Noninferiority Trial Designs for Risk Differences and Odds Ratios

Joan F. Hilton

e-mail: joan@biostat.ucsf.edu

Department of Epidemiology & Biostatistics
University of California San Francisco

Overview

Rationale for noninferiority trials:

The experimental regimen (E) has some *advantage* over the standard-of-care "control" (C) regimen that out-weighs the *disadvantage* of lower effectiveness.

Examples

- If shorter time course $\rightarrow$ fewer treatment-related adverse events (AEs), greater compliance and less development of treatment resistance
- If E more toxic than C with respect to reversible AEs, routine monitoring of patients should prevent long-term harm
- if different route of administration $\rightarrow$ less reliance on health care personnel and resources
- Suitable for people who are allergic to the control treatment

Special case: Noninferiority trials with $\pi_E = \pi_C$ under H_A

Questions to be addressed:

- Q: What are the fixed design parameters that must be specified?
- Q: What are their values?
- Q: What is the optimal value of the allocation ratio $\gamma = n_E/n_C$ – that is, the value that minimizes the overall sample size, N?

The design parameters of two-group noninferiority trials for efficacy with binary outcomes are

- a contrast parameter $\rightarrow$ Here, the odds ratio (ψ) or risk difference (δ)
- a response rate $\rightarrow$ Two values of interest: π_C and π^N

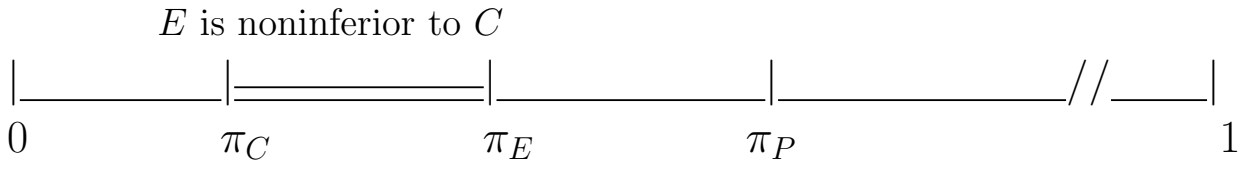
  – Positive response (eg, cure): $\delta_0 = \pi_E - \pi_C < 0 \Rightarrow E$ somewhat inferior to C

  – Negative response (eg, morbidity): $\delta_0 = \pi_E - \pi_C > 0 \Rightarrow E$ somewhat inferior to C

  – The test is one-sided.

The noninferiority contrast parameter is the "noninferiority margin."

Let the response be morbidity, so $\delta_0 > 0$.



Assume C is effective (i.e., $\pi_C < \pi_P$). Also assume E is effective (i.e., $\pi_E < \pi_P$).

We want to find the $N = n_E + n_C$ needed to detect $\delta_A - \delta_0$, with level α and power $1 - \beta$.

$$n_C = \frac{[z_{1-\alpha}\ \tilde{\sigma}_0(\hat{\delta}) + z_{1-\beta}\ \sigma_A(\hat{\delta})]^2}{(\delta_A - \delta_0)^2}$$

and

$$N = (1 + \gamma)\ n_C, \text{ where } \gamma = n_E/n_C.$$

- α and β are the type-1 and type-2 error rates,
- $z_{1-\alpha}$ and $z_{1-\beta}$ are critical values from the normal distribution

Q: What are the fixed design parameters that must be specified?

- Heuristic argument
- Derivation
- Comparison with findings of others (Table 1)

Q: What are their values?

Superiority and noninferiority trials have opposite relationships to H_0 and H_A.

Superiority design: $H_0 : \delta \geq \delta_0$ *vs.* $H_A : \delta < \delta_0$

	Contrast parameter	Response rate
Under H_0:	$\underline{\overline{\mid \delta_0 = \pi_E - \pi_C = 0 \mid}}$	π^S is estimated* (defines $\tilde{\sigma}_0$)
Under H_A:	$\delta_A = \pi_E - \pi_C < 0$	$\underline{\overline{\mid \pi_C \text{ is specified (defines } \sigma_A) \mid}}$

* $\bar{p} \equiv \hat{\pi}^S = (n_C\pi_C + n_E\pi_E)/N$, where $\hat{\pi}^S \in (\pi_C, \pi_E)$.

Noninferiority design: $H_0 : \delta \geq \delta_0$ *vs.* $H_A : \delta < \delta_0$

	Contrast parameter	Response rate
Under H_0:	$\underline{\overline{\mid \delta_0 = \pi_E - \pi_C > 0 \mid}}$	π_C is estimated (defines $\tilde{\sigma}_0$)
Under H_A:	$\delta_A = \pi_E - \pi_C = 0$	$\underline{\overline{\mid \pi^N \text{ is specified}^+ \text{ (defines } \sigma_A) \mid}}$

+ $\pi^N = (1-\omega)\tilde{\pi}_C + \omega\tilde{\pi}_E$, where $\pi^N \in (\tilde{\pi}_C, \tilde{\pi}_E)$ and $\omega \in (0, 1)$.

Constrained MLE for risk difference, for $y_C \sim Bi(n_C, \pi_C)$ and $y_E \sim Bi(n_E, \pi_E)$

Superiority: Specify π_C, $\delta_0 = 0$; estimate π^S

$$L_{H_0}(\pi) = [\pi^{y_C}(1-\pi)^{n_C - y_C}][\pi^{y_E}(1-\pi)^{n_E - y_E}], \text{ where } \pi \equiv \pi^S$$

MLE of π^S at the point-alternative of interest, $\pi_E - \pi_C \equiv \delta_A$, relates $\hat{\pi}^S$ with π_C:

$$\hat{\pi}^S = (1-\omega^S)\,\pi_C + \omega^S\,(\pi_C + \delta_A),$$
$$\text{where } \omega^S = \frac{\gamma^S}{\gamma^S + 1} = \frac{n_E}{N}.$$

$\Rightarrow \pi_C$ and δ_0 are fixed design parameters; π^S is estimated (for use in $\tilde{\sigma}_0(\hat{\delta})$). *Q:* $\gamma^S =$?

Noninferiority: Specify π^N, δ_0; estimate π_C

$$L_{H_0}(\pi_C) = \left[\pi_C^{y_C}(1-\pi_C)^{n_C - y_C}\right]\left[(\pi_C + \delta_0)^{y_E}\{1-(\pi_C+\delta_0)\}^{n_E - y_E}\right],$$

MLE of π_C at the point-alternative of interest, $\pi_E - \pi_C \equiv \delta_A = 0$, relates $\tilde{\pi}_C$ with π^N:

$$\pi^N = (1-\omega^N)\tilde{\pi}_C + \omega^N\,(\tilde{\pi}_C + \delta_0),$$
$$\text{where } \omega^N = \frac{\gamma^N}{\gamma^N + R}, \quad R = \frac{(\tilde{\pi}_C + \delta_0)\{1-(\tilde{\pi}_C+\delta_0)\}}{\tilde{\pi}_C(1-\tilde{\pi}_C)}.$$

$\Rightarrow \pi^N$ and δ_0 are fixed design parameters; π_C is estimated (for use in $\tilde{\sigma}_0(\hat{\delta})$). *Q:* $\gamma^N =$?

Superiority and noninferiority trials have opposite relationships to H_0 and H_A.

Superiority design: $H_0 : \delta \geq \delta_0$ *vs.* $H_A : \delta < \delta_0$

	Contrast parameter	Response rate	$SE(\hat{\delta})$
Under H_0:	$\underline{\mid \delta_0 = \pi_E - \pi_C = 0 \mid}$	π^S is estimated*	$\tilde{\sigma}_0 = \left[\bar{p}(1-\bar{p})\left(\frac{1}{n_C} + \frac{1}{n_E}\right)\right]^{0.5}$
Under H_A:	$\delta_A = \pi_E - \pi_C < 0$	$\underline{\mid \pi_C \text{ is specified } \mid}$	

* $\bar{p} \equiv \hat{\pi}^S = (n_C\pi_C + n_E\pi_E)/N$, where $\hat{\pi}^S \in (\pi_C, \pi_E)$.

Noninferiority design: $H_0 : \delta \geq \delta_0$ *vs.* $H_A : \delta < \delta_0$

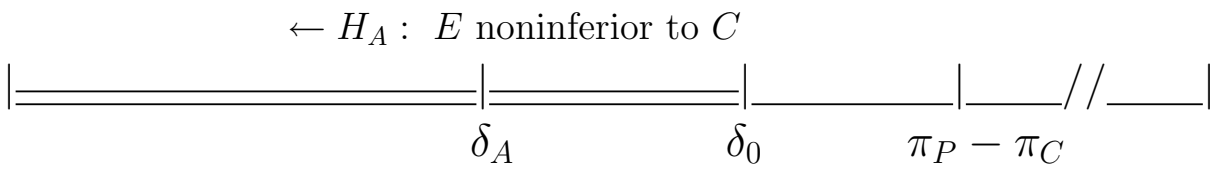
	Contrast parameter	Response rate	$SE(\hat{\delta})$
Under H_0:	$\underline{\mid \delta_0 = \pi_E - \pi_C > 0 \mid}$	π_C is estimated	
Under H_A:	$\delta_A = \pi_E - \pi_C = 0$	$\underline{\mid \pi^N \text{ is specified}^+ \mid}$	$\sigma_A = \left[\pi^N(1-\pi^N)\left(\frac{1}{n_C} + \frac{1}{n_E}\right)\right]^{0.5}$

\+ $\pi^N = (1-\omega)\tilde{\pi}_C + \omega\tilde{\pi}_E$, where $\pi^N \in (\tilde{\pi}_C, \tilde{\pi}_E)$ and $\omega \in (0,1)$.

When the response is a negative outcome, the hypotheses on the risk-difference scale are:

- Superiority trial under H_A: π_E should be lower than π_C ($\delta_A = \pi_E - \pi_C < 0$) and $\delta_A < \delta_0$.

- Noninferiority trial under H_0: Allow π_E to be higher than π_C ($\delta_0 = \pi_E - \pi_C > 0$) and $\delta_A < \delta_0$.



If the response were cure (positive outcome), the alternative hypotheses would be in the opposite direction.

At this point,

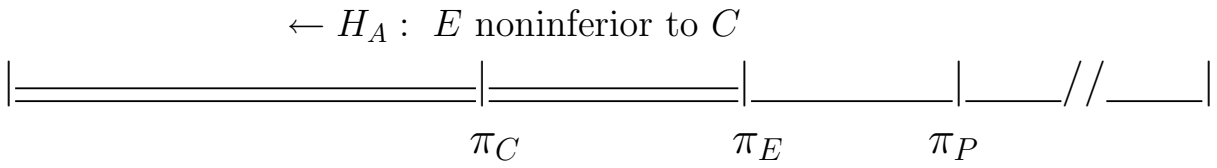
$\Rightarrow$ *A: We know what the fixed parameters are.*

$\Rightarrow$ *A: We can specify the values of the contrast parameter under both* H_0 *and* H_A*.*

When the response is a negative outcome, the hypotheses on the response-rate scale are:

- Superiority trial under H_A: π_E should be lower than $\pi_C + \delta_0$ ($\delta_0 = 0$); i.e., $\delta_A < \delta_0$.

- Noninferiority trial under H_A: π_E should be lower than $\pi_C + \delta_0$ ($\delta_0 > 0$); i.e., $\delta_A < \delta_0$.



⇒ **Superiority**: Assuming a fixed allocation ratio, $\gamma = n_E/n_C$, we can calculate the value of $\hat{\pi}^S = (n_C\pi_C + n_E\pi_E)/N$ because it is a function of known parameters. Usually, $\gamma = 1$ (balanced design).

⇒ **Noninferiority**: We still don't know the value of π^N, except it lies between $\tilde{\pi}_C$ and $\tilde{\pi}_E$.
Q: What allocation ratio should we use?

Noninferiority Algorithms

Recall that

$$\pi^N = \tilde{\pi}_C + \omega\delta_0, \quad \text{for } \delta_0 = \tilde{\pi}_E - \tilde{\pi}_C.$$

- Given design parameters $\{\pi^N, \delta_0\}$, there is a 1:1 correspondence between $\tilde{\pi}_C$ and γ.
  - Loop through $\max(\pi^N - \delta_0, .001) \leq \tilde{\pi}_C \leq \pi^N$ by .0001; solve for $\{\gamma_\theta, \tilde{\pi}_C\}$.
    - Typically, $\gamma_\theta \in (.25, 4.0)$.
  - Once $\{\gamma_\theta, \tilde{\pi}_C\}$ are found, $\tilde{\sigma}_0(\hat{\delta})$, $\sigma_A(\hat{\delta})$, and N can be calculated.
  - Find "optimal" pairs $\{\gamma_\theta, \tilde{\pi}_C\}$ such that N is minimized.
    - Because $N = (1+\gamma)\, n_C$ depends on γ both directly and through n_C, N a concave function of γ. {Small-γ,Large-n_C} and {Large-γ,Small-n_C} can yield the same N.
    - Because the algorithm must specify $\tilde{\pi}_C$ first, γ doesn't exactly equal n_E/n_C.

Preliminary Problem: π^N is unknown . . . but π_C is known (call this "π_{C_0}").

- *Midpoint Approach* (closed form): Let π^N be the midpoint between π_{C_0} and $\pi_{C_0} + \delta_0$.
- *Tailored Approach* (iterative search; outer loop): Find π^N such that $|\tilde{\pi}_C - \pi_{C_0}|$ is negligible.

Q: Does the Approach *used to select* π^N *affect the results* $\{N, \gamma\}$*?*

Table 1. Replicate results of Farrington & Manning (1990) at $\{\alpha, 1-\beta\} = \{.05, .90\}$.

$\{\pi^N, \delta_0\}$	Goal	N	Range(s) of $\{\gamma, \vec{\pi}_C\}^*$
$\{.10, .20\}$	*Find* min(N)	83	$\{1.88, .0418\} - \{2.09, .0394\}$
	Find N at $\gamma = 1.5$	85	$\{1.43, .0480\} - \{2.71, .0338\}$
	Find N at $\gamma = 1.0$	92	$(\{0.96, .0573\} - \{1.04, .0554\}), (\{3.38, .0295\} - \{3.84, .0269\})$
	Find N at $\gamma = .67$	105	$\{0.66, .0656\} - \{0.67, .0653\}$
$\{.05, .20\}$	*Find* min(N)	55	$\{1.77, .0184\} - \{3.27, .0121\}$
	Find N at $\gamma = 1.5$	58	$\{1.33, .0218\} - \{3.84, .0107\}$
	Find N at $\gamma = 1.0$	64	$\{0.94, .0260\} - \{1.06, .0246\}$
	Find N at $\gamma = .67$	74	$\{0.65, .0304\} - \{0.68, .0299\}$
$\{.01, .20\}$	*Find* min(N)	28	$\{2.16, .0028\} - \{3.59, .0019\}$
	Find N at $\gamma = 1.5$	32	$\{1.31, .0039\} - \{1.87, .0031\}$
	Find N at $\gamma = 1.0$	36	$\{0.99, .0046\} - \{1.11, .0043\}$
	Find N at $\gamma = .67$	45	$\{0.66, .0056\}$

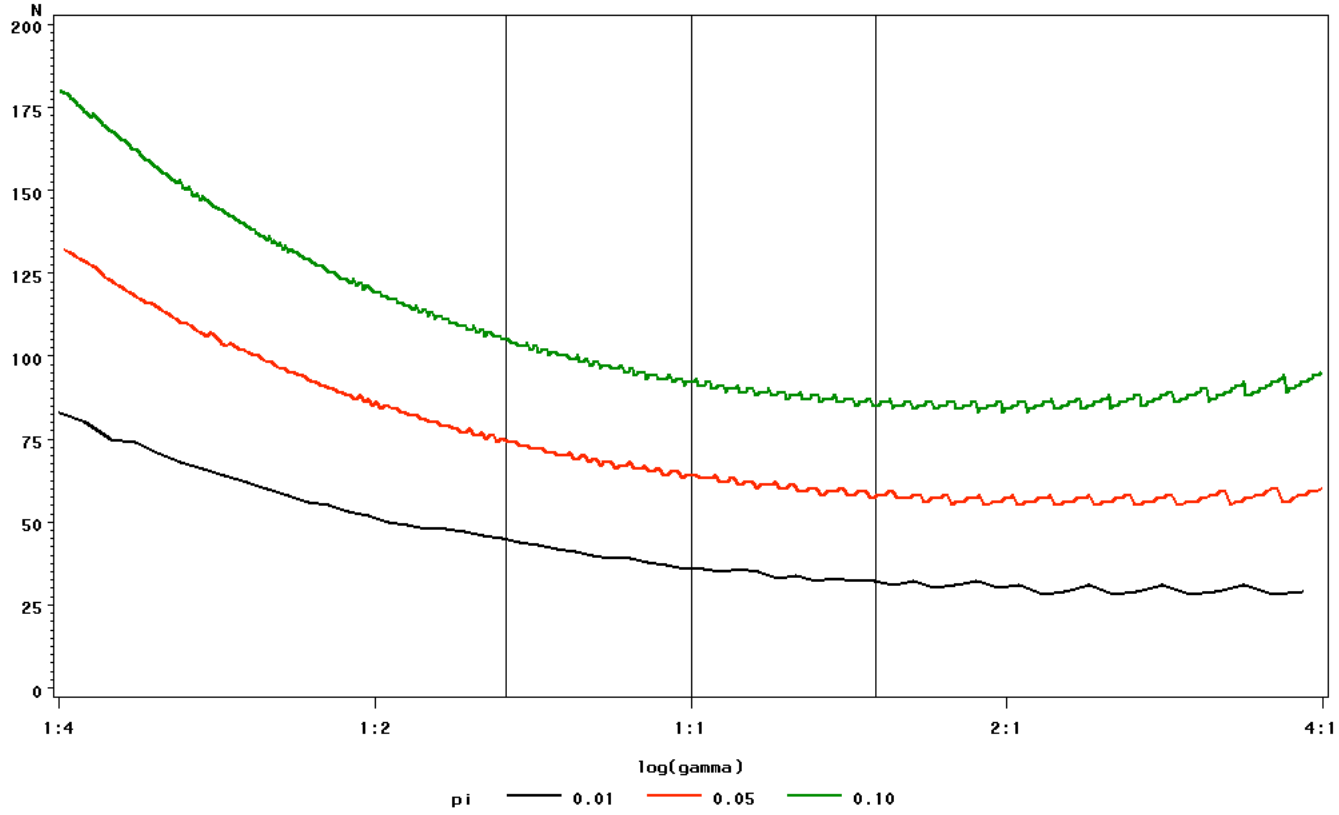
They fix γ at the design stage (on what basis?) and find corresponding N.

Algorithm results show:

- More than one value of γ yields the same N.
- Instead of pre-specifying γ, the values of γ associated with the smallest N can be found.

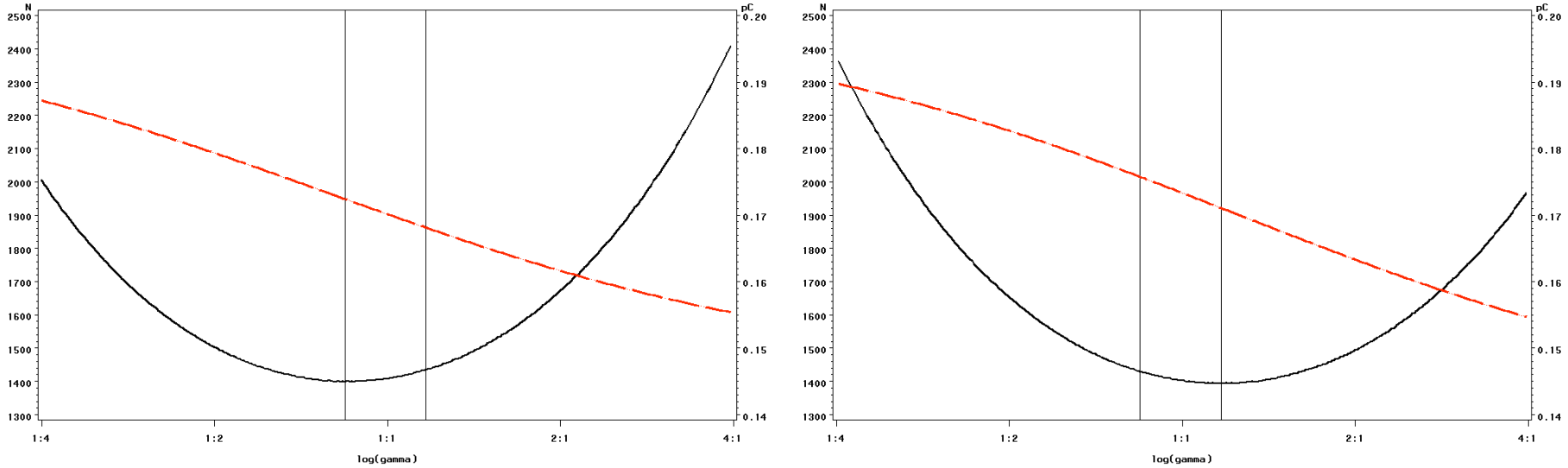
They also fix π^N (on what basis?).

Figure 1. At $\{\alpha, 1-\beta\} = \{.05, .90\}$ and $\delta_0 = .20$, $N \times \log \gamma$ relationships when $\pi^N = .10$ (green), .05 (red), and .01 (black). Vertical bars mark allocation ratios $\gamma = 0.67$, 1.0, and 1.5.



- Instead of pre-specifying γ, the values associated with the smallest N can be found.
- More than one value of γ yields the same N. We can choose among these without penalty.
- We can understand trade-off between non-optimal γ and N (e.g., choose balanced design).

Figure 2. At $\{\alpha, 1-\beta\} = \{.025, .80\}$, $N \times \log\gamma$ relationships for designs $\{\pi^N, \delta\} = \{.20, .06\}$ when the noninferiority margin is parameterized on the log-odds scale, $\theta = \log\psi$ (left), and on the risk-difference scale, $\theta = \delta$ (right). Vertical bars mark the corresponding minimum Ns.



	$\theta = \log\psi$			$\theta = \delta$		
	N	$\{\gamma, \vec{\pi}_C\}^*$	$\{\pi^N, \delta(\vec{\pi}_C)\}$	N	$\{\gamma, \vec{\pi}_C\}^*$	$\{\pi^N, \psi(\vec{\pi}_C)\}$
$\{\pi_C = .20, \psi = 1.405\}$:						
– Midpoint Approach	1,538	(.874, .202)	{.230, .0604}	1,541	(1.13, .201)	.{230, 1.404}
– Tailored Approach	1,548	(.881, .200)	{.228, .0600}	1,536	(1.14, .200)	.{229, 1.406}
$\{\pi^N = .20, \psi = 1.456\}$:	1,399	(.844, .172)	{.200, .0603}	1,393	(1.17, .171)	{.200, 1.456}

- Improperly specifying $\pi^N \equiv \pi_C$ substantially underestimates the sample size.
- When allocation ratios for $\theta = \delta$ are above 1.0, those for $\theta = \log\psi$ are below 1.0.
- *Approach* to specifying π^N affects N and should be reported in Methods.

Table 2. For $\{\alpha, 1-\beta\} = \{.025, .80\}$, (a) provides N_δ (upper) and $(\gamma_\delta;\ \vec{\pi}_C)$ (lower), and (b) provides $\psi(\vec{\pi}_C)$ (upper) and 'power of the analysis based on $\log \psi$, given a design based on δ'; $E[y_C]$ (lower).

	π^N	δ			
(a)		.01	.05	.10	.20
	.01	3,266	217	80	29
		(1.95; .0058)	(3.52; .0025)	(3.99; .0020)	(3.50; .0020)
	.05	14,936	619	170	51
		(1.16; .0451)	(1.83; .0292)	(2.66; .0184)	(3.08; .0126)
	.10	28,264	1,136	288	75
		(1.05; .0951)	(1.35; .0768)	(1.74; .0582)	(2.50; .0355)
	.20	50,232	2,007	500	123
		(1.04; .1950)	(1.10; .1762)	(1.28; .1532)	(1.67; .1125)

(b)		.01	.05	.10	.20
	.01	2.752	**22.11**	**56.68**	**129.5**
		(.672; 6)	(.056; 0)	(.007; 0)	(.002; 0)
	.05	1.237	2.860	**7.148**	**21.08**
		(.795; 312)	(.683; 6)	(.382; 1)	(.093; 0)
	.10	1.118	1.746	3.041	**8.369**
		(.799; 1,311)	(.774; 37)	(.691; 6)	(.384; 1)
	.20	1.064	1.367	1.874	3.586
		(.800; 4,806)	(.795; 169) [14]	(.778; 34)	(.699; 5)

Table 3. For the $\pi^N \times \psi$ combinations of Table 1(b), (a) provides N_ψ (upper) and $(\gamma_\psi;\ \vec{\pi}_C)$ (lower) and (b) provides $\delta(\vec{\pi}_C)$ (upper) and 'power of the analysis based on $\theta = \delta$, given a design based on $\theta = \log \psi$'; $E[y_C]$ (lower).

	π^N	δ			
(a)		.01	.05	.10	.20
	.01	3,265	600	—	—
		(.50; .0064)	(.27; .0019)		
	.05	14,909	632	206	115
		(.88; .0453)	(.53; .0315)	(.30; .0226)	(.25; .0118)
	.10	28,283	1,141	299	94
		(.92; .0952)	(.74; .0784)	(.54; .0627)	(.33; .0429)
	.20	50,229	2,019	506	129
		(.96; .1951)	(.87; .1767)	(.77; .1559)	(.59; .1217)

(b)		.01	.05	.10	.20
	.01	.0109	.0385	**.0426**	**.0371**
		(.752; 14)	(.653; 1)	(.808; 0)	(.973; 0)
	.05	.0100	.0536	.1195	.1893
		(.798; 359)	(.761; 13)	(.683; 4)	(.687; 1)
	.10	.0100	.0509	.1063	.2299
		(.800; 1,401)	(.792; 51)	(.771; 12)	(.727; 3)
	.20	.0100	.0501	.1012	.2102
		(.800; 5,000)	(.799; 1915)	(.797; 45)	(.792; 10)

Summary of designs for noninferiority trials with $\pi_E = \pi_C$ under H_A

Q: What are the fixed design parameters that must be specified?

- The contrast parameter under H_0 (i.e., the noninferiority margin).
- The common response rate, π^N.

Q: What are their values?

- The contrast parameter: should reflect subject matter knowledge and not be too large (can check on another scale, e.g., δ vs. ψ).
  - ensure that π_E still reflects efficacy (with respect to π_P)!!
  - avoid π_C too close to boundary

- The common response rate: If unknown then a suitable value can be found starting from the control-group response rate, π_C:
  - Can be found via different approaches; the *Tailored Approach* is recommended.
  - The approach used affects N and should be reported.

- For response reflecting negative outcome, $\delta \leq \pi^N$ seems to be a helpful rule of thumb.

Q: What is the optimal value of the allocation ratio – i.e., the value that minimizes N?

- As $\pi^N \rightarrow 0.5$, $\gamma \rightarrow 1.0$.

- As $\pi^N \rightarrow 0$ (or 1) and/or δ_0 (or ψ_0) increases,
  - γ_δ increases
  - γ_ψ decreases

  Extreme values of γ have very low power because $E[y_C] \rightarrow 0$ (or 1).

- The overall sample size, N, can be substantially smaller at the optimal value of γ than at an arbitrary value.

- The optimal γ and corresponding N must be found iteratively via an algorithm: http://www.epibiostat.ucsf.edu/biostat/joan/