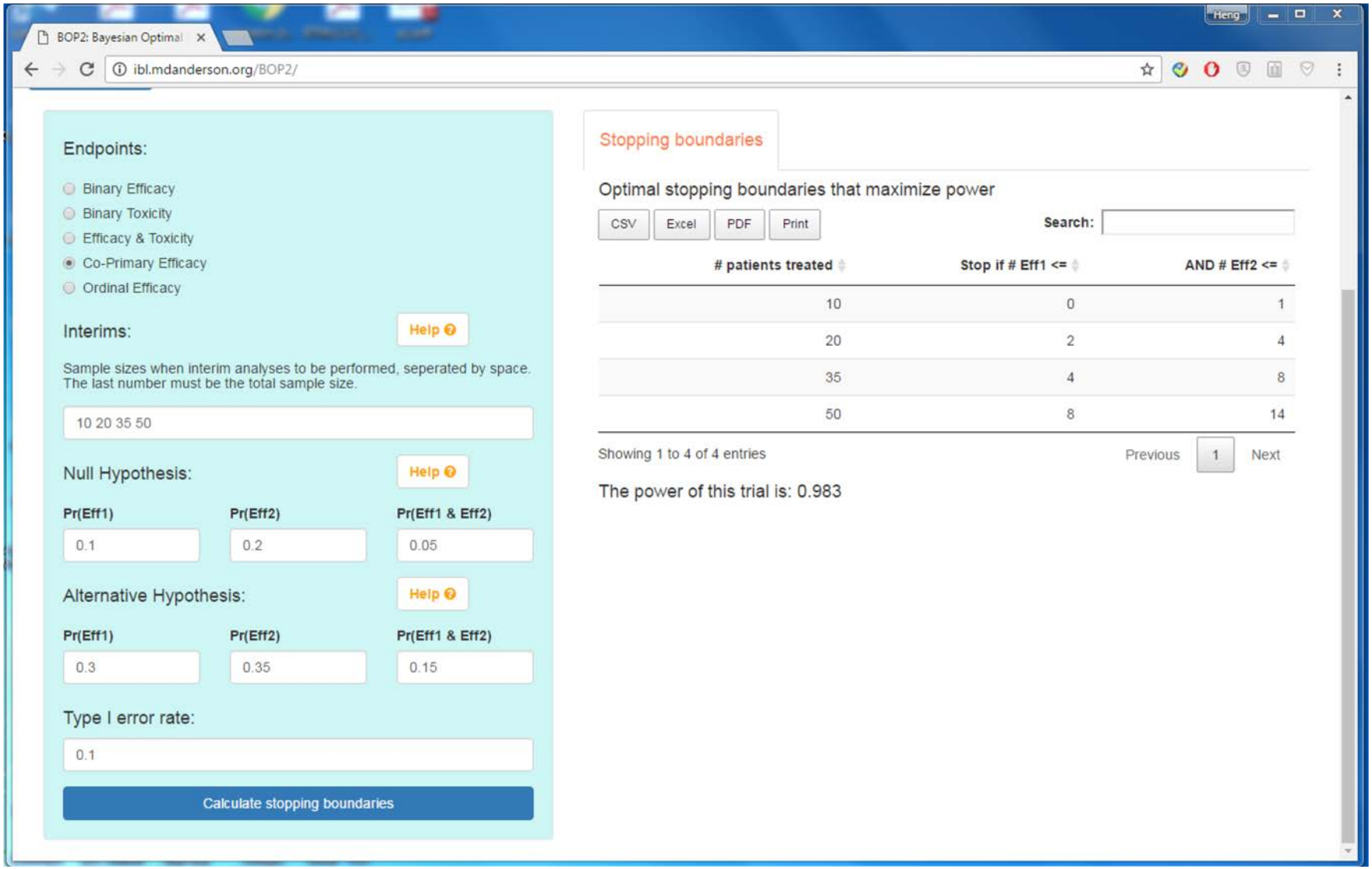
Below are the operating characteristics for monitoring the response rate based on 10000 simulations using the BOP2 web application, which is available at http://www.trialdesign.org.

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

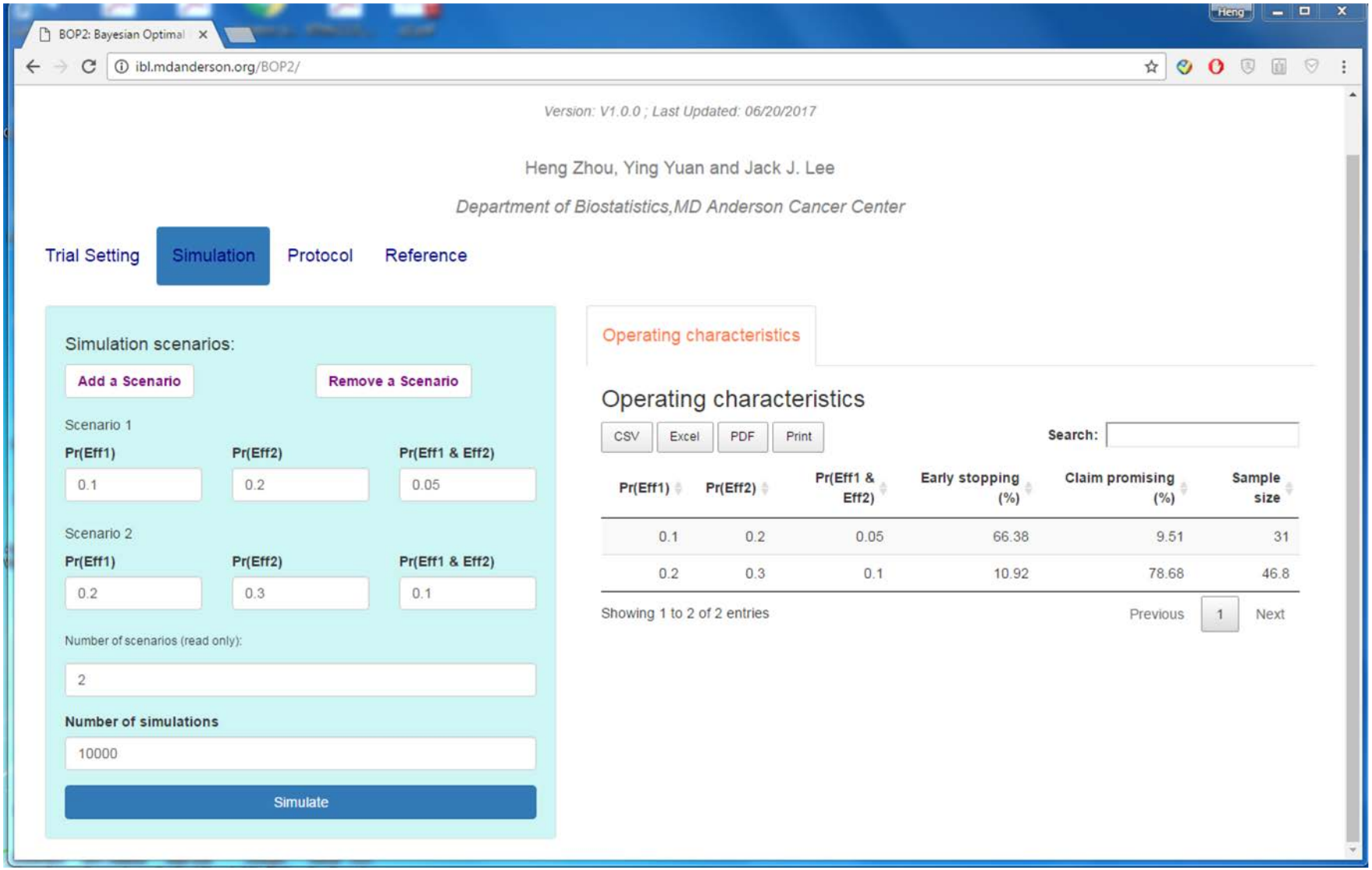
BOP2 App: Co-primary Endpoints



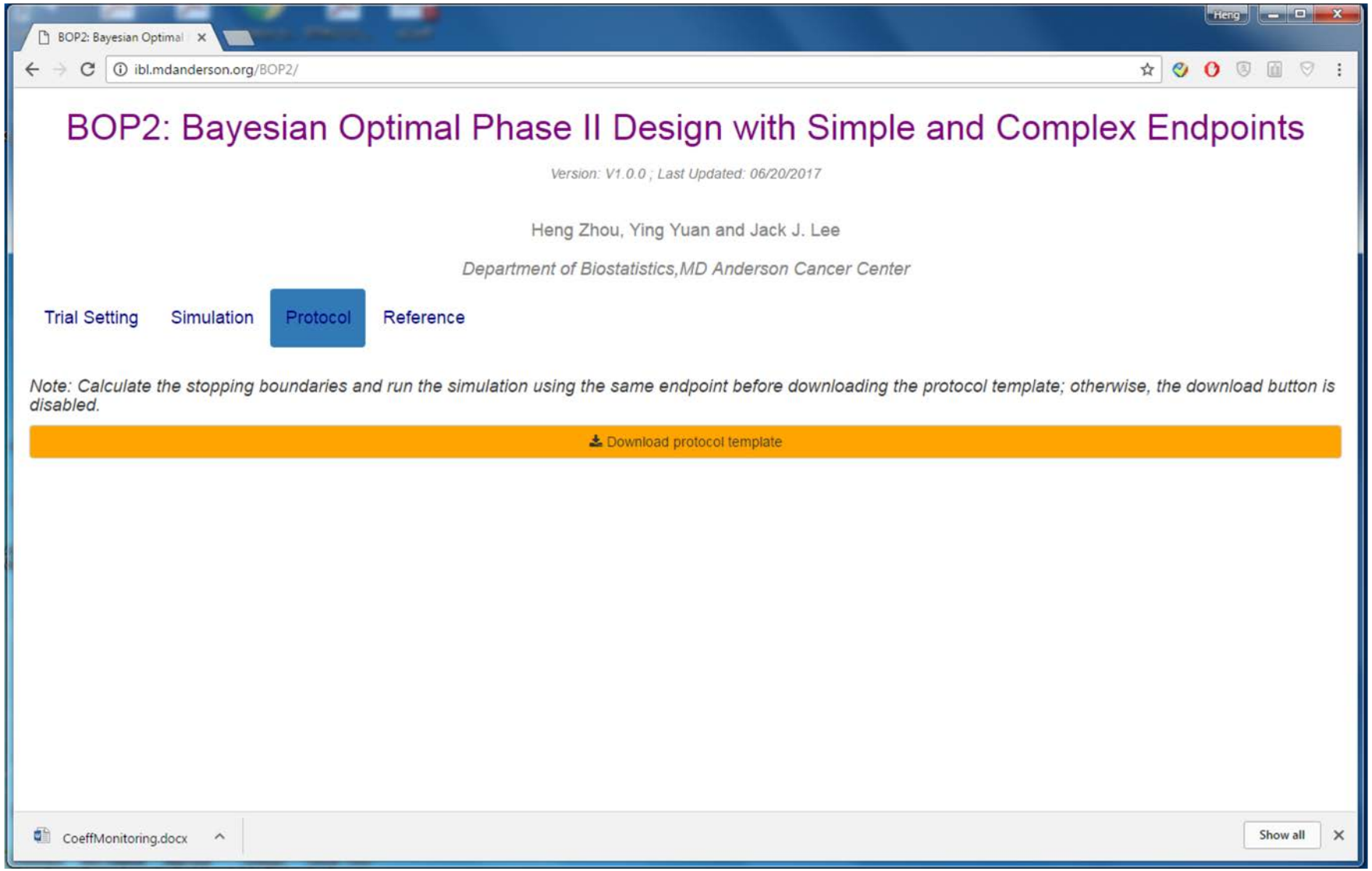
BOP2 App: Co-primary Endpoints



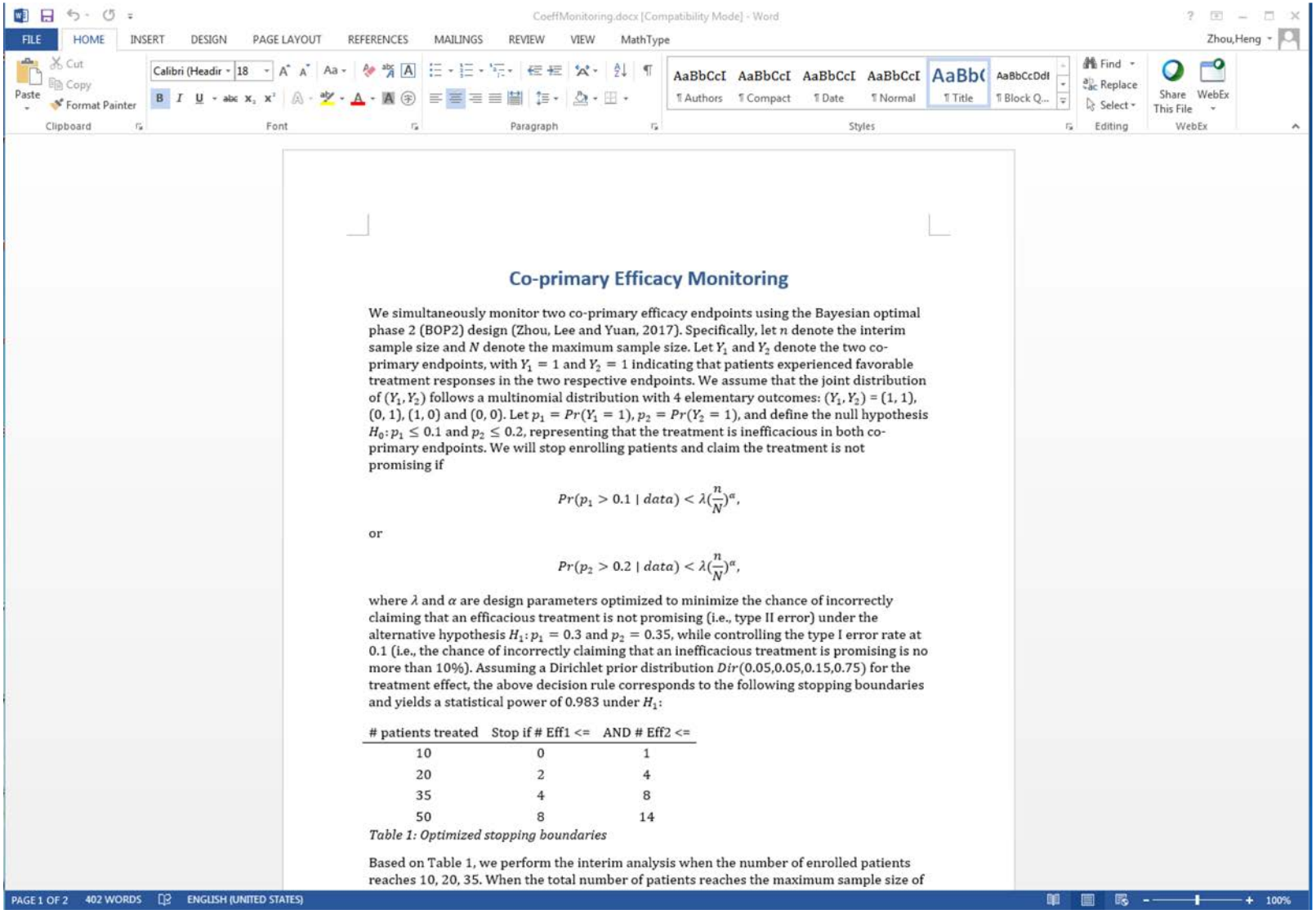
BOP2 App: Co-primary Endpoints



BOP2 App: Co-primary Endpoints



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

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

or

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

where λ and α 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

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.


Thank You!