- Predictive prob. of future evidence against H_0 (i.e., the PP of trial success):

$$\text{PP} = \text{E}[Pr(p > p_0 \mid \mathbf{X}, \mathbf{Y}) > \theta_T \mid \mathbf{X}]$$

⋆⋆**Case 1:** A **high** PP means that the treatment is likely to be efficacious by the end of the study given S_x.

$\Rightarrow$ The trial should be stopped early due to efficacy

⋆⋆**Case 2:** A **small** PP means that the treatment may not have sufficient activity.

$\Rightarrow$ The trial should be stopped early due to futility

⋆⋆**Case 3:** A ~~**not high or small**~~ PP means that the treatment may not have sufficient activity.

$\Rightarrow$ The trial should be continued because the current data are not yet conclusive.

† **Algorithm 4.1** Phase IIA basic PP design

Introduce two thresholds, θ_L (small positive number, ≥ 0) and θ_U (large positive number, ≤ 1).

- **Step 1:** If $PP < \theta_L$, stop the trial and reject the alternative hypothesis (stop the trial due to futility).

  i.e., $PP < \theta_L \Leftrightarrow$ It is unlikely the response rate will be larger than p_0 at the end of the trial given the current data.

- **Step 2:** If $PP > \theta_U$, stop the trial and reject the null hypothesis (stop the trial due to efficacy).

  i.e., $PP > \theta_U \Leftrightarrow$ The current data suggest that if the same trend continues, we will have a high probability of concluding that the treatment is efficacious at the end of study.

- **Step 3:** Otherwise, continue to the next stage until reaching $N_{\max}$

† **Example 4.2:** To assess the efficacy of a particular new treatment.

- The current standard treatment yields a response rate of approximately 20% ($\Rightarrow \ p_0 = 0.2$)
- The target response rate of the new regimen is 40% ($\Rightarrow \ p_1 = 0.4$).
- Constraint both α and $\beta \leq 0.1$

Now we will design a trial using Simon's optimal approach and PP's approach.

† **Example 4.2:** Simon's optimal design

- $n_1 = 17$, $r_1 = 3$, $N_{\text{max}} = 37$, $r = 10$
- Probability of early termination under H_0: $PET(p_0) = 0.55$
- Expected sample size under H_0: $\text{E}(N \mid p_0) = 26.02$
- The resulting $\alpha = 0.095$ and $\beta = 0.097$.

† **Example 4.2:** Sequential stopping: Use PP to define early stopping for futility (and efficacy)

Idea: Fix the prior for p, type I and II error rates ($\alpha = \beta = 0.01$). For simplicity, we let $\theta_U = 1.0$ (no stopping due to efficacy, of course we can relax this).

$$\text{PP} = \text{E}[Pr(p > p_0 \mid \mathbf{X}, \mathbf{Y}) > \theta_T \mid \mathbf{X}] < \theta_L \quad \Rightarrow \quad \text{Stop!}$$

Search for θ_L and θ_T that yield the prespecified α and β! e.g.,

- **0. Initialize:** Fix thresholds θ_T , θ_L and θ_U, prior parameters a_0, b_0 and N_{max}

  For example, $(a_0, b_0) = (0.2, 0.8)$ (equivalent to 1 patient), $p_0 = 0.20$, $p_1 = 0.40$, $\theta_L = 0.001$, $\theta_U = 1.0$ (no stopping due to efficacy), $\theta_T = 0.9$, $N_{\text{max}} = 36$.

† **Example 4.2:** Do simulation studies to evaluate α and β with given θ_T and θ_L!

- **1. Initial stage:** Enroll first 10 patients, record $x_1, \ldots, x_{10}$. Set $n = 10$.

- **2. Sequential stopping:** Evaluate

$$\text{PP} = \text{E}[Pr(p > p_0 \mid \mathbf{X}, \mathbf{Y}) > \theta_T \mid \mathbf{X}].$$

  If PP $< \theta_L$, stop for futility. Otherwise, continue with step 3.

- **3. Next patient:** Record outcome for next patient, increment $n \equiv n + 1$. Repeat with step 2.

$p_1 = 0.2 \rightarrow H_0$
$0.4 \Rightarrow H_a$

† Tab 4-2: $\alpha = \beta = 0.1$, $p_0 = 0.2$, Be(0.2, 0.8) for p

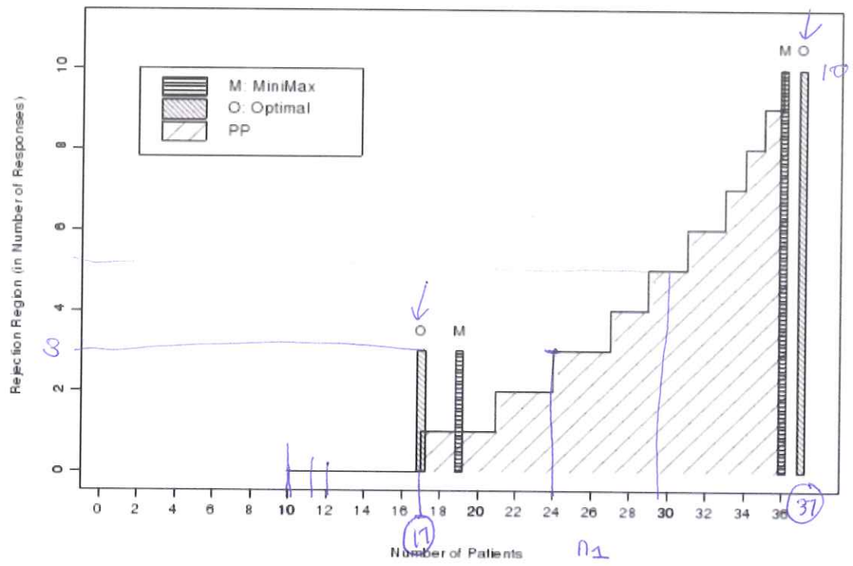
Simon's Minimax/Optimal Two-Stage designs:

	r_1/n_1	r/N_{max}	$PET(p_0)$	$E(N\|p_0)$	α	β
Minimax	3/19	10/36	0.46	28.26	0.086	0.098
→ Optimal	3/17	10/37	0.55	26.02	0.095	0.097

Predictive Probability-based designs:

θ_L	θ_T	r/N_{max}	$PET(p_0)$	$E(N\|p_0)$	α	β
		NA/35			0.126	0.093
		NA/35			0.074	0.116
0.001	[0.852,0.922]	10/36	0.86	27.67	0.088	0.094
0.011	[0.830,0.908]	10/37	0.85	25.13	0.099	0.084
0.001	[0.876,0.935]	11/39	0.88	29.24	0.073	0.092
0.001	[0.857,0.923]	11/40	0.86	30.23	0.086	0.075
0.003	[0.837,0.910]	11/41	0.85	30.27	0.100	0.062
0.043	[0.816,0.895]	11/42	0.86	23.56	0.099	0.083
0.001	[0.880,0.935]	12/43	0.88	32.13	0.072	0.074
0.001	[0.862,0.924]	12/44	0.87	33.71	0.085	0.059
0.001	[0.844,0.912]	12/45	0.85	34.69	0.098	0.048
0.032	[0.824,0.898]	12/46	0.86	26.22	0.098	0.068
0.001	[0.884,0.936]	13/47	0.89	35.25	0.071	0.058
0.001	[0.868,0.925]	13/48	0.87	36.43	0.083	0.047
0.001	[0.850,0.914]	13/49	0.86	37.86	0.095	0.038
0.020	[0.832,0.901]	13/50	0.86	30.60	0.100	0.046

† Fig 4.1: Comparison with Optimal Simon 2-Stage, the PP design allows more flexible and frequent monitoring.



† Proper Bayes Designs

- **Analysis:** Bayesian model (prior and likelihood...) for inference.
- **Design:** Define decisions (protocol) by tracking posterior probabilities of clinically meaningful events.
- **Design parameters:** Thresholds for probabilities etc. Consider frequentist properties to fix design parameters. θ_T θ_L θ_U
- **Sequential stopping:** Use PP to define early stopping for futility (and efficacy)

Jack Lee

† Stopping for Futility and Efficacy (BCLM 4.3.1)

Thall, Simon & Estey (1995, StatMed), Thall & Russell (1998, Bmcs)

IIa

- **Model:** e.g. $y_i \in \{0, 1\}$

  tumor response under new therapy (E): $Pr(y_i = 1) = \theta_E$

  standard of care (S): $\theta_S \equiv 15\%$

- **Event:** π_n: posterior probability of clinically meaningful event.

  e.g. $\pi_n = p(\theta_E > \theta_S + \delta \mid y)$ where δ: offset fixed by the investigator, reflects the minimum clinically meaningful improvement.

† Stopping for Futility and Efficacy (contd)

- **Decision boundaries:** Specify design parameters $\{(L_n, U_n), n = 1, 2, \ldots\}$ for π_n

declare E is promising

$$\text{decision} = \begin{cases} \text{stop \& E} & \text{if } \pi_n > U_n \\ \text{continue} & \text{if } L_n < \pi_n < U_n \\ \text{stop \& not E} & \text{if } \pi_n < L_n \end{cases}$$

Not promising

"Stop & (not) E" indicates stop and declare E (not) promising.

Use, for example, $\{(L_n, U_n) \equiv (0.05, 0.95)\}$.

- **Frequentist properties:** Evaluate type-I error, power, etc.
- **Design pars:** Adjust design pars (e.g., L_n, U_n) to achieve desired frequentist properties.

† In specific, *how* to specify design pars (e.g. L_n, U_n)?

- Set L_n and U_n at some reasonable first choice, say, $L_n = 1\%$, $U_n = 80\%$.
- Consider two scenarios and do simulations under each.

  ⋆⋆ a null scenario $S0$ with $\theta_E = \theta_S$: Examine how many times we yield the (wrong) conclude that E is promising ($\Rightarrow$ Type I error).

  ⋆⋆ an alternative scenario $S1$ with $\theta_E = \theta_S + \delta$: Examine how many times we yield the (correct) conclude that E is promising ($\Rightarrow$ Power=1 - Type II error).

- A sequence of iterative corrections will eventually lead to a set of bounds that achieve desirable operating characteristics.

† Binary stopping for futility, efficacy and toxicity: Thall et al. (1995)

- **Toxicity:** Extend the framework to include monitoring for toxicity.
- **Elementary events:** Let CR: efficacy event & Tox: toxicity event.

  Define a set of elementary events as a partition of possible outcomes. For example,

$$A_1 = (CR, TOX), \qquad A_2 = (\text{ no } CR, TOX),$$
$$A_3 = (CR, \text{ no } TOX), \qquad A_4 = (\text{ no } CR, \text{ no } TOX)$$

such that efficacy and toxicity are unions of these events:

$$\text{efficacy} = A_1 \cup A_3; \quad \text{toxicity} = A_1 \cup A_2$$

$\sum_{j=1}^{4} p_{Tj} = 1 \quad 0 < p_{Tj} < 1$

- **Probability model:** $p(A_j) = p_{Tj}$ under treatment $T \in \{E, S\}$.
- **Prior:** $(p_{T1}, p_{T2}, p_{T3}, p_{T4}) \sim \text{Dir}(\theta^0_{T1}, \ldots, \theta^0_{T4})$, $T \in \{E, S\}$.
- **Posterior:** Let y^n_j = number of patients with A_j

$$p(p_{E1}, \ldots, p_{E4} \mid y^n) = \text{Dir}(\theta^n_{E1}, \ldots, \theta^n_{E4})$$

with updated parameters $\theta^n_{Ej} = \theta^0_{Ej} + y^n_j$.
And $\theta^n_{Sj} = \theta^0_{Sj}$, since no patients are assigned to S.

- **Posterior probs:** Let $\eta_T(CR) = p_{T1} + p_{T3}$ and $\eta_T(TOX) = p_{T1} + p_{T2}$ ($p(A_1)$, $p(A_3)$; $p(A_1) + p(A_2)$)

$$\begin{aligned} p(\eta_T(CR) \mid y^n) &= \text{Be}(\theta^n_{T1} + \theta^n_{T3}, \theta^n_{T2} + \theta^n_{T4}) \\ p(\eta_T(TOX) \mid y^n) &= \text{Be}(\theta^n_{T1} + \theta^n_{T2}, \theta^n_{T3} + \theta^n_{T4}) \end{aligned}$$

- **Decision rule:** Use thresholds on posterior probabilities of clinically meaningful events:

$$\begin{aligned} \pi_n(CR) &= Pr(\eta_E(CR) > \eta_S(CR) + \delta_{CR} \mid y^n), \\ \pi_n(TOX) &= Pr(\eta_E(TOX) > \eta_S(TOX) + \delta_{TOX} \mid y^n) \end{aligned}$$

Sequential stopping rule

$$\text{decision} = \begin{cases} \text{stop for futility} & \text{if } \pi_n(CR) \leq L_n(CR) \\ \text{stop for toxicity} & \text{if } \pi_n(TOX) > U_n(TOX) \\ \text{stop for efficacy} & \text{if } \pi_n(CR) > U_n(CR) \\ \text{continue} & \text{otherwise} \end{cases}$$

- **Computation:**

  ⋆⋆ After each patient cohort, $\pi_n(\cdot)$ are updated.

  ⋆⋆ evaluation of $\pi_n(CR)$ requires integration with respect to the two independent Be r.v's

† Proper Bayes– Algorithm 4.2

- **0. Initialization:** Initialize θ_E^0.

  Set $n = 0$ (# patients), $y_j^n = 0$ (# events A_j), and $k = 4$ (cohort size). Fix simulation truth $p_{Ej}^o \equiv Pr(A_j)$

- **1. Posterior updating:** $\theta_n = \theta_{Ej} + y_j^n$.

- **2. Evaluate posterior probabilities:** Evaluate $\pi_n(CR)$ and $\pi_n(TOX)$ by Monte Carlo simulation.

- **3. Sequential stopping:** Evaluate the decision rule

  ⋆⋆ If $\pi_E(CR) < L_n(CR)$, then stop for lack of efficacy.

  ⋆⋆ If $\pi_E(TOX) > U_n(TOX)$, then stop for excessive toxicity.

  ⋆⋆ If $\pi_E(CR) > U_n(CR)$, then stop for efficacy.

  ⋆⋆ If $n_1 > N_{\max}$, then stop for maximum enrollment.

  Otherwise continue with Step 4.

† Proper Bayes– Algorithm 4.2: contd

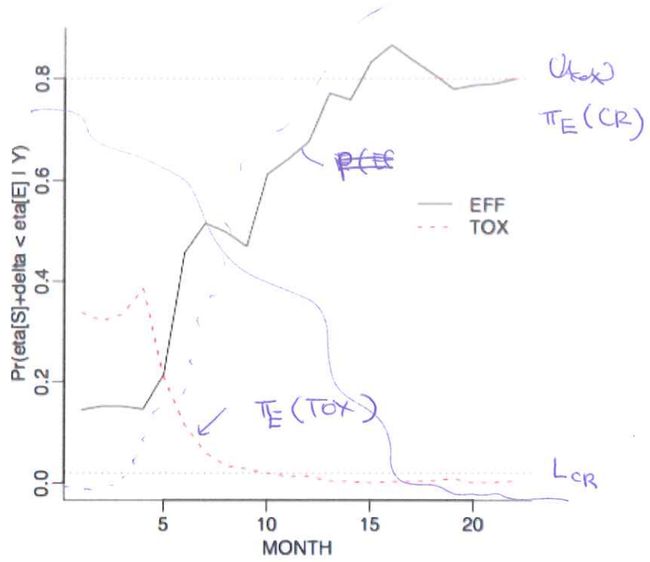
- **4. Next cohort:** If $n < N_{\max}$, then recruit a new cohort using $Pr(x_i = j) = p^o_{Ej}$. Update $n \equiv n + k$. Repeat from step 1.

† **Example 4.3**: Thall, Simon and Estey (1995)

- **Study:** bone marrow transplant (BMT), phase II trial of post-transplant prophylaxis for graft versus host disease (GVHD: A condition that occurs when donor bone marrow or stem cells attack the recipient from googling). $CR = \overline{G}$
- **Efficacy Endpoint:** GVHD within 100 days post transplant, $\overline{G}$ = no GVHD within 100 days (CR)
- **Toxicity Endpoint:** transplant rejection, T = transplant rejection within 100 days ("TOX")
- **Elementary outcomes:** $A_1 = \overline{GT}$, $A_2 = \bar{G}T$, $A_3 = G\bar{T}$, $A_4 = GT$.
  $\overline{C}\,\overline{T}$ $C\overline{T}$ CT $\overline{C}\,\overline{T}$

† **Example 4.3**: Thall, Simon and Estey (1995) – contd

- **Prior:** $p(p_{E1}, \ldots, p_{E4}) = \text{Dir}(\theta_{E1}, \ldots, \theta_{E4})$ with $\theta_E \propto (2.037, 6.111, 30.555, 2.037)$ and $\sum \theta_{Ej} = 4$.
- **Design:** $\delta_{CR} = 20\%$, $\delta_{TOX} = 5\%$, and $N_{\max} = 75$. $L_{CR} = 2\%$ and $U_{TOX} = 80\%$ (both constant over n). U_{CR} is not used (i.e., no early stopping for efficacy).
- **Simulation Truth:** $\eta_E(TOX) = 10\%$ and $\eta_E(CR) = 40\%$ vs $\eta_S(TOX) = 20\%$ and $\eta_S(CR) = 20\%$



* Posterior estimated probabilities $\pi_N(CR)$ (black, solid line) and $\pi_n(TOX)$ (red, dashed line) against month

† Adaptive Randomization and Dose Allocation (BCLM Section 4.4)

- Consider a randomized PhaseIIB multi-arm clinical trial
  - ⋆⋆ Multiple arms?
  - – different treatment (experimental treatment vs control)
  - – different doses or schedules of the same agent.

- **Randomization:** multi-arm trials usually include randomization, but need not be uniform!

- **Types of adaptation:** we consider outcome-adaptive designs (not covarite-adaptive or deterministic adaptive designs)

play-the-winner

† **Study:**

- Consider a 2 or multi-arm study, with treatment specific parameters θ_k, $k = 1, \ldots, K$, with larger (efficacy) or smaller (toxicity) values desirable.

- For example, suppose a binary efficacy outcome Y_i (CR) for patient i allocated to treatment k, i.e., $d_i = k$

$$\theta_k = p(Y_i = 1 \mid d_i = k)$$

For Bayesian flavor, $\theta_k \overset{indep}{\sim} \text{Be}(\alpha_k, \beta_k)$

We have a posterior probability distribution of $\boldsymbol{\theta}$.

† Adaptive Randomization: Thall and Wathen (2007, Europ J. Cancer)

Adaptive allocation: basic rule for treatment allocation

$$p(d_i = k \mid \mathbf{y}) \propto \{p(\theta_k = \max \theta_j \mid \mathbf{y})\}^c \qquad (AA)$$

$\theta_1, \ldots, \theta_K$

Details: usually,

- only start adaptive allocation after minimum sample size, $N_{\min}$ (safeguard against excessively adapting away from a trt arm)
- **Software:** AR package, for download from the MDACC software site; (AA) for binary and TITE outcome – see page 156 for more.

If we have two arms, A, B

$$P(\theta_A = \max_j \theta_j \mid y) = P(\theta_A > \theta_B \mid y)$$

$$P(d_i = A \mid y) \propto (P(\theta_A > \theta_B \mid y))^c$$

$$P(d_i = B \mid y) \propto (P(\theta_B > \theta_A \mid y))^c$$

$$= (1 - P(\theta_A > \theta_B \mid y))^c$$

Adaptive allocation: basic rule for treatment allocation

$$p(d_i = k \mid \mathbf{y}) \propto \{p(\theta_k = \max \theta_j \mid \mathbf{y})\}^c \qquad (AA)$$

- power c

  $\star\star$ $c = 0 \Rightarrow$ equal randomization.

  $\star\star$ larger value of c: potentially greater deviation from equal randomization.

  $\star\star$ power c close to 1.0 and perhaps no bigger than 2.

  $\star\star$ Based on empirical evidence, Thall and Wathen (2007) recommend $c = n/(2N_{\max})$ where $N_{\max}$: max # of patients and n: # of currently enrolled patients.

  $*$ That is, c is increasing gradually as more patients are accrued.

† Algorithm 4.3: Phase II AR design — Early stopping + Adaptive randomization.

Early stopping: early loser (allow re-activation), drop arms for futility (no re-activation), declare early winner (and stop trial):

- *Early loser:* $p(\theta_k > \theta_{j \neq k} \mid \mathbf{y}) < p_L$

  ⋆⋆ If the prob that trt arm k is the best falls below p_L, then arm k is declared a loser and is suspended.

  ⋆⋆ $p_L = 0.1$ or less

  ⋆⋆ trt arm k can return to active status later in the trial if the other arms grow worse and arm k becomes competitive again.

† Algorithm 4.3: Phase II AR design (contd)

- *Early winner:* $p(\theta_k > \theta_{j \neq k} \mid \mathbf{y}) > p_U$
  - ⋆⋆ If the prob that trt arm k is the best exceeds p_U, then arm k is declared the winner and the trial is stopped early.
  - ⋆⋆ p_U fairly large
- *Final winner:* $p(\theta_k > \theta_{j \neq k} \mid \mathbf{y}) > p_U^\star$
  - ⋆⋆ If, *after all patients have been evaluated,* the prob that trt arm k is the best exceeds $p_U^\star$, arm k is the winner. If no trt makes this criterion, AR dose not make a final decision.
  - ⋆⋆ Usually $p_U^\star < p_U$ (stronger evidence for early decision).

† Algorithm 4.3: Phase II AR design– contd

- *Futility:* $p(\theta_k > \theta_{\min} \mid \mathbf{y}) < p_L^\star$

  ⋆⋆ If the prob that trt arm k is better than $\theta_{\min}$ (minimally tolerable response rate), then arm k is declared futile and will no accrue more patients (can't be re-activated).

  ⋆⋆ $p_L^\star$: quite small such as 0.1 or less

† **Example 4.4:** Sensitizer trial – to assess the efficacy of a "sensitizer"

- **Study:** Phase II trial, 2 arms, sensitizer concurrently with chemo. ⇒ drug-plus-sensitizer vs drug alone
- **Outcome:** CR (by 28 days post treatment)
- **Prior:** Beta priors, with prior mean 0.55 (sd: 0.10) for control and 0.75 (sd: 0.13) for sensitizer.
- **Design:** $c = 1$, $p_L = 0.025$, $p_U = 0.975$, $p_U^\star = 0.90$, $\theta_{\text{min}} = 0.5$.
  ⋆⋆ $N_{\text{max}} = 60$ and the first 14 patients are fairly randomized before the adaptive randomization begins.

* Tab 4.3 Results: Under 3 scenarios, null, alternative and optimistic alternative, we find the following operating characteristics:

Arm	True Pr (success)	Pr (select)	Pr(select early)	Pr(stop early)	# Patients (2.5%,97.5%)
Scenario 1: null. Average trial length: 22.5 months					
Arm1	0.55	0.01	0	0.11	19.6 (5, 38)
Arm2	*0.55*	0.16	0.11	0	35.6 (8, 53)
Scenario 2: alternative. Average trial length: 16.4 months					
Arm1	0.55	0	0	0.55	10.1 (4, 22)
Arm2	*0.7*	0.74	0.55	0	30.8 (4, 51)
Scenario 3: optimistic. Average trial length: 10.8 months					
Arm1	0.55	0	0	0.89	7.01 (4, 16)
Arm2	*0.8*	0.96	0.89	0	20.1 (4, 51)

† Outcome Adaptive Randomization with Delayed Survival Response — Huang et al. (2009, StatMed)

- **Delayed response** T**:** desired final response (for example, progression-free survival, PFS)

  ⋆⋆ common in clinical trials, e.g., "GVHD within 100 days", any survival endpoint (PFS, OS etc.), "improvement in stroke score by week 13", etc.

  ⇒ Investigators have to wait for the response of the currently recruited patients before being able to make a decision about sequential stopping or treatment allocation for the next patient.

- **Bayesian Paradigm:** always conditional on *observed* data, including censored response etc.

  ⇒ Investigators can proceed on the basis of the partial information.