A Hybrid Design for Non-inferiority Trials with Binary Outcomes

George YH Chi*, Janssen R&D

Abstract

For non-inferiority (NI) trials with binary outcomes, the current practice is to determine a fixed margin δ_o at an assumed control response rate $p_{C,o}$. The pair $(\delta_o, p_{C,o})$ is usually derived from historical data and expert clinical knowledge, an approach that was recommended in the 2010 FDA draft NI guidance. FDA posed the following question to the November 2011 FDA Anti-infective Advisory Committee. Should the same margin be used in the event the true control response rate $p_C \neq p_{C,o}$? If not, then what margin should one use and what strategy can one propose to prospectively address this problem at the design stage?

A hybrid design for the rate difference measure is considered here for addressing the above problem. The NI hypothesis of the hybrid design is defined by a special kind of linear margin. This special linear margin is derived within the framework of the theory of inferiority index for Bernoulli distributions [Li ans Chi (2011) and Chi and Koch (2013)] as follows. From the given pair $(\delta_o, p_{C,o})$, one can determine its degree of stringency through the index function $\rho_{S,o} = g^*_{RD}(\delta_o, p_{C,o})$. Then, by setting the index $\rho_S = \rho_{S,o}$ in the margin function g^{*-1}_{RD}, one derives the specific margin function $\delta_{RD}(p_C|\rho_{S,o}) = g^{*-1}_{RD}(\rho_{S,o}, p_C)$. It is clear that $\delta_{RD}(p_{C,o}|\rho_{S,o}) = \delta_o$. This linear margin integrates the given fixed margin δ_o at the assumed control response rate $p_{C,o}$ with a variable component. This integration is accomplished by finding the line tangent to the above specific margin function $\delta_{RD}(p_C|\rho_{S,o}) = g^{*-1}_{RD}(\rho_{S,o}, p_C)$ at the assumed control response rate $p_{C,o}$. This tangent line defines the special linear margin $\mathcal{L}(p_C|\rho_{S,o}, p_{C,o}) = \delta_{RD}(p_{C,o}|\rho_{S,o}) + \frac{\partial g^{*-1}_{RD}(\rho_{S,o}, p_{C,o})}{\partial p_C}(p_C - p_{C,o})$, which integrates the given fixed margin $\delta_{RD}(p_{C,o}|\rho_{S,o}) = \delta_o$ with the variable linear term $\frac{\partial g^{*-1}_{RD}(\rho_{S,o}, p_{C,o})}{\partial p_C}(p_C - p_{C,o})$. This is the reason why this specific linear margin is called a hybrid margin. Unlike the case of a general variable margin [Zhang (2006)], the focus of this linear margin is still on δ_o at $p_{C,o}$, but it also retains some of the flexibility of a variable margin. Through a study of the performance characteristics of the test statistic associated with the hybrid NI hypothesis, management of its type I error rate and power is possible. The design is illustrated with an application to a hospital acquired/ventilation associated bacterial pneumonia (HABP/VABP) trial. This hybrid design can potentially address the above questions posed by the FDA.

Key words: binary outcomes, fixed margin, hybrid design, hybrid margin, linear margin, margin function, non-inferiority, variable margin.

1. Background and Statement of the Problem

The 1992 FDA Anti-Infective Points-to-Consider Guideline defines the following function $\delta(p_C)$ as the margin function for the rate difference measure δ_{RD}: $\delta(p_C) = -0.20$, if $p_C \leq 0.80$, $= -0.15$, if $0.80 < p_C \leq 0.90$, and $= -0.10$, if $p_C > 0.90$. As discussed by Bristol (1994) and Weng and Liu (1994), there

*Corresponding author: gchi@its.jnj.com

are two problems with this Guideline - the discontinuous nature of this margin function and its retrospective implementation. To address these two issues, Lewis [Röhmel (1998, 2001)] suggested that a margin function should be continuous and the test statistic associated with the non-inferiority (NI) hypothesis defined by the margin function should have relatively constant power over a wide range of the control response rate p_C. The latter property apparently is thought to be required in order to avoid the problem discussed in Weng and Liu (1994). However, it is clear from the discussion in Chi and Koch (2013) that such constant power is unlikely to extend beyond the range (0.30, 0.70). Of course, the reason for the guideline's retrospective implementation is because one does not know for sure at the design stage what the expected control response rate would be from the NI trial.

In order to search for an appropriate continuous function with the desired properties, Tu (1998), Senn (2000) and Garrett (2003) proposed continuous margin functions that are derived from the following functional relationship $\delta_{RD} = h(\delta_{OR}, p_C) = (1 - \delta_{OR})p_C/(1 + \delta_{OR}p_C/(1 - p_C))$ between the rate difference measure δ_{RD} and the odds ratio measure $\delta_{OR} = \frac{p_T(1-p_C)}{p_C(1-p_T)}$ by assigning the respective values of 0.43, 0.55 and 0.50 for the odds ratio measure [Chi and Koch (2013)]. Röhmel on the other hand defined three different continuous margin functions involving the control response rate p_C and its variance σ_C^2 [Röhmel (1998, 2001)]. Phillips (2003) proposed a linear margin function that is consistent with the FDA step function at its discontinuity points, while Kim and Xu (2004) considered a piecewise linear margin function which is derived by piecing together separate segments that are constructed based on clinical judgments. A common feature of these margin functions is that they are designed to be consistent with the 1992 FDA guideline at the discontinuity points of its step function. Munk, Skipka and Stratmann (2005) and Zhang (2006) further derived the asymptotic test statistic for a variable margin NI hypothesis where the variable margin is defined by a general margin function that satisfies certain regularity conditions. They applied their results to the various margin functions discussed above.

However, the problem is not to find the most general margin function, but rather in finding a natural family of margin functions that have the necessary properties. This family of margin functions is derived in Chi and Koch (2013).

The U.S. Food and Drug Administration issued a draft guidance to the industry on NI [US FDA (2010)] trials recommending a fixed margin approach. Relative to the rate difference measure δ_{RD}, this approach essentially involves two steps. The first step is to obtain an estimate of the control response rate and a conservative estimate of the effect of the control based on available relevant historical studies involving the control and placebo called M_1. The second step then involves the specification of a retention fraction λ and the NI margin is then set equal to $M_2 = -(1-\lambda)M_1$, which represents the maximum amount of loss of the control effect that is judged to be clinically acceptable at the estimated control response rate. Should the trial succeed in excluding this margin, then one would surmise that the new treatment would have shown superiority to a placebo had a placebo been present – provided the assumption on the assay sensitivity of the trial and the assumption on the constancy of the historical control effect both hold true. For a NI trial with binary outcome, the fixed margin is also determined in association with an assumed control response rate for p_C, which is also estimated from the historical data. This approach reflects FDA's perspective and emphasis on a margin that is based empirically on the best available relevant historical data and expert clinical judgment. Although obvious, it is important and of present interest to point out that such empirically based fixed margins are not derived through any margin function.

This new guidance effectively recommends a fixed margin approach to address the issues inherent in the 1992 anti-infective guideline. This fixed margin approach was recently used by the Agency in the November 2011 Anti-infective Advisory Committee discussion of the design of HABP and VABP trials [US FDA (20110]. In the briefing book provided to the Committee members, FDA presented the

following information based on two historical placebo-controlled studies and five recent active control studies.

	Mortality Rate	95% CI
"Placebo"	62%	(52%, 71%)
Control	20%	(18%, 23%)

Following the fixed margin approach recommended in the 2010 FDA Draft NI Guidance, an estimate of the control effect is given by *CE* = [(0.52 – 0.23) – 0.09] = 0.20, which was obtained by taking the difference between a conservative estimate of the mortality rate under placebo (52%) and a conservative estimate of the mortality rate under control (23%) and then subtract 9% to account for factors that may impact on the underlying assumptions of constancy and assay sensitivity. The proposed NI margin for δ_{RD} is set at $\delta_o = -CE \times \frac{1}{2} = -0.20 \times \frac{1}{2} = -0.10$, where the fraction of one-half was applied based on clinical judgment regarding the appropriate size of the margin.

One of the questions FDA posed to the Anti-infective Advisory Committee is the following:

Question: *Discuss whether the fixed margin of* $\delta_o = -0.10$ *derived above would still be appropriate in the event the expected control response rate* p_C *from the trial actually deviates from the assumed rate of* $p_{C,o} = 0.80$. *If not appropriate, then what margin should one use instead? What prospective strategy can one use to address this question at the design stage?*

This question points to the same problem that was inherent in the 1992 Anti-infective guideline. The Committee did not have an answer to this question. The hybrid NI design to be proposed in this paper is an attempt to address this question.

This paper acknowledges the regulatory agency's preference for relying on empirical data and expert clinical judgment in the determination of a fixed NI margin. However, it also recognizes the need for an objective assessment of the stringency of this margin or any other margin however it is derived. In addition, it also realizes the fact that the actual control response rate from the NI trial may very well differ from what one may have expected it to be at the design stage. Hence the fixed margin that is derived based on an assumed control response rate at the design stage may need to be changed depending upon the actual control response rate. If the actual control response rate deviates somewhat from the assumed control response rate, then the margin should also be expected to be different. This paper proposes a hybrid design to prospectively address these issues. This hybrid design will be defined by explicitly integrating the fixed margin derived using the approach recommended in the 2010 FDA draft NI guidance and the variable margin approach as discussed in Zhang (2006). This integration is accomplished through the theory of inferiority index for Bernoulli distributions as discussed in Chi and Koch (2013) by an interactive application of the pair of index and margin functions (g^*_{RD}, g^{*-1}_{RD}) as defined in that paper. The notations used there will be adopted throughout this paper. This hybrid NI design has the following nice features: (i) it retains the given fixed margin and control response rate that were derived from historical data and expert clinical judgment according to the FDA recommended approach, and (ii) it possesses the flexibility to allow the margin to change in the event the observed control response rate from the trial deviates somewhat from the assumed control response rate. However, this flexibility comes at the cost of an increase in sample size.

2. Fixed Margin and Margin Function for Bernoulli Distributions

For the setting of this paper, consider a randomized parallel non-inferiority (NI) trial with a treatment T and an active control C. Let X represent a binary outcome of interest, X_T and X_C are the outcome under treatment T and active control C. Let $X_T \sim Bernoulli(p_T)$ and $X_C \sim Bernoulli(p_C)$, and T is worse than C if $p_T < p_C$. For reasons as discussed in Chi and Koch (2013) in this Proceeding, the focus in this paper will be on the rate difference measure $\delta_{RD} = p_T - p_C$. As shown in Chi and Koch (2013), there is a natural function g^*_{RD} that asymptotically links the standard inferiority index ρ_S under the normal distributions to the rate difference measure δ_{RD} and the control response rate p_C given by

$$\rho_S = g^*_{RD}(\delta_{RD}, p_C). \tag{1}$$

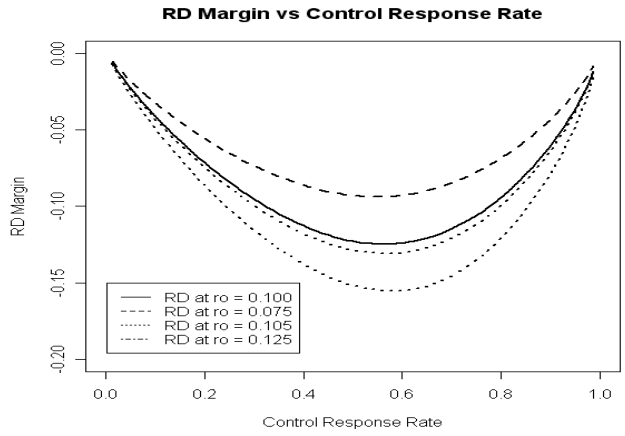
The standard inferiority index ρ_S is a probability and can be viewed as providing a measure of the degree of stringency for a given margin δ_{RD} at a given control response rate p_C through the function g^*_{RD}. The link function g^*_{RD} will be called an *index function*. Now as shown in Chi and Koch (2013), associated with the index function g^*_{RD}, there is an inverse function g^{*-1}_{RD} that assigns a margin $\delta_{RD} = g^{*-1}_{RD}(\rho_S, p_C)$ at a given control response rate p_C with a desired degree of stringency ρ_S. Upon setting the standard inferiority index ρ_S at a desired threshold $\rho_{S,o}$ in the function $\delta_{RD} = g^{*-1}_{RD}(\rho_S, p_C)$, one obtains a specific margin function

$$\delta_{RD}(p_C|\,\rho_{S,o}) = g^{*-1}_{RD}(\rho_{S,o}, p_C) \tag{2}$$

with the stringency given by $\rho_{S,o}$. The specific margin function $\delta_{RD}(p_C|\rho_{S,o}) = g^{*-1}_{RD}(\rho_{S,o}, p_C)$ can be viewed as that level curve on the surface of the index function $\rho_S = g^*_{RD}(\delta_{RD}, p_C)$ realized upon setting $\rho_S = \rho_{S,o}$. Figure 1 plots the margin functions defined by (2) for a few specific stringency levels.

FIGURE 1

A Family of Margin Functions for δ_{RD}



Zhang (2006) has considered variable margins defined by general margin functions. In particular, using the margin function (2), one can define the following variable margin NI hypothesis:

$$H_o\colon \delta_{RD} \le \delta_{RD}(p_C|\,\rho_{S,o}) = g^{*-1}_{RD}(\rho_{S,o}, p_C) \quad \text{vs.} \quad H_o\colon \delta_{RD} > \delta_{RD}(p_C|\,\rho_{S,o}) = g^{*-1}_{RD}(\rho_{S,o}, p_C). \tag{3}$$

The focus of this paper is not on the variable margin NI hypothesis as defined in (3), but on a hybrid margin which explicitly integrates a given fixed margin as derived from the FDA recommended approach into a linear margin derived from the margin function (2) as discussed below.

Clearly, the margin function (2) defines fixed margins at all control response rate p_C with the same degree of stringency given by $\rho_{S,o}$. Thus, at a given control response rate $p_C = p_{C,o}$, the margin function (2) defines a fixed margin $\delta_{RD,o} = \delta_{RD,o}(p_{C,o}|\rho_{S,o}) = g_{RD}^{*-1}(\rho_{S,o}, p_{C,o})$ with the degree of stringency $\rho_{S,o}$. One can then define a fixed margin NI hypothesis based on this fixed margin as follows:

$$H_o: \delta_{RD} \leq \delta_{RD,o} \quad \text{vs.} \quad H_o: \delta_{RD} > \delta_{RD,o}. \tag{4}$$

Now it is important to note that in practice as recommended by the FDA draft NI guidance, the control response rate and the fixed NI margin are usually derived from relevant historical data and expert clinical knowledge. For instance, suppose that such a pair of fixed margin and control response rate $(\delta_o, p_{C,o})$ has been derived. Then, the index function g_{RD}^* given in (1) would yield a degree of stringency $\rho_{S,o} = g_{RD}^*(\delta_o, p_{C,o})$. Now with this degree of stringency $\rho_{S,o}$, the inverse link function g_{RD}^{*-1} given in (2) defines a specific margin function $\delta_{RD}(p_C|\rho_{S,o}) = g_{RD}^{*-1}(\rho_{S,o}, p_C)$ so that $\delta_o = \delta_{RD}(p_{C,o}|\rho_{S,o}) = g_{RD}^{*-1}(\rho_{S,o}, p_{C,o})$. It is through such interactive application of the pair of index and margin functions (g_{RD}^*, g_{RD}^{*-1}) that the potential usefulness of the inferiority index can be further realized as to be illustrated in the hybrid NI design to be proposed below.

In Section 3, a hybrid margin and its corresponding hybrid NI hypothesis will be defined. In Section 4, the performance of the test statistic associated with the hybrid NI hypothesis will be discussed. An application to the design of HABP/VABP trials will be illustrated in Section 5. The paper concludes with a summary discussion.

3. A Hybrid Design for NI Trials with Binary Outcomes

Suppose that one is contemplating the design of a parallel randomized double-blind NI trial comparing an experimental treatment T to an active control C with binary outcome using the rate difference δ_{RD} as the effect measure. Assume that a fixed margin δ_o has been determined at an assumed control response rate of $p_{C,o}$. Such a pair $(\delta_o, p_{C,o})$ could have been derived using the fixed margin approach recommended by the FDA as discussed earlier which involves using historical data and expert clinical judgment. Such δ_o has typically been derived without reference to any margin function.

Then, normally one would define the corresponding fixed margin NI hypothesis as

$$H_{RD,o}: \delta_{RD} \leq \delta_o \quad \text{vs.} \quad H_{RD,a}: \delta_{RD} > \delta_o. \tag{5}$$

However, for obvious reason, this fixed margin NI hypothesis will not provide the proper framework for addressing the FDA question described in the beginning. An alternative strategy is proposed below.

From the index function (1), the standard inferiority index value for the pair $(\delta_o, p_{C,o})$ is given by

$$\rho_{S,o} = g_{RD}(\delta_o, p_{C,o}). \tag{6}$$

This inferiority index value $\rho_{S,o}$ represents a degree of stringency that reflects the amalgamation of the information based on the historical data and the expert clinical judgment.

Now by specifying this same index value $\rho_{S,o}$ from (6) in the margin function (2), one obtains the following specific continuously differentiable margin function

$$\delta_{RD}(p_C|\rho_{S,o}) = g_{RD}^{*-1}(\rho_{S,o}, p_C) \tag{7}$$

with the stringency $\rho_{S,o}$. This margin function defines margin with the same stringency $\rho_{S,o}$ at all p_C's. In particular, this function when evaluated at the assumed control response rate $p_{C,o}$ would yield the fixed margin δ_o as expected, i.e.,

$$\delta_{RD}(p_{C,o}|\rho_{S,o}) = g_{RD}^{*-1}(\rho_{S,o}, p_{C,o}) = \delta_o. \tag{8}$$

It is of interest to point out that the pair $(\delta_o, p_{C,o})$ was originally derived from empirical data and expert clinical knowledge and not from a margin function, but now through an interactive application of the index function (1) and the margin function (2), a specific margin function with the empirically determined degree of stringency $\rho_{S,o}$ from (6) can be defined by setting the inferiority index $\rho_S = \rho_{S,o}$ in the function $\delta_{RD} = g_{RD}^{*-1}(\rho_S, p_C)$ as shown in (7). This specific margin function when evaluated at the assumed control response rate $p_{C,o}$ yields the given fixed margin δ_o. It should be emphasized that the degree of stringency $\rho_{S,o}$ reflects the information gathered based on empirical data and expert clinical knowledge through the pair $(\delta_o, p_{C,o})$.

It follows from the implicit function theorem that the derivative of the margin function (7) exists and is given by

$$\frac{\partial \delta_{RD}(p_C|\rho_{S,o})}{\partial p_C} = \frac{\partial g_{RD}^{*-1}(\rho_{S,o}, p_C)}{\partial p_C} = -\frac{\frac{\partial g_{RD}^*}{\partial p_C}}{\frac{\partial g_{RD}^*}{\partial \delta_{RD}}}. \tag{9}$$

Now consider the first order Taylor expansion of the margin function (7) around the point $p_{C,o}$ as illustrated by Figure 2.

$$\begin{aligned}\mathcal{L}(p_C|\rho_{S,o}, p_{C,o}) &= g_{RD}^{*-1}(\rho_{S,o}, p_C)|_{p_C=p_{C,o}} + \frac{\partial g_{RD}^{*-1}(\rho_{S,o}, p_C)}{\partial p_C}|_{p_C=p_{C,o}}(p_C - p_{C,o}) \\ &= \delta_{RD}(p_{C,o}|\rho_{S,o}) + \frac{\partial g_{RD}^{*-1}(\rho_{S,o}, p_C)}{\partial p_C}(p_C - p_{C,o}) = \delta_o + \frac{\partial g_{RD}^{*-1}(\rho_{S,o}, p_{C,o})}{\partial p_C}(p_C - p_{C,o})\end{aligned} \tag{10}$$

The expression in the approximation $\mathcal{L}(p_C|\rho_{S,o}, p_{C,o})$ in (10) is equal to the fixed margin δ_o plus the linear term $\frac{\partial g_{RD}^{*-1}(\rho_{S,o}, p_{C,o})}{\partial p_C}(p_C - p_{C,o})$. If the true control response rate p_C from the trial turns out to be equal to $p_{C,o}$, then $\mathcal{L}(p_{C,o}|\rho_{S,o}, p_{C,o}) = \delta_o$. But when p_C deviates from $p_{C,o}$, then the margin is equal to the given fixed margin $\mathcal{L}(p_C|\rho_{S,o}, p_{C,o}) = \delta_o$ plus the perturbation, $\frac{\partial g_{RD}^{*-1}(\rho_{S,o}, p_{C,o})}{\partial p_C}(p_C - p_{C,o})$, which represents an adjustment to the margin δ_o due to the deviation. Hence, the linear function defines a hybrid margin $\mathcal{L}(p_C|\rho_{S,o}, p_{C,o})$, so called because it integrates the given fixed margin δ_o at $p_{C,o}$ with a variable term $\frac{\partial g_{RD}^{*-1}(\rho_{S,o}, p_{C,o})}{\partial p_C}(p_C - p_{C,o})$.

Now a natural question to ask is how stringent is the linear margin $\mathcal{L}(p_C|\rho_{S,o}, p_{C,o})$? The specific margin function $\delta_{RD}(p_C|\rho_{S,o})$ has the stringency $\rho_{S,o}$, so the linear margin function $\mathcal{L}(p_C|\rho_{S,o}, p_{C,o})$ cannot be at this stringency level except at $p_C = p_{C,o}$. But the important point to note is that this linear margin has approximately the same degree of stringency $\rho_{S,o}$ as the margin function $\delta_{RD}(p_C|\rho_{S,o})$ in a certain

interval around $p_{C,o}$. For the HABP/VABP example discussed above, with $(\delta_o, p_{C,o}) = (-0.10, 0.80)$, this interval is approximately (0.75, 0.90) as shown in Table 1.

FIGURE 2

Linear Margins

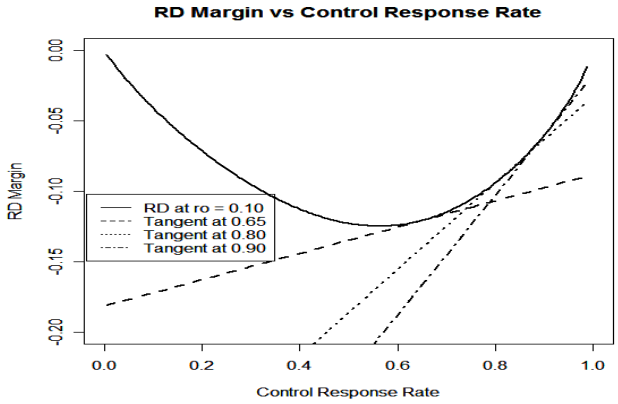


TABLE 1

$$\delta_{RD}(p_C|\rho_{S,o}) \ vs. \ \mathcal{L}(p_C|\rho_{S,o}, p_{C,o})$$

Comparing the Margin Functions $\delta_{RD}(p_C\|\rho_{S,o})$ and $\mathcal{L}(\rho_{S,o}, p_C)$*		
True Control Response Rate	Taylor Expansion around $P_{C,o}$= 0.80	
	Margin Function	
P_C	$\delta_{RD}(p_C\|\rho_{S,o})$	$\mathcal{L}(\rho_{S,o}, p_C)$
0.50	-0.1228	-0.1859
0.55	-0.1246	-0.1706
0.60	-0.1240	-0.1553
0.65	-0.1257	-0.1399
0.70	-0.1146	-0.1246
0.75	-0.1058	-0.1093
0.80	0.0939	0.0939
0.85	-0.0789	-0.0767
0.90	-0.0610	-0.0614
0.95	-0.0360	-0.0460
* Linear Margin derived from Taylor expansion		

Therefore, now one may consider the following hybrid NI hypothesis as approximately equivalent to the NI hypothesis in (3) for certain interval of p_C:

$$H_{RD,o}: \ \delta_{RD} \leq \mathcal{L}(p_C|\rho_{S,o}, p_{C,o}) \quad \text{vs.} \quad H_{RD,a}: \ \delta_{RD} > \mathcal{L}(p_C|\rho_{S,o}, p_{C,o}). \tag{11}$$

The hybrid NI hypothesis (11) can be equivalently written as

$$H_{RD,o}: \ \delta_{RD} - \mathcal{L}(p_C|\rho_{S,o}, p_{C,o}) \leq 0 \ \text{ vs. } \ H_{RD,a}: \ \delta_{RD} - \mathcal{L}(p_C|\rho_{S,o}, p_{C,o}) > 0. \tag{12}$$

Now consider a binary outcome trial and let $\{X_{T,i}\}_{i=1}^{n}$ and $\{X_{C,i}\}_{i=1}^{n}$ be two independent random Bernoulli samples, where $X_{T,i} \sim Bernoulli\ (p_T)$ and $X_{C,j} \sim Bernoulli\ (p_C)$. Let $\hat{p}_T = \sum_{i=1}^{n} X_{T,i}\ /n$ and $\hat{p}_C = \sum_{j=1}^{n} X_{C,j}\ /n$ denote the sample means of X_T and X_C respectively.

Consider the statistic

$$\hat{\Delta}_{RD} = \left[\hat{\delta}_{RD} - \mathcal{L}(\hat{p}_C|\rho_{S,o}, p_{C,o})\right]. \quad (13)$$

Then, $E(\hat{\Delta}_{RD}) = E\left[\hat{\delta}_{RD} - \mathcal{L}(\hat{p}_C|\rho_{S,o}, p_{C,o})\right] = \delta_{RD} - \mathcal{L}(p_C|\rho_{S,o}, p_{C,o})$. Let

$$\hat{\Delta}_{RD,o} = \left[\hat{\Delta}_{RD} - E(\hat{\Delta}_{RD}|H_o)\right]. \quad (14)$$

One has the following theorem.

Theorem 1: The statistic $\sqrt{n}\,\hat{\Delta}_{RD,o}$ is asymptotically normal $N(0, \Sigma(p_C|H_o))$ at the boundary of the inferiority null of (11) or (12), where

$$\Sigma^2(p_C|H_o) = \Sigma^2_{RD,o}(p_{C,o}|H_o) + \left[2\left(\frac{\partial g^{*-1}_{RD}(\rho_{S,o}, p_{C,o})}{\partial p_C}\right) + \left(\frac{\partial g^{*-1}_{RD}(\rho_{S,o}, p_{C,o})}{\partial p_C}\right)^2\right] p_C(1-p_C), \quad (15)$$

and $\Sigma^2_{RD,o}(p_{C,o}|H_o) = \left((1+\sigma_o^2)\left[\sigma^2_{C,o} + \frac{\delta^2_{RD,o}}{16\sigma_C^2\sigma_o^2}\right]\right) + 2(1-2p_{C,o})\left[\frac{\delta_{RD,o}}{2\sigma_{C,o}}\right]^2 (p_C - p_{C,o}) - \left[\frac{\delta_{RD,o}}{\sigma_{C,o}}\right](1-2p_C)\sigma_C$ is the variance of the statistic under the fixed margin NI hypothesis (5) with the fixed margin $\delta_{RD,o} = \delta_{RD}(p_{C,o}|\rho_{S,o})$, $\sigma_o^2 = \frac{\sigma^2_{T,o}}{\sigma^2_{C,o}}$, where $\sigma^2_{C,o} = p_{C,o}(1-p_{C,o})$ and $\sigma^2_{T,o} = p_{T,o}(1-p_{T,o})$ with $p_{T,o} = p_{C,o} + \delta_{RD}(p_{C,o}|\rho_{S,o})$.

Proof: The proof follows from the central limit theorem and a derivation of the asymptotic variance of

$$\begin{aligned}\sqrt{n}\,\hat{\Delta}_{RD,o} &= \sqrt{n}\left[\hat{\delta}_{RD} - \mathcal{L}(\hat{p}_C|\rho_{S,o}, p_{C,o})\right] = \sqrt{n}\left[\hat{\delta}_{RD} - \frac{\partial g^{*-1}_{RD}(\rho_{S,o}, p_{C,o})}{\partial p_C}(\hat{p}_C - p_C)\right] - \mathcal{L}(p_C|\rho_{S,o}, p_{C,o}) \\ &= \sqrt{n}\left[\hat{\delta}_{RD} - \frac{\partial g^{*-1}_{RD}(\rho_{S,o}, p_{C,o})}{\partial p_C}(\hat{p}_C - p_C)\right] - \left[\delta_{RD}(p_{C,o}|\rho_{S,o}) + \frac{\partial g^{*-1}_{RD}(\rho_{S,o}, p_{C,o})}{\partial p_C}(p_C - p_{C,o})\right] \\ &= \sqrt{n}\left[\hat{\delta}_{RD} - \delta_{RD}(p_{C,o}|\rho_{S,o})\right] - \left[\frac{\partial g^{*-1}_{RD}(\rho_{S,o}, p_{C,o})}{\partial p_C}(\hat{p}_C - p_C) - \frac{\partial g^{*-1}_{RD}(\rho_{S,o}, p_{C,o})}{\partial p_C}(p_C - p_{C,o})\right]\end{aligned}$$

by applying Theorem 3 in Chi and Koch (2013) with the fixed margin $\delta_o = \delta_{RD}(p_{C,o}|\rho_{S,o})$ and calculating the cross-product term. ■

Hence, the test statistic,

$$\hat{T}_{RD,o} = \frac{\sqrt{n}\,\hat{\Delta}_{RD,o}}{\sqrt{\Sigma(\hat{p}_C|H_o)}} \sim N(0,1), \quad (16)$$

where p_C is substituted by the sample proportion $\hat{p}_C$. The hybrid inferiority hypothesis in (11) or (12) may be rejected at the $\alpha = 0.025$ significance level if the test statistic $\hat{T}_{RD,o} = \sqrt{n}\,[\hat{\delta}_{RD} - \mathcal{L}(\hat{p}_C|\rho_{S,o}, p_{C,o})]\,/\Sigma(\hat{p}_C|H_o) > 1.96$.

4. Performance of the Test Statistic $\hat{T}_{RD,o}$

It should be pointed out that the focus of the hybrid NI design is still on the margin $\delta_o = \delta_{RD}(p_{C,o}|\rho_{S,o})$ at the assumed control response rate $p_C = p_{C,o}$, even though one has added the flexibility in the event the true control response rate p_C deviates somewhat from $p_{C,o}$. Therefore, it would be of particular interest to investigate the performance of the test $\hat{T}_{RD,o}$ at $p_{C,o}$.

4.1 Simulation of the Type I Error Rate

The type I error rate of $\hat{T}_{RD,o}$ is given by

$$\alpha(p_C) = Pr(\hat{T}_{RD,o} > 1.96|H_o) = 1 - Pr(\hat{T}_{RD,o} < 1.96|H_o) = 1 - \Phi\left(\frac{\sqrt{n}\,\hat{\Delta}_{RD,o}}{\sqrt{\Sigma(p_C|H_o)}}\right). \quad (17)$$

Figure 3 displays the simulated type I error rate as a function of the true control response rate p_C . It shows that at the one-sided nominal significance level of 0.025, the type I error rate will be somewhat inflated when the true control response rate $p_C \leq p_{C,o}$. This should be expected because the true p_C is unknown and is being estimated by $\hat{p}_C$. Furthermore, for $p_C < p_{C,o}$, the margin becomes more liberal, whereas for $p_C > p_{C,o}$ the margin becomes tighter.

In light of the type I error rate inflation when $p_C = p_{C,o}$, one can control this by lowering the nominal significance level. Figure 4 shows that if the nominal significance level is lowered to approximately α = 0.0205, then the simulated overall type I error rate when $p_C = p_{C,o}$ is roughly controlled at 0.025.

FIGURE 3

Simulated Overall Type I Error Rate vs p_C

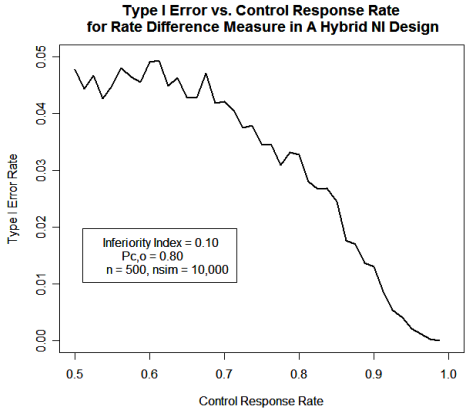


FIGURE 4

Adjusted Simulated Overall Type I Error Rate vs p_C

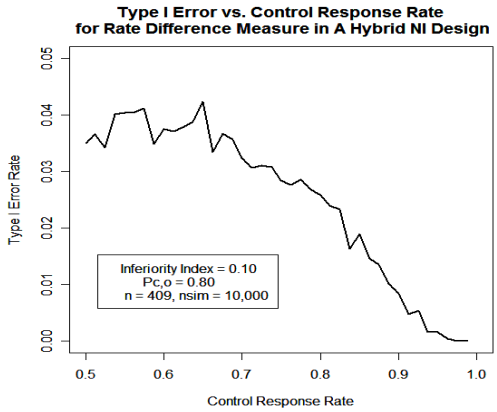


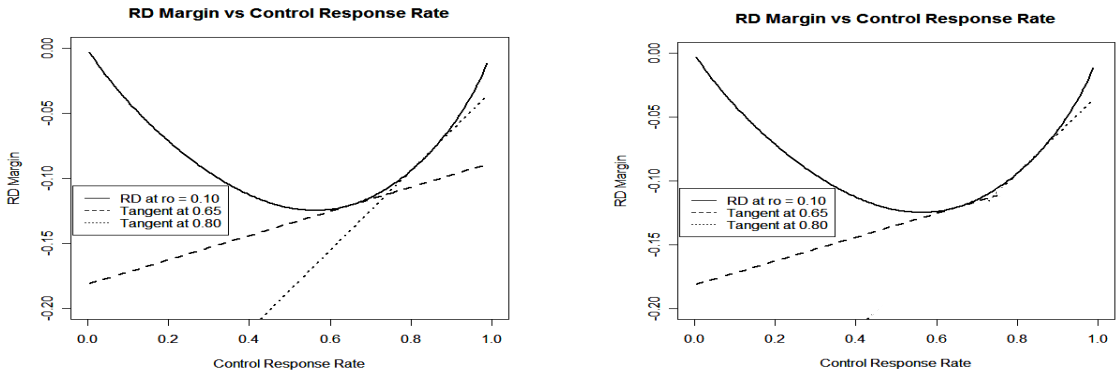
TABLE 2
Adjusted and Unadjusted Type I Error Rates

Unadjusted and Adjusted Type I Error Rate			
True Control Response Rate	Critical Value	Taylor Expansion Point	
		0.80	
P_C	C_a	1.96	2.05*
0.50		0.0477	0.0350
0.55		0.0448	0.0404
0.60		0.0490	0.0375
0.65		0.0429	0.0423
0.70		0.0420	0.0324
0.75		0.0344	0.0284
0.80		0.0327	0.0258
0.85		0.0245	0.0189
0.90		0.0131	0.0084
0.95		0.0022	0.0016
* Adjusted for type I error inflation			

Another way to minimize the type I error inflation for $p_C < p_{C,o}$ is by tightening the margin in this region. This can be accomplished by performing a second Taylor expansion of the margin function $\delta_{RD}(p_C|\rho_{S,o}) = g_{RD}^{*-1}(\rho_{S,o}, p_C)$ at some point $p_{C,1} < p_{C,o}$ and then by joining the two linear margins together to form a piecewise linear margin as shown in Figure 5.

FIGURE 5

Piecewise Linear Margin



4.2 Power Function

To derive the power function for the test statistic $\hat{T}_{RD,o}$, one notes that under the specific alternative H_a: $\delta_{RD}(p_C) \equiv 0$, it follows from (13) that

$$E(\hat{\Delta}_{RD}|H_a) = E[\hat{\delta}_{RD} - \mathcal{L}(\hat{p}_C|\rho_{S,o}, p_{C,o})|H_a] = -\mathcal{L}(p_C|\rho_{S,o}, p_{C,o}) \quad (18)$$

Now let,

$$\hat{\Delta}_{RD,a} = \hat{\Delta}_{RD} - E(\hat{\Delta}_{RD}|H_a) = \hat{\Delta}_{RD} + \mathcal{L}(p_C|\rho_{S,o}, p_{C,o}) = \left[\hat{\delta}_{RD} - \frac{\partial g_{RD}^{*-1}(\rho_{S,o}, p_{C,o})}{\partial p_C}(\hat{p}_C - p_C)\right]. \quad (19)$$

Then, it follows that under the specific alternative H_a: $\delta_{RD}(p_C) \equiv 0$,
$\sqrt{n}\,\hat{\Delta}_{RD,a} \sim N(0, \Sigma(p_C|H_a))$, where the asymptotic variance $\Sigma(p_C|H_a)$ is given by

$$\Sigma(p_C|H_{sa}) = \left[2 + 2\left(\frac{\partial g_{RD}^{*-1}(\rho_{S,o}, p_{C,o})}{\partial p_C}\right) + \left(\frac{\partial g_{RD}^{*-1}(\rho_{S,o}, p_{C,o})}{\partial p_C}\right)^2\right] p_C(1 - p_C) \quad (20)$$

Thus,

$$\hat{T}_{RD,a} = \frac{\sqrt{n}\,\hat{\Delta}_{RD,a}}{\sqrt{\Sigma(p_C|H_a)}} \sim N(0,1), \quad (21)$$

Now, to derive the power formula, one has

$$\hat{T}_{RD,o} = \frac{\sqrt{n}\,\hat{\Delta}_{RD,o}}{\sqrt{\Sigma(p_C|H_o)}} = \frac{\sqrt{n}[\hat{\delta}_{RD} - \mathcal{L}(\hat{p}_C|\rho_{S,o}, p_{C,o})]}{\sqrt{\Sigma(p_C|H_o)}}$$

$$= \frac{\sqrt{n}\left(\left[\hat{\delta}_{RD} - \left(\delta_{RD}(p_{C,o}|\rho_{S,o}) + \frac{\partial g_{RD}^{*-1}(\rho_{S,o}, p_{C,o})}{\partial p_C}(\hat{p}_C - p_{C,o}) \right) \right] \right)}{\sqrt{\Sigma(p_C|H_o)}}, \text{ from (10)}$$

$$= \frac{\sqrt{n}\left(\hat{\Delta}_{RD,a} - \left[\delta_{RD}(p_{C,o}|\rho_{S,o}) + \frac{\partial g_{RD}^{*-1}(\rho_{S,o}, p_{C,o})}{\partial p_C}(p_C - p_{C,o}) \right] \right)}{\sqrt{\Sigma(p_C|H_o)}}, \text{ from (19)}$$

$$= \frac{\sqrt{n}\left(\hat{\Delta}_{RD,a} - \mathcal{L}(p_C|\rho_{S,o}, p_{C,o}) \right)}{\sqrt{\Sigma(p_C|H_o)}}, \text{ from (10)}$$

$$= \hat{T}_{RD,a} \frac{\sqrt{\Sigma(p_C|H_a)}}{\sqrt{\Sigma(p_C|H_o)}} - \frac{\sqrt{n}[\mathcal{L}(p_C|\rho_{S,o}, p_{C,o})]}{\sqrt{\Sigma(p_C|H_o)}}. \quad (22)$$

Therefore, it follows that the power function is given by,

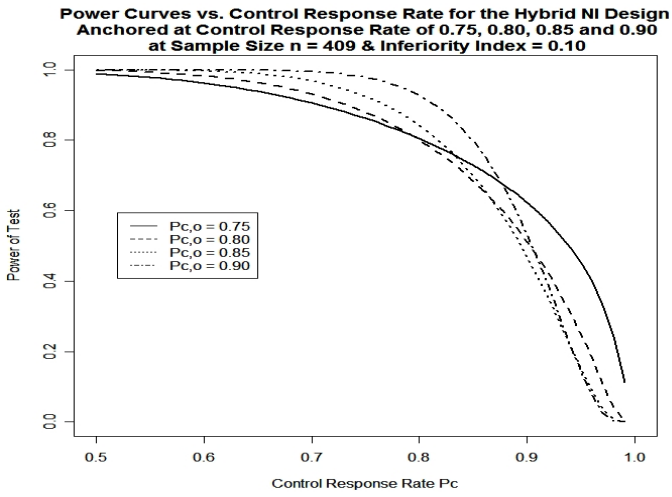
$$1 - \beta = 1 - Pr\left(\hat{T}_{RD,o} < 1.96 \middle| H_a\right) = 1 - Pr\left(\hat{T}_{RD,a} \frac{\sqrt{\Sigma(p_C|H_a)}}{\sqrt{\Sigma(p_C|H_o)}} - \frac{\sqrt{n}\, \mathcal{L}(p_C|\rho_{S,o}, p_{C,o})}{\sqrt{\Sigma(p_C|H_o)}} < 1.96 \right)$$

$$= 1 - Pr\left(\hat{T}_{RD,a} < \frac{1.96\sqrt{\Sigma(p_C|H_o)} + \sqrt{n}\, \mathcal{L}(p_C|\rho_{S,o}, p_{C,o})}{\sqrt{\Sigma(p_C|H_a)}} \right) = 1 - \Phi\left(\frac{1.96\sqrt{\Sigma(p_C|H_o)} + \sqrt{n}\, \mathcal{L}(p_C|\rho_{S,o}, p_{C,o})}{\sqrt{\Sigma(p_C|H_a)}} \right). \quad (23)$$

Now the power function plot in Figure 6 shows that the power drops off quickly when $p_C > p_{C,o}$ due to the dramatic change in variance as $p_C \to 0$ or 1. The deflation in type I error rate for $p_C > p_{C,o}$ might be a desirable feature since it raises a natural barrier on rejecting the inferiority null when the true control response rate p_C is much greater than the assumed control response rate $p_{C,o}$.

FIGURE 6

Power Functions



4.3 Sample Size Calculation

From (21), the sample size formula is derived by setting

$$-z_{1-\beta} = \frac{1.96\sqrt{\Sigma(p_C|H_o)} + \sqrt{n}\,\mathcal{L}(p_C|\rho_{S,o}, p_{C,o})}{\sqrt{\Sigma(p_C|H_a)}}.$$

Solving for *n*, one obtains,

$$n = \frac{\left(1.96\sqrt{\Sigma(p_C|H_o)} + z_{1-\beta}\sqrt{\Sigma(p_C|H_a)}\right)^2}{\mathcal{L}^2(p_C|\rho_{S,o}, p_{C,o})}. \quad (24)$$

FIGURE 7

Sample Size per Group

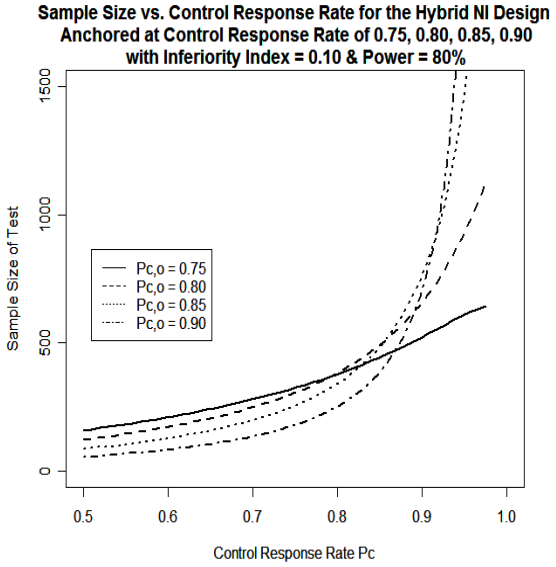


FIGURE 8

Sample Size per Group Adjusted at $\alpha = 0.0205$

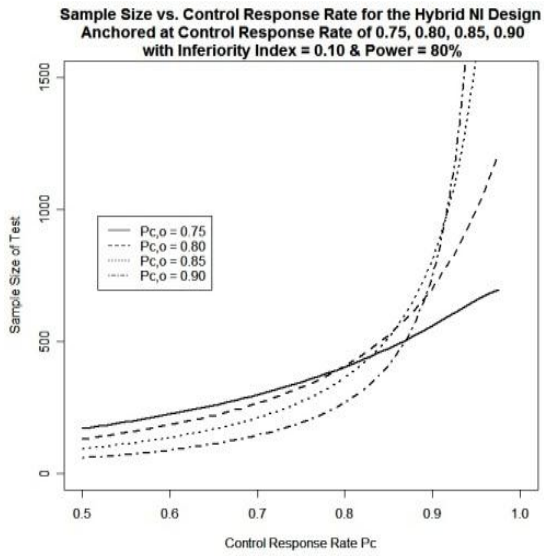


TABLE 3
Adjusted and Unadjusted Sample Size per Group

Sample Size per Group		
True Control Response Rate	Taylor Expansion around $P_{C,o} = 0.80$	
	Critical Value	
P_C	1.96	2.05*
0.50	124	132
0.55	146	155
0.60	174	184
0.65	208	221
0.70	251	266
0.75	307	327
0.80	384	409
0.85	493	526
0.90	659	705
0.95	931	1000
* Adjusted for type I error inflation at $P_{C,o} = 0.80$		

5. Application to HABP/VABP Trials

Now returning to the HABP/VABP example discussed at the beginning of this paper.

Thus, with the given fixed margin of $\delta_o = -0.10$ at an estimated survival rate of 80% (equivalent to a 20% mortality rate), one can find the degree of stringency for $(\delta_o, p_{C,o}) = (-0.10, 0.80)$ to be given through the index function (1) by the index value of $\rho_{S,o} = g^*_{RD}(\delta_{RD,o}, p_{C,o}) = g^*_{RD}(-0.10, 0.80) = 0.1057$. Now, for convenience, consider an index level of $\rho_{S,o} = 0.10$ instead of the actual index level of 0.1057, since type I error simulations, power plots and sample size calculations presented previously used the index level of 0.10. This is equivalent to considering a margin of $\delta_o = -0.0939$ instead of the original margin $\delta_o = -0.10$, at $p_{C,o} = 0.80$. Now upon setting the inferiority index level to $\rho_{S,o} = 0.10$ in the margin function (2), one obtains the special margin function $\delta_{RD}(p_C|0.10) = g^{*-1}_{RD}(0.10, p_C)$ with the degree of stringency specified by $\rho_{S,o} = 0.10$. Applying the Taylor expansion, one finds the linear margin function to be

$$\mathcal{L}(p_C|\rho_{S,o}, p_{C,o}) = \mathcal{L}(p_C|0.10, 0.80) = \delta_{RD}(p_C|0.10) + \frac{\partial g^{*-1}_{RD}}{\partial p_C}(0.10, 0.80)(p_C - 0.80)$$

$$= -0.0939 + 0.3066\,(p_C - 0.80).$$

The hybrid NI hypothesis is then defined by

$$\begin{aligned} &H_o\colon\ \delta_{RD} - [-0.0939 + 0.3066\,(p_C - 0.80)] \le 0 \\ &\qquad\qquad vs. \\ &H_a\colon\ \delta_{RD} - [-0.0939 + 0.3066\,(p_C - 0.80)] > 0. \end{aligned} \tag{25}$$

Based on the hybrid design that has just been discussed in the preceding sections, to test the hybrid NI hypothesis at the significance level of $\alpha = 0.025$ at $p_C = 0.80$ with a power of 80%, an adjusted sample size of $n = 409$ subjects per group would be needed (see Table 3). Now the sample size per group needed for a HABP/VABP trial is given by 682/584 reflecting an adjustment for a 60%/70% microbilogic evaluability rate, or for a total sample size N of 1364/1168 which is greater than the corresponding fixed margin NI hypothesis with a total sample size of 944/809 for HABP/VABP trials [see Chi and Koch (2013)] reflecting a 44.5%/44.4% increase.

Thus, one can see that the flexibility realized in a hybrid NI design is gained at the cost of about a 45% increase in the sample size that was originally needed for a fixed margin.

Now the hybrid design with its hybrid NI hypothesis given by (11) has a linear variable margin $\mathcal{L}(p_C|\rho_{S,o}, p_{C,o})$ that allows the true control response rate p_C to deviate somewhat from the assumed control response rate of 0.80 at the design stage. If the true control response rate $p_C > 0.80$, then from the type I error simulations, one knows that the probability of rejecting the null of (11) is low and very low when $p_C > 0.90$. However, with the given sample size, the test still has about 60% power in rejecting the margin given by $\mathcal{L}(0.90|0.10, 0.80) = -0.0614$ at $p_C = 0.90$, which is very comparable to the margin $\delta_{RD}(0.90|\rho_{S,o}) = g^{-1}_{RD}(0.10, 0.90) = -0.0610$ based on the margin function (2) as shown in Table 1. The power of the test also decreases rapidly as p_C moves away from 0.80 towards 1. However, if the true $p_C < 0.80$, then there is still some inflation in the type I error rate despite the adjustment. Without adjustment by lowering the nominal significance level further from $\alpha = 0.0205$, one may consider constructing a piecewise linear margin by joining another linear margin $\mathcal{L}(p_C|\rho_{S,o}, p_{C,1})$ derived

from first order Taylor expansion of the same margin function $\delta_{RD}(p_C|\rho_{S,o})$ at another point $p_{C,1}$, where $0.50 < p_{C,1} < p_{C,o}$, with the original linear margin $\mathcal{L}(p_C|\rho_{S,o}, p_{C,o})$.

6. Summary Discussion

A hybrid NI hypothesis is proposed in this paper to address a question posed to the November 2011 Anti-infective Advisory Committee by the FDA. The Agency raised the question as to what one should do about the derived fixed margin δ_o when the expected control response rate p_C turns out to be different from the assumed control response rate $p_{C,o}$. Should one use the same margin or a different margin? If one is to use a different margin, then what should that margin be? Is there a prospective strategy that one can use to address this problem?

The hybrid NI hypothesis proposed in this paper is defined by a special linear margin, $\mathcal{L}(p_C|\rho_{S,o}, p_{C,o})$, which is a linear margin derived by applying the Taylor expansion to a specific margin function at the assumed control response rate $p_{C,o}$, where the specific margin function is defined by setting the index $\rho_S = \rho_{S,o}$ in the margin function (2), where the index value $\rho_{S,o}$ is derived from the index function evaluated at the given pair $(\delta_o, p_{C,o})$ by $\rho_{S,o} = g^*_{RD}(\delta_o, p_{C,o})$. Specifically, the linear margin is given by $\mathcal{L}(p_C|\rho_{S,o}, p_{C,o}) = \delta_{RD}(p_{C,o}|\rho_{S,o}) + \frac{\partial g^{*-1}_{RD}(\rho_{S,o}, p_C)}{\partial p_C}(p_C - p_{C,o})$, where $\delta_{RD}(p_{C,o}|\rho_{S,o}) = \delta_o$ is the given fixed margin at $p_{C,o}$. Thus, if the true control response rate $p_C = p_{C,o}$, then the hybrid margin reduces to the given fixed margin, but if the true control response rate $p_C \neq p_{C,o}$, then the hybrid margin adjusts the given fixed margin $\delta_{RD}(p_{C,o}|\rho_{S,o}) = \delta_o$ by the quantity $\frac{\partial g^{*-1}_{RD}(\rho_{S,o}, p_C)}{\partial p_C}(p_C - p_{C,o})$. It is shown that for the example with $(\delta_o, p_{C,o}) = (-0.10, 0.80)$, the linear margin closely approximates the margin function $\delta_{RD}(p_C|\rho_{S,o})$ for p_C over a fairly wide range (0.75, 0.90).

However, the type I error rate simulation shows that there is some inflation at $p_C = p_{C,o} = 0.80$. This inflation can be controlled by adjusting the significance level to $\alpha \approx 0.0205$. However, this adjustment results in an increase of about 45% in sample size per group. Thus, the flexibility in the hybrid NI hypothesis comes at the expense of a substantial increase in sample size. The simulations show a deflationary trend in the type I error rate for $p_C > p_{C,o}$. This may be considered as a desirable property from a regulatory perspective. The type I error rate inflation observed for $p_C < p_{C,o}$ even after this adjustment may be minimized by considering applying the Taylor expansion of the same margin function $\delta_{RD}(p_C|\rho_{S,o})$ around a second point at $p_{C,1} < p_{C,o}$. One then constructs a piecewise linear margin function from the two linear margins by joining them at their point of intersection. The test statistic for a piecewise linear margin and its asymptotic convergence has been studied by Zhang (2006). It may be of interest to consider such investigation for the special margin function $\delta_{RD}(p_C|\rho_{S,o})$ discussed above.

Acknowledgment: Appreciation is expressed to Kim DeWoody for her continued interest and support of this research.

References

Bristol DR (1996). Determining equivalence and the impact of sample size in anti-infective studies: A point to consider. *Journal of Biopharmaceutical Statistics* 6 (3): 319-326.

Chi, GYH and Koch, GG (2013). Inferiority index and margin function for non-inferiority trials with binary outcomes. 2013 *Proceedings of the American Statistical Association, Biopharmaceutical Section*.

Garrett, A.D. (2003), "Therapeutic equivalence: fallacies and falsification". *Statistics in Medicine* 22: 741-762.

Kim MY and Xue X (2004). Likelihood ratio and a Bayesian approach were superior to standard non-inferiority analysis when the non-inferiority margin varied with the control event rate. *Journal of Clinical Epidemiology* 57: 1253-1261.

Li, G and Chi, GYH (2011). Inferiority index and margin in non-inferiority trials. *Statistics in Biopharmaceutical Research* 3 (2): 288-301.

Munk, A, Skipka, B and Stratmann, B (2005). Testing general hypotheses under binomial sampling: the two sample case – asymptotic theory and exact procedures. *Computational Statistics & Data Analysis* 49: 723-739.

Phillips, KF (2003). A new test of non-inferiority for anti-infective trials. *Statistics in Medicine* 22: 201-212.

Röhmel, J (1998). Therapeutic equivalence investigations: statistical considerations. *Statistics in Medicine* 17: 1703-1714.

Röhmel, J (2001). Statistical considerations of FDA and CPMP rules for the investigation of new anti-bacterial products. *Statistics in Medicine* 20: 2561-2571.

Senn, S. (2000), "Consensus and controversy in pharmaceutical statistics (with discussion)", *Journal of the Royal Statistical Society, Series D* 49: 135-176.

Tu, D. (1998), "On the use of the ratio or the odds ratio of cure rates in therapeutic equivalence clinical trials with binary endpoints". *Journal of Biopharmaceutical Statistics* 8 (2): 263-282.

US Food and Drug Administration:

- Anti-Infective Points to Consider Guidance (1992).
- Guidance to Industry: Non-Inferiority Clinical Trials (2010)
- Anti-Infective Drugs Advisory Committee Meeting Materials, November 3-4, 2011. http://www.fda.gov/AdvisoryCommittees/CommitteesMeetingMaterials/ Drugs/Anti-InfectiveDrugsAdvisoryCommittee/ ucm242307.htm (Last accessed February 7, 2013)

Office of Communication/Division of Drug Information/CDER/FDA, W051, Room 2001,10903 New Hampshire Avenue, Silver Spring, MD, 20993.

Weng, C.S.W. and Liu J.P. (1994). Some pitfalls in sample size estimation for an anti-infective study. 1994 Proceedings of the Pharmaceutical Section of the ASA, 56-60.

Zhang, Z (2006). Non-inferiority testing with a variable margin. *Biometrical Journal* 48 (6): 948-965.