2.2 Phase II Clinical Trials

Phase II clinical trials usually are conducted to assess

- feasibility of treatment
- side effects and toxicity
- logistics of administration and cost
- dose finding (lowest dose level with good efficacy)

Major issue: Is there enough evidence of efficacy of the new drug to move to phase III?

Surrogate markers are often used.

Usually, one-arm (no comparison)

Example: Suppose a new drug is developed for patients with lung cancer. Ultimately, we would like to know whether this drug will extend the life of lung cancer patients as compared to currently available treatments. Establishing the effect of a new drug on survival would require a long study with relatively large number of patients and thus may not be suitable as a screening mechanism. Instead, during phase II, the effect of the new drug may be assessed based on tumor shrinkage in the first few weeks of treatment. If the new drug shrinks tumors sufficiently for a sufficiently large proportion of patients, then this may be used as evidence for further testing.

In this example,
Overall (or disease-free) survival time = **clinical endpoint**
tumor shrinkage = **surrogate markers**

Other examples of surrogate markers are

- Lowering blood pressure or cholesterol for patients with heart disease
- Increasing CD4 counts or decreasing viral load for patients with HIV disease

Caution in using **surrogate markers**:

- surrogacy is difficult to establish.
- If the new drug has no effect on the **surrogate markers**, it is probably more likely that the new drug will have no effect on the final clinical endpoint.
- However, sometimes it is possible that the new drug may have effect on the surrogate markers but have no effect on the final clinical endpoint.

Statistical Issues and Methods:

Goal: estimate the effect of the new drug on some endpoint (a surrogate marker, safety endpoint, etc) with enough precision to decide whether we investigate the new drug in phase III.

Examples:

- probability of a random patient responding to treatment (response has to be unambiguously defined)
- probability that a treated patient has side effects
- average decrease in blood pressure over a two week period

Consider a **binary** endpoint:

whether or not a patient responds to the new drug; whether or not a patient will have side effects, etc.

Suppose n patients are treated with the new drug:

$$X \sim bin(n, \pi)$$

- $X =$ total number of patients who respond to the new drug
- $\pi =$ population response rate (if the whole patient population is given the new drug.)

One objective: estimate π with enough precision.

Properties of a binomial distribution:

- $E(X) = n\pi$, where $E(\cdot)$ denotes the expectation of a random variable.
- $Var(X) = n\pi(1-\pi)$, where $Var(\cdot)$ denotes the variance of a random variable.
- $P(X = k) = \left(\begin{array}{c} n \\ k \end{array} \right) \pi^k (1-\pi)^{n-k}$, where $P(\cdot)$ denotes the probability of an event, and $\left(\begin{array}{c} n \\ k \end{array} \right) = \frac{n!}{k!(n-k)!}$
- Denote the sample proportion by $p = X/n$, then
  - $E(p) = \pi$
  - $Var(p) = \pi(1-\pi)/n$
- When n is sufficiently large, the distribution of the sample proportion $p = X/n$ is well approximated by a normal distribution

with mean π and variance $\pi(1-\pi)/n$.

$$p \sim N(\pi, \pi(1-\pi)/n)$$

- A large sample $(1-\alpha)$ CI of π can be constructed as

$$p \pm z_{\alpha/2}\{p(1-p)/n\}^{1/2}.$$

- A large sample 95% CI of π:

$$p \pm 1.96\{p(1-p)/n\}^{1/2}.$$

- Can be used to calculate sample size n.

Example: Suppose our best guess for the response rate of a new drug is about 35%; if we want the precision of our estimator to be such that the 95% confidence interval is within 15% of the true π, then we need

$$1.96 \left\{ \frac{(.35)(.65)}{n} \right\}^{1/2} = .15,$$

or

$$n = \frac{(1.96)^2(.35)(.65)}{(.15)^2} = 39 \text{ patients.}$$

Exact Confidence Intervals

If $n\pi$ or $n(1-\pi)$ is small, then the normal approximation may not be accurate.

$\Longrightarrow$ **Exact CI**.

Definition: A $(1-\alpha)$-th confidence region (interval) for π: $\mathcal{C}(k)$ ($k =$ observed # of response) such that

$$P_\pi\{\mathcal{C}(X) \supset \pi\} \geq 1-\alpha, \text{ for all } 0 \leq \pi \leq 1.$$

Question: How to find $\mathcal{C}(k)$ for given k responses out of n patients?

Consider testing the following hypothesis:

$$H_0 : \pi = \pi_0 \quad v.s. \quad H_a : \pi \neq \pi_0$$

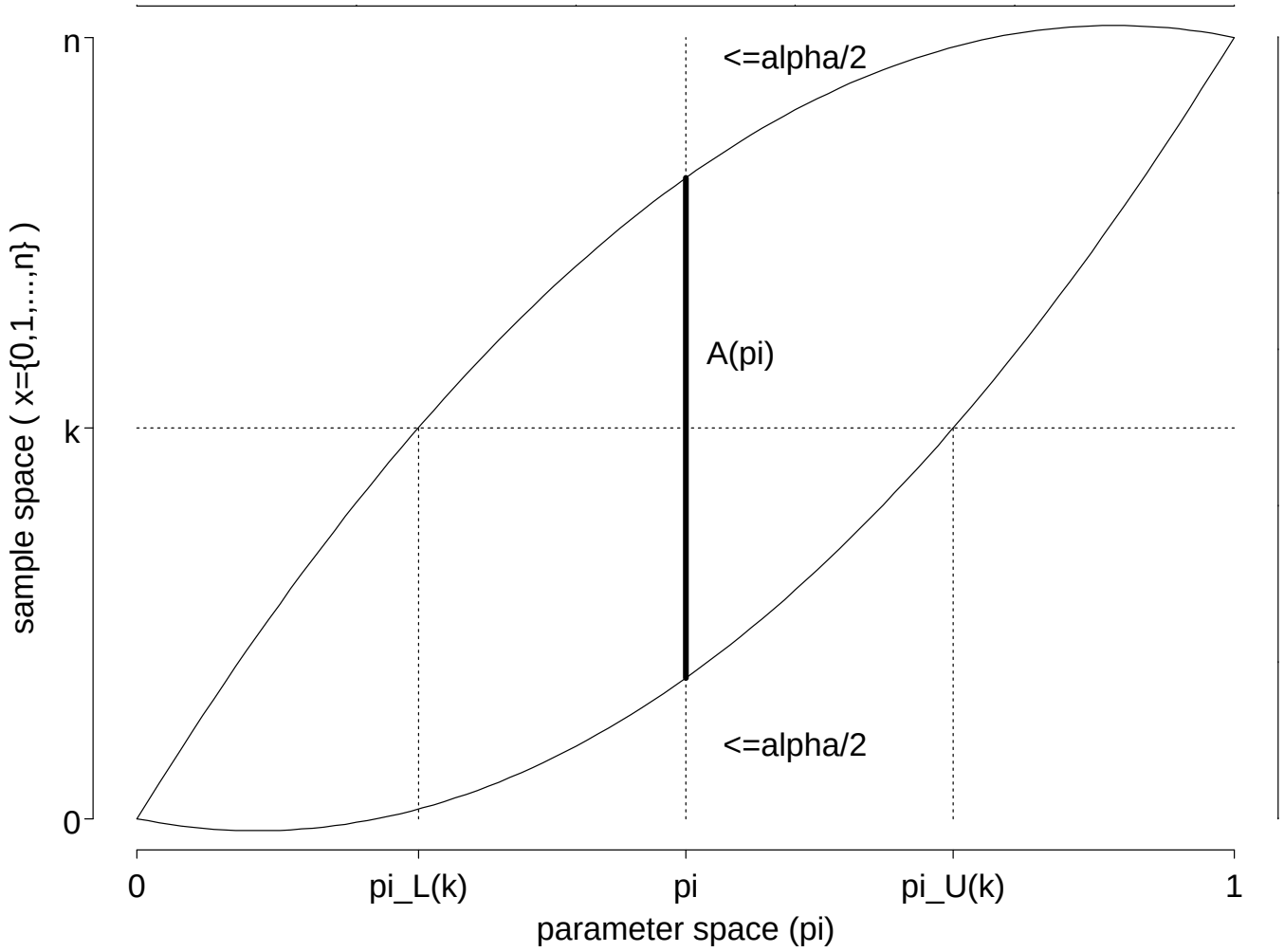
for some π_0.

- Intuitively, we would reject H_0 if k is too small or too large.
- Equivalently, we would accept H_0 if k is neither too small nor too large; that is, there is an interval (or region) $\mathcal{A}(\pi_0)$ (determined by π_0) such that we would not reject (and hence accept) H_0 if $k \in \mathcal{A}(\pi_0)$.
- $\mathcal{A}(\pi_0)$ is called the **acceptance region**.
- If we set the type I error probability of the above testing procedure at α, then we have:

$$P[X \in \mathcal{A}(\pi_0)|H_0] \geq 1 - \alpha \quad \text{for } X \sim bin(n, \pi_0).$$

- For given observed k, solving $k \in \mathcal{A}(\pi_0)$ will usually give us an interval $\mathcal{C}(k) = [\pi_L(k), \pi_U(k)]$, which is the exact $(1-\alpha)$ CI of π.
- Suppose we observe k responses out of n patients, then for any $\pi_0 \in [\pi_L(k), \pi_U(k)]$, we would not reject H_0
- That is, the above CI consists of all values of π_0 which is consistent with H_0 given k.
- **Question:** How to find $\pi_L(k)$ and $\pi_U(k)$?

Figure 1: *Exact confidence intervals*



- The figure indicates that for given k, $\pi_L(k)$ and $\pi_U(k)$ have to satisfy

$$P_{\pi_L(k)}[X \geq k] = \sum_{j=k}^{n} \binom{n}{j} \pi_L(k)^j \{1 - \pi_L(k)\}^{n-j} = \alpha/2,$$

$$P_{\pi_U(k)}[X \leq k] = \sum_{j=0}^{k} \binom{n}{j} \pi_U(k)^j \{1 - \pi_U(k)\}^{n-j} = \alpha/2.$$

- $\pi_L(k)$ and $\pi_U(k)$ can be solved using binomial tables or through statistical software.
- When $k = 0$, the first equation has no solution. Set $\pi_L(k) = 0$. $\pi_U(k)$ can be solved.
- When $k = n$, the second equation has no solution. Set $\pi_U(k) = 1$. $\pi_L(k)$ can be solved.

- **Remark:** Since X has a discrete distribution, the way we define the $(1-\alpha)$-th confidence interval above will yield

$$P_\pi\{\pi \in [\pi_L(k), \pi_U(k)]\} > 1 - \alpha$$

(strict inequality) for most values of $0 \leq \pi \leq 1$. Strict equality cannot be achieved because of the discreteness of the binomial random variable.

Example: Suppose in a Phase II clinical trial, 3 of 19 patients respond to α-interferon treatment for multiple sclerosis.

- 95% CI of π based Normal approximation:

$$\frac{3}{19} \pm 1.96 \left(\frac{\frac{3}{19} \times \frac{16}{19}}{19} \right)^{1/2} = [-.006, .322].$$

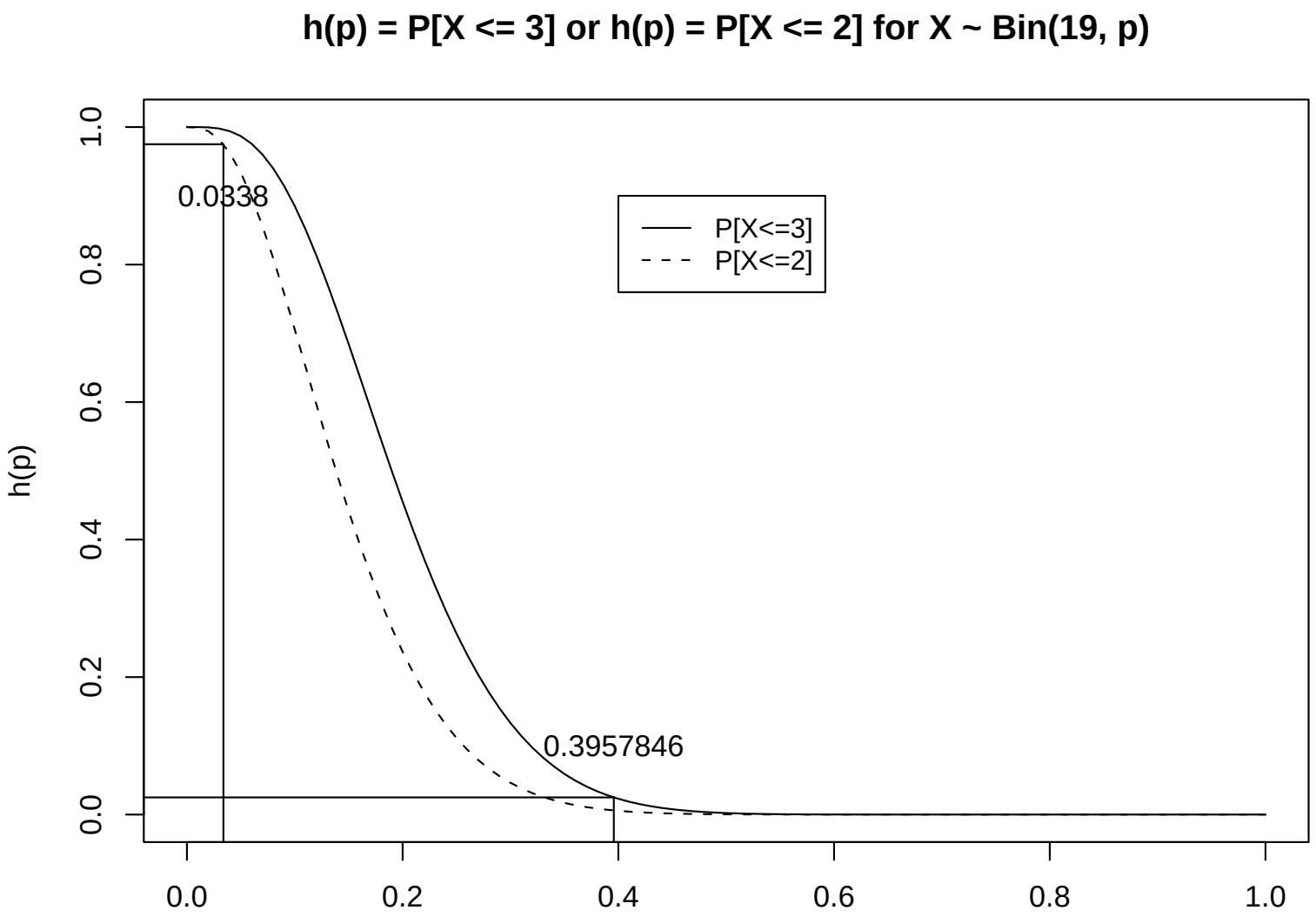
- Exact 95% CI: need to find out $\pi_L(3)$ and $\pi_U(3)$ $(n = 19, k = 3)$ such that

$$P_{\pi_L(3)}(X \geq 3) = .025 \Longleftrightarrow P_{\pi_L(3)}(X \leq 2) = .975$$
$$P_{\pi_U(3)}(X \leq 3) = .025.$$

$\Longrightarrow$ (using binomial tables)

$$\pi_L(3) \approx .03, \quad \pi_U(3) \approx .40$$

Figure 2: *Find Exact CI*



- SAS function for $P[X \leq k]$ where $X \sim bin(n, p)$:
  `probbnml(p, n, k)`
- R function for $P[X \leq k]$:
  `pbinom(k, n, p)`

```
options ls=80 ps=200 nodate;

data binprob;
  do pi=0.35 to 0.45 by 0.01;
    prob = probbnml(pi, 19, 3);
    output;
  end;
run;

proc print data=binprob;
run;
```

Obs	pi	prob
1	0.35	0.059140
2	0.36	0.049483
3	0.37	0.041180
4	0.38	0.034083
5	0.39	0.028053
6	0.40	0.022959
7	0.41	0.018683
8	0.42	0.015115
9	0.43	0.012156
10	0.44	0.009717
11	0.45	0.007719

Gehan's Two-Stage Design

One **goal** of a phase II trial is to discard ineffective treatments early.

Gehan's Two-Stage Design achieves this goal with 2 stages in a trial:

- Stage I: Give the new treatment to n_0 patients. If no patient responds, declare the treatment ineffective.
- Stage II: If at least one patient responds in stage I, add $n - n_0$ patients and count the total number of patients responding to the new treatment. Calculate point estimate of π and construct a CI for π.

How to determine n_0 and n?

- Determine n_0: denote $X = \#$ responses out of n_0 patients.

$$P[X = 0] = (1 - \pi)^{n_0}.$$

- $\pi_0 =$ minimal efficacy; that is, if $\pi \geq \pi_0$, we want to investigate the new drug in phase III.
- We would like to control the probability of discarding the new drug early if in fact it is promising.
- That is, n_0 has to satisfy:

$$P[X = 0] = (1 - \pi)^{n_0} \leq \alpha_0 \quad \text{for all } \pi \geq \pi_0,$$

where α_0 is our tolerance.

- Since $P[X = 0] = (1 - \pi)^{n_0}$ is an deceasing function of π, only need

$$(1 - \pi_0)^{n_0} \leq \alpha_0.$$

$$\Longrightarrow$$

$$n_0 \log(1 - \pi_0) \leq \log(\alpha_0)$$

$$\Longrightarrow$$

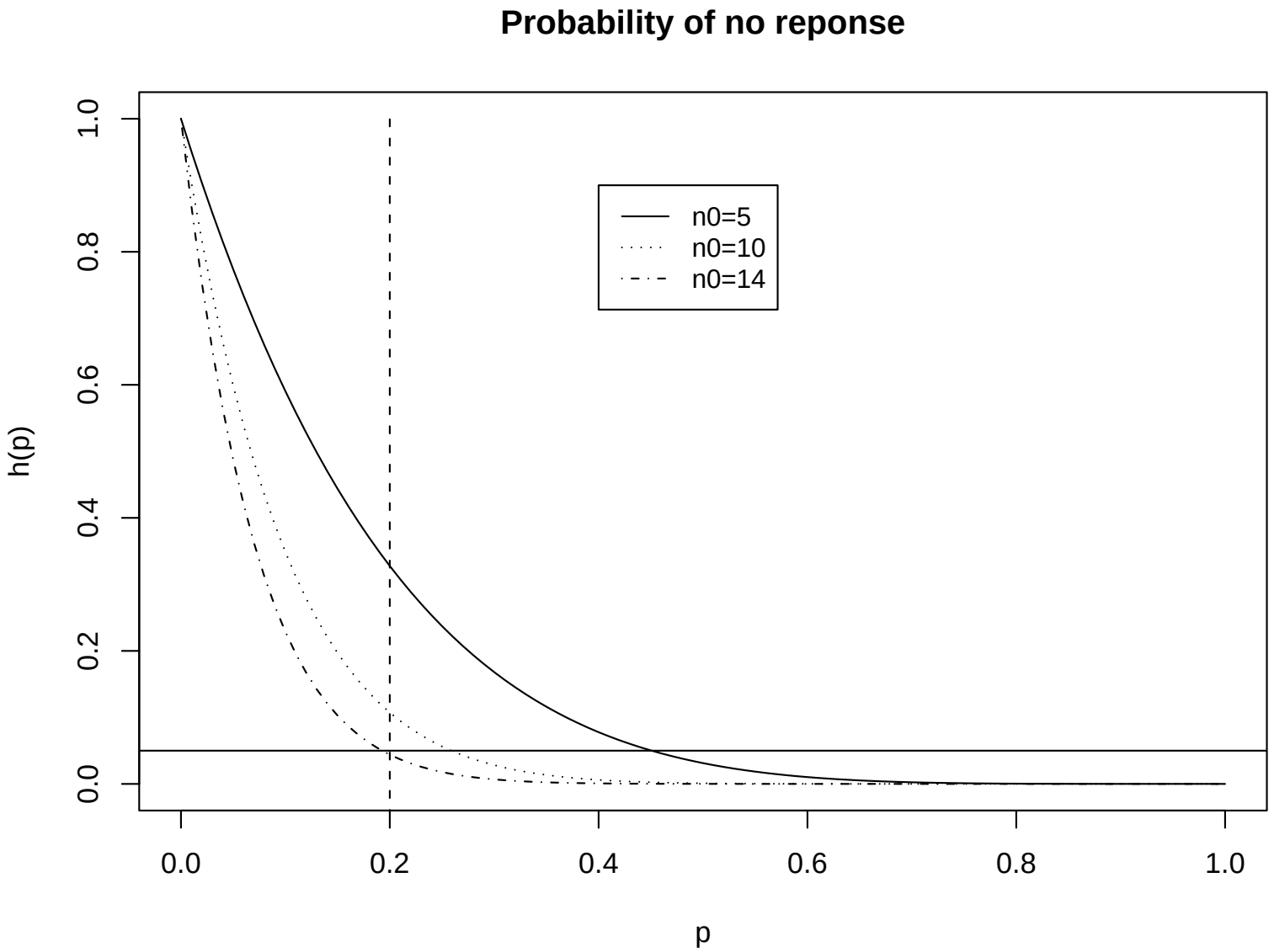
$$n_0 \geq \frac{\log(\alpha_0)}{\log(1 - \pi_0)}$$

- For example, $\pi_0 = 0.2$, $\alpha_0 = 0.05$, then

$$n_0 \geq \frac{\log(0.05)}{\log(1 - 0.2)} = 14 (\text{round up}).$$

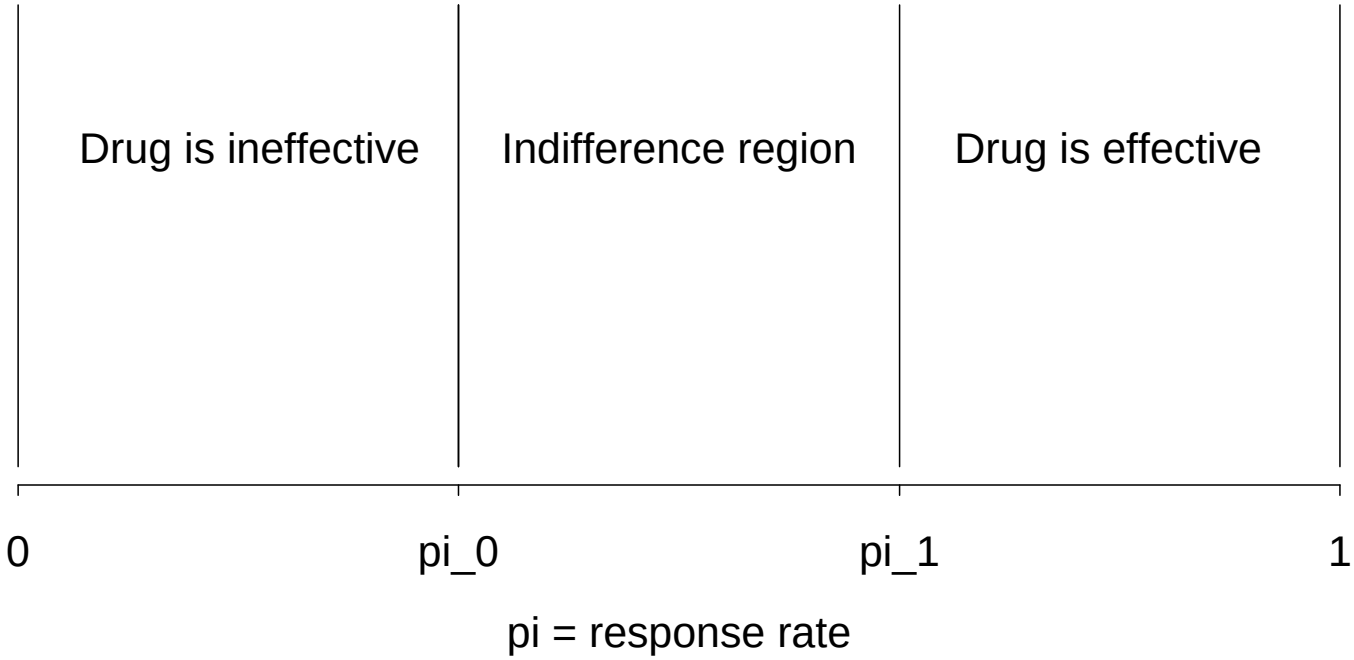
- Determine n: based on precision of 95% CI.
- For example, want to be 95% sure that the estimate is within $\pm 15\%$ of the minimum $\pi_0 = 0.2$:

$$1.96 \left(\frac{.2 \times .8}{n} \right)^{1/2} = .15, \text{ or } n = 28.$$


Probability of no reponse
h(p)
p
0.0
0.2
0.4
0.6
0.8
1.0
n0=5
n0=10
n0=14

Simon's Two-Stage Design

Suppose two values $\pi_0 < \pi_1$ are pre-specified such that



- If $\pi \leq \pi_0$, then we want to declare the drug ineffective with high probability, say $1 - \alpha$, where α is taken to be small.
- If $\pi \geq \pi_1$, then we want to consider this drug for further investigation with high probability, say $1 - \beta$, where β is taken to be small.

The values of α and β are generally taken to be between .05 and .20.

Simon's two-stage design proceeds as follows: Integers n_1, n, r_1, r, with $n_1 < n$, $r_1 < n_1$, and $r < n$ are chosen (to be described later) and

- n_1 patients are given treatment in the first stage. If r_1 or less respond, then declare the treatment a failure and stop.
- If more than r_1 respond, then add $(n - n_1)$ additional patients for a total of n patients.
- At the second stage, if the total number that respond among all n patients is greater than r, then declare the treatment a success; otherwise, declare it a failure.
- Of course, we stop the trial at stage 1 if the number of responses among n_1 patients is greater than r and declare the treatment a success.

$X_1 =$ the number of responses in stage 1 (out of n_1 patients)
$X_2 =$ the number of responses in stage 2 (out of $n_2 = n - n_1$ patients)

$$X_1 \sim b(n_1, \pi), \quad X_2 \sim b(n_2, \pi) \quad (X_1 \text{ and } X_2 \text{ are ind}).$$

- Declare the new drug a failure if

$$(X_1 \leq r_1) \text{ or } \{(X_1 > r_1) \text{ and } (X_1 + X_2 \leq r)\}$$

- The new drug is declared a success if

$$\{(X_1 > r_1) \text{ and } (X_1 + X_2) > r)\}.$$

- Design constraints $\Longrightarrow$

$$P\{(X_1 > r_1) \text{ and } (X_1 + X_2 > r)\} \leq \alpha \quad \text{for all } \pi \leq \pi_0 \quad (2.1)$$

$$P\{(X_1 > r_1) \text{ and } (X_1 + X_2 > r)\} \geq 1 - \beta \quad \text{for all } \pi \geq \pi_1 \quad (2.2)$$

- Denote the power function by

$$h(\pi) = P\{(X_1 > r_1) \text{ and } (X_1 + X_2 > r)|\pi\}.$$

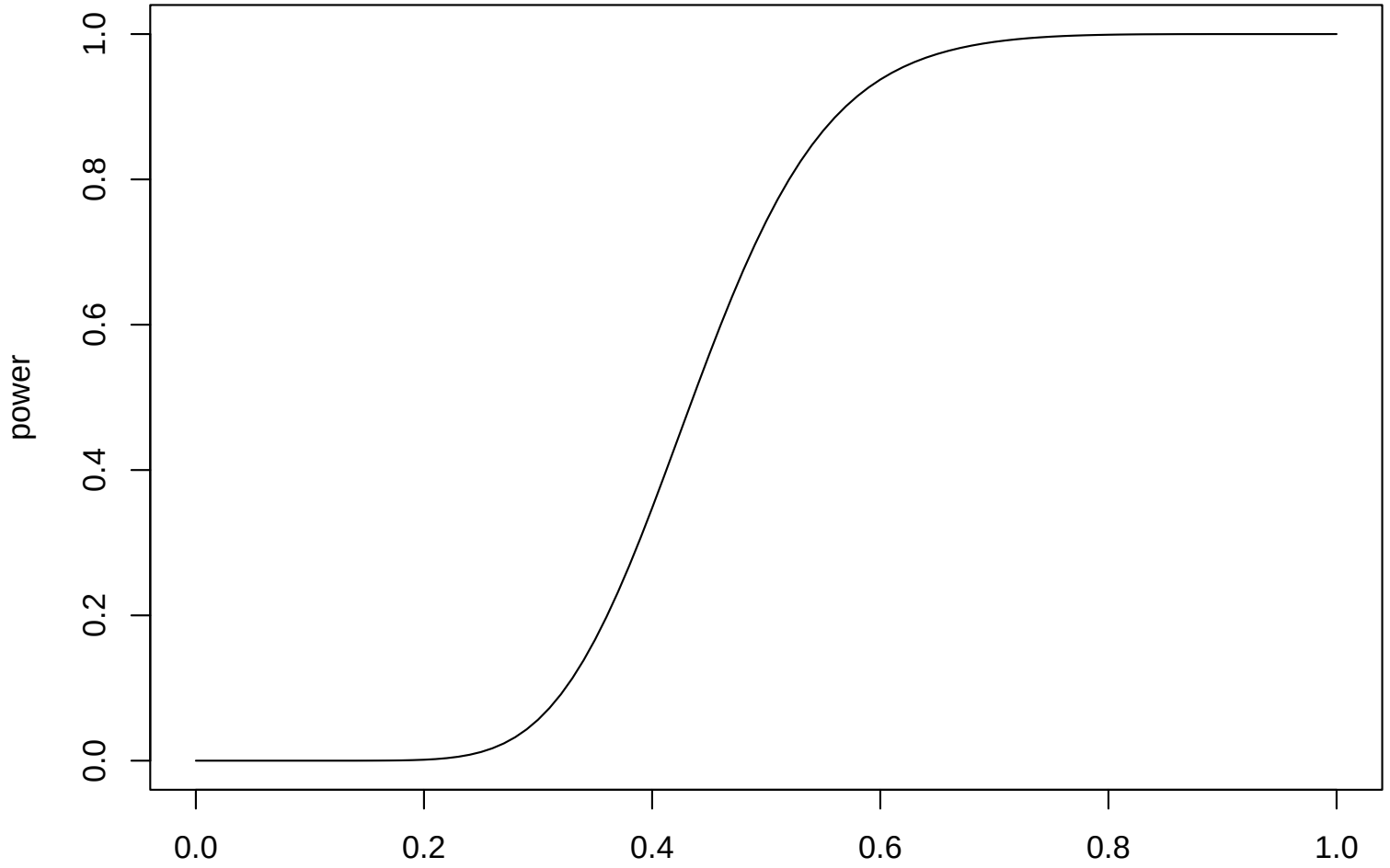
- It can be shown that $h(\pi)$ is an increasing function of π for any n_1, r_1, n, r.

- Therefore, criteria (2.1) and (2.2) are equivalent to

$$P\{(X_1 > r_1) \text{ and } (X_1 + X_2 > r)|\pi = \pi_0\} = \alpha$$

$$P\{(X_1 > r_1) \text{ and } (X_1 + X_2 > r)|\pi = \pi_1\} = 1 - \beta.$$

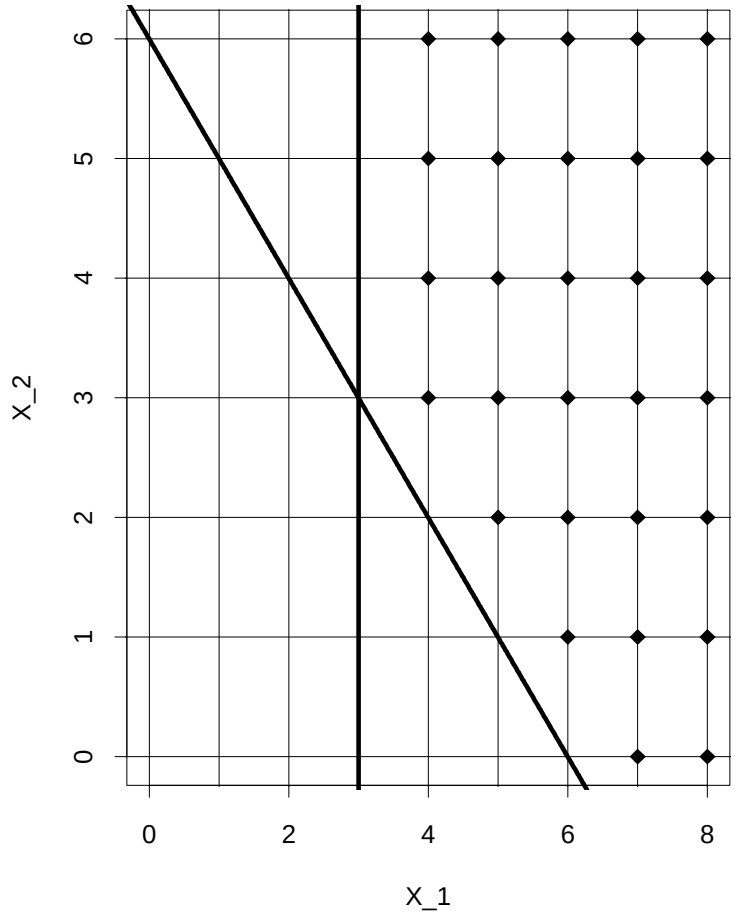
Figure 3: *Power Function of Simon's Design*



- How to calculate $P\{(X_1 > r_1) \text{ and } (X_1 + X_2 > r)|\pi\}$ for each π?
- By independence of X_1 and X_2:

$$
\begin{aligned}
& P\{(X_1 > r_1) \text{ and } (X_1 + X_2 > r)|\pi\} \\
= & \sum_{m1>r_1, m_1+m_2>r} P[X_1 = m_1, X_2 = m_2] \\
= & \sum_{m1>r_1, m_1+m_2>r} P[X_1 = m_1]P[X_2 = m_2]
\end{aligned}
$$

Figure 4: *Example:* $n_1 = 8,\ n = 14,\ X_1 > 3,\ and\ X_1 + X_2 > 6$



- Many combinations of (r_1, n_1, r, n) satisfy (2.1) and (2.2).
- "**Optimal design**" is the one that has smallest expected sample size when $\pi = \pi_0$ (when the new drug is ineffective)
- The expected sample size when $\pi = \pi_0$:

$$\begin{aligned} & n_1 P(\text{stopping at stage 1}) + nP(\text{did not stop at stage 1}) \\ = \quad & n_1\{P(X_1 \leq r_1 | \pi = \pi_0) + P(X_1 > r | \pi = \pi_0)\} \\ & + nP(r_1 + 1 \leq X_1 \leq r | \pi = \pi_0) \end{aligned}$$

- Through computer search, the **optimal design** can be identified.
- Simon, R. (1989). Optimal two-stage designs for Phase II clinical trials. *Controlled Clinical Trials*. 10: 1-10.
- Software can be downloaded from Dr. Simon's website (see class website)

Table 1 Designs for $p_1 - p_0 = 0.20$[a]

		Optimal Design				Minimax Design			
		Reject Drug if Response Rate				Reject Drug if Response Rate			
p_0	p_1	$\leq r_1/n_1$	$\leq r/n$	EN(p_0)	PET(p_0)	$\leq r_1/n_1$	$\leq r/n$	EN(p_0)	PET(p_0)
0.05	0.25	0/9	2/24	14.5	0.63	0/13	2/20	16.4	0.51
		0/9	2/17	12.0	0.63	0/12	2/16	13.8	0.54
		0/9	3/30	16.8	0.63	0/15	3/25	20.4	0.46
0.10	0.30	1/12	5/35	19.8	0.65	1/16	4/25	20.4	0.51
		1/10	5/29	15.0	0.74	1/15	5/25	19.5	0.55
		2/18	6/35	22.5	0.71	2/22	6/33	26.2	0.62
0.20	0.40	3/17	10/37	26.0	0.55	3/19	10/36	28.3	0.46
		3/13	12/43	20.6	0.75	4/18	10/33	22.3	0.50
		4/19	15/54	30.4	0.67	5/24	13/45	31.2	0.66
0.30	0.50	7/22	17/46	29.9	0.67	7/28	15/39	35.0	0.36
		5/15	18/46	23.6	0.72	6/19	16/39	25.7	0.48
		8/24	24/63	34.7	0.73	7/24	21/53	36.6	0.56
0.40	0.60	7/18	22/46	30.2	0.56	11/28	20/41	33.8	0.55
		7/16	23/46	24.5	0.72	17/34	20/39	34.4	0.91
		11/25	32/66	36.0	0.73	12/29	27/54	38.1	0.64
0.50	0.70	11/21	26/45	29.0	0.67	11/23	23/39	31.0	0.50
		8/15	26/43	23.5	0.70	12/23	23/37	27.7	0.66
		13/24	36/61	34.0	0.73	14/27	32/53	36.1	0.65
0.60	0.80	6/11	26/38	25.4	0.47	18/27	24/35	28.5	0.82
		7/11	30/43	20.5	0.70	8/13	25/35	20.8	0.65
		12/19	37/53	29.5	0.69	15/26	32/45	35.9	0.48
0.70	0.90	6/9	22/28	17.8	0.54	11/16	20/25	20.1	0.55
		4/6	22/27	14.8	0.58	19/23	21/26	23.2	0.95
		11/15	29/36	21.2	0.70	13/18	26/32	22.7	0.67

[a]For each value of (p_0, p_1), designs are given for three sets of error probabilities (α, β). The first, second and third rows correspond to error probability limits (0.10, 0.10), (0.05, 0.20), and (0.05, 0.10) respectively. For each design, EN(p_0) and PET(p_0) denote the expected sample size and the probability of early termination when the true response probability is p_0.

Table 2 Designs for $p_1 - p_0 = 0.15$[a]

		Optimal Design				Minimax Design			
		Reject Drug if Response Rate				Reject Drug if Response Rate			
p_0	p_1	$\leq r_1/n_1$	$\leq r/n$	EN(p_0)	PET(p_0)	$\leq r_1/n_1$	$\leq r/n$	EN(p_0)	PET(p_0)
0.05	0.20	0/12	3/37	23.5	0.54	0/18	3/32	26.4	0.40
		0/10	3/29	17.6	0.60	0/13	3/27	19.8	0.51
		1/21	4/41	26.7	0.72	1/29	4/38	32.9	0.57
0.10	0.25	2/21	7/50	31.2	0.65	2/27	6/40	33.7	0.48
		2/18	7/43	24.7	0.73	2/22	7/40	28.8	0.62
		2/21	10/66	36.8	0.65	3/31	9/55	40.0	0.62
0.20	0.35	5/27	16/63	43.6	0.54	6/33	15/58	45.5	0.50
		5/22	19/72	35.4	0.73	6/31	15/53	40.4	0.57
		8/37	22/83	51.4	0.69	8/42	21/77	58.4	0.53
0.30	0.45	9/30	29/82	51.4	0.59	16/50	25/69	56.0	0.68
		9/27	30/81	41.7	0.73	16/46	25/65	49.6	0.81
		13/40	40/110	60.8	0.70	27/77	33/88	78.5	0.86
0.40	0.55	16/38	40/88	54.5	0.67	18/45	34/73	57.2	0.56
		11/26	40/84	44.9	0.67	28/59	34/70	60.1	0.90
		19/45	49/104	64.0	0.68	24/62	45/94	78.9	0.47
0.50	0.65	18/35	47/84	53.0	0.63	19/40	41/72	58.0	0.44
		15/28	48/83	43.7	0.71	39/66	40/68	66.1	0.95
		22/42	60/105	62.3	0.68	28/57	54/93	75.0	0.50
0.60	0.75	21/34	47/71	47.1	0.65	25/43	43/64	54.4	0.46
		17/27	46/67	39.4	0.69	18/30	43/62	43.8	0.57
		21/34	64/95	55.6	0.65	48/72	57/84	73.2	0.90
0.70	0.85	14/20	45/59	36.2	0.58	15/22	40/52	36.8	0.51
		14/19	46/59	30.3	0.72	16/23	39/49	34.4	0.56
		18/25	61/79	43.4	0.66	33/44	53/68	48.5	0.81
0.80	0.95	5/7	27/31	20.8	0.42	5/7	27/31	20.8	0.42
		7/9	26/29	17.7	0.56	7/9	26/29	17.7	0.56
		16/19	37/42	24.4	0.76	31/35	35/40	35.3	0.94