Chapter 13
Inferiority Index, Margin Functions, and Hybrid Designs for Noninferiority Trials with Binary Outcomes

George Y. H. Chi

Abstract In the design of noninferiority (NI) trials with binary outcomes, two basic problems are invariably present. The first problem pertains to the appropriateness of a fixed margin. The two-step fixed margin approach recommended in the Food and Drug Administration (FDA) guidance to industry on NI trials (US FDA, Guidance to industry: non-inferiority clinical trials, 2010) relies on the availability of relevant historical data and expert clinical knowledge and experience to provide the assurance that the derived fixed margin is appropriate. Nonetheless, it still needs an objective measure for assessing its stringency. The FDA approach has its merit in that the fixed margin is determined empirically using the best control response rate and control effect estimates and the best clinical judgment. This feature should be retained in a new design. However, once this fixed margin has been determined, one is faced with the second problem of what appropriate margin to use when the control rate from the NI trial differs from the estimated control response rate. This question was raised by the FDA Anti-infective Division at the November 2011 Anti-infective Advisory Committee meeting. A hybrid design for NI trials with binary outcomes is proposed here that integrates the FDA's fixed margin approach with a variable margin by applying the theory of inferiority index developed for Bernoulli distributions. The inferiority index is an objective measure of the relative stringency of a margin, and it can be used to define a special margin function that retains the empirical nature of the fixed margin but also allows the margin to vary.

13.1 Introduction

In the late 1980s, Food and Drug Administration's (FDA's) Anti-infective Division received submissions that include many active control studies. The Division was wrestling with the difficult issue of how to set the noninferiority (NI) margin. Its efforts resulted in the 1992 FDA Anti-infective Points-to-Consider Guidance (US FDA 1992) which reflects the Division's best thinking at the time. The guidance recognized that the nature of the problem lies in the fact that the margin is depended upon

G. Y. H. Chi (✉)
Janssen R&D, LLC, Raritan, NJ 08869, USA
e-mail: gchi@its.jnj.com

Z. Chen et al. (eds.), *Applied Statistics in Biomedicine and Clinical Trials Design*,
ICSA Book Series in Statistics, DOI 10.1007/978-3-319-12694-4_13

the true control response rate which is generally unknown. Therefore, the guidance provided the following margin function to be used for guiding the selection of the NI margin for the rate difference (RD) measure $\delta_{RD} = p_T - p_C$, where p_T and p_C are the response rates for treatment and control, respectively, in a parallel randomized active control trial. The margin function is actually a step function defined to take the value of -0.20, for $p_C \leq 0.80$, the value of -0.15, for $0.80 < p_C \leq 0.90$, and the value of -0.10, for $p_C > 0.90$. However, since the control response rate p_C is not known ahead of time, it is estimated from the NI trial data. The margin function is then applied retrospectively using the estimate $\hat{p}_C$ obtained from the NI trial. Various authors, including Weng and Liu (1994), Bristol (1996), Röhmel (1998, 2001), and Senn (2000) had discussed problems associated with the discontinuous nature of this function and the retrospective nature of its application. Röhmel and Senn also proposed different continuous margin functions. Munk, Skipka and Stratmann (2005) and Zhang (2006) considered NI hypotheses with variable margins that are defined by general margin functions with some regularity properties. However, the concept of a margin function did not receive its due attention from the regulatory authority, probably due to a lack of justification for the choice of margin function and a lack of an accompanied methodology.

In the FDA Guidance to Industry on NI Trials (US FDA 2010), a two-step fixed margin approach is recommended. In this approach, an estimate of the control response rate p_C and an estimate of the control effect (CE) are obtained first from available and relevant historical data. Then from the knowledge and experience of clinical experts, a fraction of the CE estimate is determined as the fixed margin which represents the amount of CE loss that can be tolerated or deemed clinically irrelevant. This two-step fixed margin approach is empirically based and reflects the best clinical judgment as to the degree of stringency required. This is the current practice for most NI trials. However, this two-step fixed margin approach cannot address the question as to what would be the appropriate margin to use in the event the true control response rate p_C from the current NI trial appears to differ from the empirically based estimate of the control response rate p_C. Indeed, at the November 2011 FDA Anti-Infective Advisory Committee meeting discussing the design of hospital-acquired and ventilation-associated bacterial pneumonia (HABP/VABP) NI trials, the FDA Anti-Infective Division posed the following questions among others to the Committee. First, is the fixed margin derived using the two-step procedure for the HABP/VABP trials appropriate? Second, what margin should one use in the event the control response rate p_C from the NI trial appears to deviate from the empirically based estimate? However, the Committee did not provide an answer to this question.

In this chapter, a hybrid NI design for the RD measure that is defined by a special linear margin is presented to address the above two questions raised by the FDA Anti-Infective Division. In Sect. 13.2, the convergence theorem for the test statistic associated with a general fixed margin NI hypothesis is established for the RD measure. This test statistic is more efficient than the classical Wald test and comparable to the likelihood ratio test because it captures the heterogeneity of variance at the boundary of the inferiority null hypothesis. This convergence theorem is used later to

establish the convergence theorem for the test statistic associated with hybrid NI hypothesis. In Sect. 13.3, it is shown that there is an index function linking the standard inferiority index under the normal distributions to the RD measure and the control response rate. Upon setting the index at a specific value in its inverse function, one derives a margin function with a degree of stringency specified by that index. This margin function also accommodates the potential heterogeneity of variance through the variance ratio. Then, in Sect. 13.4, through an application of the index and margin function in tandem, it is shown how one can integrate a given fixed margin into a linear margin that can be used to define a variable margin NI hypothesis which will be termed a hybrid NI hypothesis. This hybrid design has the explicit degree of stringency as measured by the index function at the empirically determined fixed margin and control response rate. In addition, it can accommodate the adjustment of the margin in the event the control response rate from the NI trial deviates somewhat from the empirically based estimate of the control response rate. This hybrid design therefore can address both questions posed by the FDA Anti-Infective Division discussed earlier and is consistent with the spirit stated in the Investigational New Drug (IND) Application Format and Content (US FDA 2013) which mentions among other things that "a protocol for a phase 2 or 3 investigation should be designed in such a way that, if the sponsor anticipates that some deviation from the study design may become necessary as the investigation progresses, alternatives or contingencies to provide for such deviation are built into the protocols at the outset." The performance of the test statistic associated with the hybrid NI hypothesis is investigated and its results discussed, and an application to the design of HABP/VABP trials is given. The chapter concludes with a discussion.

13.2 The Scaled Relative Difference Measure and the Relative Difference Measure

In this section, the scaled rate difference (SRD) measure for Bernoulli distributions is first introduced and the related convergence theorem for the test statistic associated with its fixed margin NI hypothesis is proved. The corresponding convergence theorem for the test statistic associated with the fixed margin NI hypothesis for the RD is then deduced. The reason the scaled difference measure is important is because it takes into account potential differences in the variance through the variance ratio. This property is then passed to the RD measure through its relationship with the SRD measure. The reason this is important is because under Bernoulli distributions, at the boundary of the inferiority null, the variances are different since the treatment and control have different response rates. Furthermore, the slope of the variance function of Bernoulli distributions changes dramatically outside the range of (0.30, 0.70) as the response rate approaches 0 or 1 (Chi and Koch 2012). This property is then also captured in the hybrid design for the RD measure.

13.2.1 The Scaled Relative Difference Measure

The scaled rate difference (SRD) measure plays an interesting and important role in the development of the hybrid design which will become evident later. Consider an active control trial with a treatment T, a control C and a clinical outcome X of interest. Assume that the smaller the value of X, the worse is the outcome. Let X_T and X_C denote outcomes on subjects treated with T and C, and $F_{X_T}(t)$ and $F_{X_C}(t)$ denote their distributions with means μ_T and μ_C and variances σ_T^2 and σ_C^2 respectively.

Let $\delta_{SRD} = \frac{\mu_T - \mu_C}{\sigma_C}$ denote the SRD measure and $\delta_{SRD,\,o}$ denote a fixed NI margin for δ_{SRD}. The adjective "relative" is used to emphasize the fact that the measure is defined relative to the control C. However, for simplicity, it may henceforth simply be referred to as the scaled difference measure. Then, an NI hypothesis for the scaled difference measure δ_{SRD} defined by the margin $\delta_{SRD,\,o}$ is given by Eq. 13.1.

$$H_{SRD,\,o}\colon \delta_{SRD} \le \delta_{SRD,\,o} \quad \text{vs} \quad .H_{SRD,a}\colon \delta_{SRD} > \delta_{SRD,\,o}. \tag{13.1}$$

Now, if furthermore, the distribution F_{X_T} and F_{X_C} have finite third and fourth central moments denoted, respectively, by $\mu_T^{(3)}$ and $\mu_C^{(3)}$, and $\mu_T^{(4)}$ and $\mu_C^{(4)}$, then Li and Chi (2011) proved Theorem 1.

For simplicity, it suffices for the purpose of this chapter to assume that the variance ratio $\sigma^2 = \frac{\sigma_T^2}{\sigma_C^2}$ is known and $\sigma^2 = \sigma_o^2$ for some fixed number σ_o^2. In the general setting where the variance ratio $\sigma^2 = \frac{\sigma_T^2}{\sigma_C^2}$ is not known, the forthcoming discussion can be similarly developed with an appropriate adjustment to the asymptotic variance given in Eq. 13.2 resulting in a correspondingly larger variance. This general case will be discussed elsewhere.

Theorem 1 Assuming that the variance ratio $\sigma^2 = \frac{\sigma_T^2}{\sigma_C^2}$ is known and $\sigma^2 = \sigma_o^2$ for some fixed number σ_o^2. Then at the boundary of the inferiority null hypothesis in Eq. 13.1, the test statistic defined by $\hat{T}_{SRD} = \sqrt{n}(\hat{\delta}_{SRD} - \delta_{SRD,\,o})$ converges asymptotically to a normal distribution $N(0, \Sigma^2_{SRD,\,o})$, where the asymptotic variance $\Sigma^2_{SRD,\,o}$ is given by Eq. 13.2,

$$\Sigma^2_{SRD,\,o} = \left(1 + \sigma_o^2\right) + \frac{\delta^2_{SRD,\,o}}{16}\left[\frac{\mu_T^{(4)} - \sigma_T^4}{\sigma_T^4} + \frac{\mu_C^{(4)} - \sigma_C^4}{\sigma_C^4}\right] - \frac{\delta_{SRD,\,o}}{2}\left[\frac{\mu_T^{(3)}}{\sigma_C \sigma_T^2} + \frac{\mu_C^{(3)}}{\sigma_C^3}\right]. \tag{13.2}$$

The following two corollaries follow directly from Theorem 1.

Corollary 1 *When $X_T \sim N\left(\mu_T, \sigma_T^2\right)$ and $X_C \sim N\left(\mu_C, \sigma_C^2\right)$ are normally distributed, then assuming that the variance ratio $\sigma^2 = \frac{\sigma_T^2}{\sigma_C^2}$ is known and $\sigma^2 = \sigma_o^2$ for some fixed number σ_o^2 then at the boundary of the inferiority null hypothesis in Eq. 13.1, the test statistic $\hat{T}_{SRD} = \sqrt{n}(\hat{\delta}_{SRD} - \delta_{SRD,\,o})$ converges asymptotically to*

a normal distribution $N(0,\ \Sigma^2_{SRD,\,o})$ *where the asymptotic variance* $\Sigma^2_{SRD,\,o}$ *is given by Eq. 13.3,*

$$\Sigma^2_{SRD,\,o} = \left(1 + \sigma_o^2\right) + \frac{\delta^2_{SRD,\,o}}{4}. \tag{13.3}$$

Proof The proof follows from Theorem 1 by noting that for normal distributions, their third central moments are $\mu_T^{(3)} = 0$ and $\mu_C^{(3)} = 0$ and their fourth central moments are $\mu_T^{(4)} = 3\sigma_T^4$ and $\mu_C^{(4)} = 3\sigma_C^4$ respectively. ■

Corollary 2 *When* $X_T \sim Bernoulli(p_T)$ *and* $X_C \sim Bernoulli(p_C)$ *are Bernoulli distributed, assuming that the variance ratio* $\sigma^2 = \frac{\sigma_T^2}{\sigma_C^2}$ *is known and* $\sigma^2 = \sigma_o^2$ *for some fixed number* σ_o^2 *then at the boundary of the inferiority null hypothesis in Eq. 13.1, the test statistic* $\hat{T}_{SRD} = \sqrt{n}(\hat{\delta}_{SRD} - \delta_{SRD,\,o})$ *converges asymptotically to a normal distribution* $N(0,\ \Sigma^2_{SRD,\,o})$ *where the asymptotic variance* $\Sigma^2_{SRD,\,o}$ *is given by Eq. 13.4,*

$$\Sigma^2_{SRD,\,o} = \left(1 + \sigma_o^2\right)\left[1 + \frac{\delta^2_{SRD,\,o}}{16\sigma_C^2\sigma_o^2}\right] + \frac{\delta^2_{SRD,\,o}}{2}. \tag{13.4}$$

Proof The proof follows from Theorem 1 by noting that for the Bernoulli distributions, their third central moments $\mu_T^{(3)} = p_T(1 - p_T)(1 - 2p_T)$ and $\mu_C^{(3)} = p_C(1 - p_C)(1 - 2p_C)$, and their fourth central moments $\mu_T^{(4)} = p_T(1 - p_T)(1 - 2p_T)^2 + \sigma_T^4$ and $\mu_C^{(4)} = p_C(1 - p_C)(1 - 2p_C)^2 + \sigma_C^4$, respectively. ■

Now for obvious reason, under Bernoulli distributions, the scaled difference measure δ_{SRD} will be called the SRD measure. The importance of the SRD measure under normal distributions or the SRD measure under the Bernoulli distributions is that by definition, it accommodates for potential differences between the variance of the treatment and the variance of the control through their variance ratio. Hence, it is the natural parameter to consider if one cannot assume homogeneity of variance. This property is captured in the asymptotic variance of their test statistics associated with the fixed margin NI hypothesis in Eq. 13.1. In the next section, it will be shown how this property can be transferred to the RD measure under the Bernoulli distributions, which is a measure that is more commonly used in practice. Analogous derivation can be done for all other binary effect measures as discussed in Chi and Koch (2012), but will not be discussed here.

13.2.2 The Rate Difference Measure

The rate difference (RD) measure $\delta_{RD} = \mu_T - \mu_C$ under normal distributions has been discussed in Chi (2012) within the context of the Behrens–Fisher problem under the NI hypothesis. It will be further dealt with elsewhere in the context of design of

bioequivalence study for highly variable drugs. For the purpose of this chapter, the focus is on the RD measure $\delta_{RD} = p_T - p_C$ under the Bernoulli distributions.

Under the Bernoulli distributions, the relationship between the SRD measure δ_{SRD} and the RD measure δ_{RD} is given by Eq. 13.5,

$$\delta_{RD} = f_{RD}(\delta_{SRD}, p_C) = \sigma_C \delta_{SRD} = \sqrt{p_C(1-p_C)}\,\delta_{SRD}. \tag{13.5}$$

Let $\delta_{SRD,\,o}$ be a fixed NI margin for δ_{SRD} associated with a given control response rate $p_{C,o}$ Then, Eq. 13.5 indicates that

$$\delta_{RD,o} = \sqrt{p_{C,o}(1-p_{C,o})}\,\delta_{SRD,\,o} \tag{13.6}$$

is the corresponding NI margin for the RD measure δ_{RD} at the same control response rate $p_{C,o}$.

Let the NI hypothesis for the RD measure δ_{RD} corresponding to the NI hypothesis in Eq. 13.1 for the SRD measure δ_{SRD} be defined by

$$H_{RD,o}\colon \delta_{RD} \le \delta_{RD,o} \quad \text{vs} \quad H_{RD,a}\colon \delta_{RD} > \delta_{RD,o}. \tag{13.7}$$

Then, using the test statistic $\hat{T}_{SRD} = \sqrt{n}(\hat{\delta}_{SRD} - \delta_{SRD,\,o})$ in Corollary 2 for the SRD measure as the pivoting statistic, one can derive Theorem 2 for the RD measure δ_{RD}.

Theorem 2 When $X_T \sim Bernoulli(p_T)$ and $X_C \sim Bernoulli(p_C)$ are Bernoulli distributed, assuming that the variance ratio $\sigma^2 = \frac{\sigma_T^2}{\sigma_C^2}$ is known and $\sigma^2 = \sigma_o^2$ for some fixed number σ_o^2 then at the boundary of the inferiority null hypothesis in Eq. 13.7, the test statistic defined by $\hat{T}_{RD} = \sqrt{n}(\hat{\delta}_{RD} - \delta_{RD,o})$ asymptotically converges to a normal distribution $N(0,\ \Sigma^2_{RD,o})$, where the asymptotic variance $\Sigma^2_{RD,o}$ is given by Eq. 13.8,

$$\Sigma^2_{RD,o} = \left[\left(\sigma^2_{C,o} + \sigma^2_{T,o}\right)\left(1 + \frac{\delta^2_{RD,o}}{16\sigma^2_{C,o}\sigma^2_{T,o}}\right) + \frac{\delta^2_{RD,o}}{2}\right] - \left[\left(1 - 2p_{C,o}\right)\delta_{RD,o}\right], \tag{13.8}$$

where $\sigma^2_{C,o} = p_{C,o}(1-p_{C,o})$ and $\sigma^2_{T,o} = \sigma^2_{C,o}\sigma^2_o$.

Proof The result follows from an application of the Taylor theorem to the function $\delta_{RD} = f_{RD}(\sigma_C,\ \delta_{SRD})$ given by Eq. 13.5 and the test statistic $\widehat{T}_{SRD}$ of Corollary 2, and calculating the product term. ■

It is of interest to point out that the asymptotic variance of $\widehat{T}_{RD}$ takes into consideration the variance differences through the relationship between δ_{RD} and δ_{SRD} as given by Eq. 13.7 to arrive at Theorem 2. Equation 13.8 shows that the asymptotic variance adjusts for the rate of change of the variance function for the Bernoulli distribution at $p_{C,o}$, since $(1 - 2p_{C,o}) = \frac{\mathrm{d}}{\mathrm{d}p_C}\, p_C(1-p_C)|_{p_C = p_{C,o}}$. This is important because as discussed in Chi and Koch (2012), the variance of the Bernoulli distribution decreases to 0 as response rate approaches 1 and the rate of change in the variance function of Bernoulli distributions begins to accelerate when the

response rate exceeds 0.7 and dramatically so as the response rate approaches 1 (or 0). As shown in Chi and Koch (2012), Theorem 2 for the test statistic $\hat{T}_{RD}$ is already an improvement over the corresponding classical Wald test for control response rate in the range (0.5, 1). This improvement is quite substantial for control response rate p_C that approaches 1 due to the fact that the difference in the variance at the boundary of the inferiority null has been taken into account in the test statistic $\hat{T}_{RD}$. As just noted, this difference in variance at the boundary of the inferiority null needs to be accounted for since the rate of change of the variance function of the Bernoulli distribution changes dramatically as the control response rate p_C approaches 1. One can show that this improvement is a result of the fact that the inequality $\left[\left(\sigma_{C,o}^2 + \sigma_{T,o}^2\right) \frac{\delta_{RD,o}^2}{16\sigma_{C,o}^2\sigma_{T,o}^2} + \frac{\delta_{RD,o}^2}{2}\right] < \left[\left(1 - 2p_{C,o}\right)\delta_{RD,o}\right]$ holds for $0.5 < p_{C,o} < 1$ at the boundary of the inferiority null. In addition, within this range of (0.5, 1), the performance of the test statistic $\hat{T}_{RD}$ is comparable to the likelihood ratio test as shown in Chi and Koch (2012).

Remark 1 It should be pointed out that similar results can be established for other binary effect measures, including odds ratio, log odds ratio, relative risk and relative risk reduction by utilizing the corresponding relationship between the SRD measure δ_{SRD}, and each of these binary effect measures analogous to that given by Eq. 13.5 between δ_{SRD} and δ_{RD}. Details of these derivations may be found in Chi and Koch (2012). They are outside the scope of this chapter and is not discussed further here.

In the above derivation thus far, the fixed margins $\delta_{SRD,\,o}$ or $\delta_{RD,o}$ are assumed to have been given and are associated with a given assumed control response rate $p_{C,o}$. For example, $\delta_{SRD,\,o}$ or $\delta_{RD,o}$ could have been determined through the FDA's two-step fixed margin approach (US FDA 2010). But the fixed margins $\delta_{SRD,\,o}$ and $\delta_{RD,o}$ are generally not given by an explicit margin function of the control response rate. The desire to have such a function is apparent in the 1992 FDA Anti-Infective Guidance (US FDA 1992), where it was suggested that a step function, as mentioned earlier, linking the control response rate p_C to the RD measure δ_{RD} should be used, albeit it was retrospectively implemented. Since then, other continuous margin functions have been proposed by various authors as discussed in Chi and Koch (2012). Can an explicit margin function be derived between δ_{RD} and p_C in a natural way that has all the desired properties? The answer is yes, and it is shown in Sect. 13.3 that the SRD measure δ_{SRD} again plays a critical role in establishing such an explicit margin function through its relationship to the inferiority index and the control response rate. Then, in Sect. 13.4, it is shown how to use the empirically derived fixed margin $\delta_{RD,o}$ and control response rate $p_{C,o}$ to define a special margin functions for δ_{RD} with an empirically based degree of stringency. This special margin function for δ_{RD} is then used to integrate the given empirically derived fixed margin into a linear margin called the hybrid margin.

13.3 The Inferiority Index and Margin Function

The definition of an inferiority index between two distributions was defined in Li and Chi (2011) as follows. Again consider an active control trial with a treatment T, a control C, and a clinical outcome X of interest. Assume that the smaller the value of X, the worse is the outcome. Let X_T and X_C denote outcomes on subjects treated with T and C, and $F_{X_T}(t)$ and $F_{X_C}(t)$ denote their cumulative distributions, respectively.

Definition The *inferiority index of the distribution* F_{X_T} *relative to the distribution* F_{X_C} is the quantity

$$\rho = \rho(F_{X_T}, F_{X_C}) = Sup_{-\infty<t<\infty} [F_{X_T}(t) - F_{X_C}(t)]. \tag{13.9}$$

The inferiority index $\rho(F_{X_T}, F_{X_C})$ measures the one-sided maximum separation between the distributions F_{X_T} and F_{X_C} and represents the excess proportion of subjects under treatment T compared to that under treatment C that responded prior to some point t^* at which the maximum separation occurs. Since $0 \le \rho < 1$ is a probability, it can be viewed as an index measuring the *degree of inferiority* of F_{X_r} relative to F_{X_C}. The inferiority index defined in Eq. 13.9 is simply the one-sided distributional analogue of the Kolmogorov–Smirnov statistics. It reflects the distributional differences resulting from various moment differences between the two distributions. For other related distributional concepts, one may refer to the discussion in Li and Chi (2011). An important and useful property of $\rho(F_{X_T}, F_{X_C})$ is that it is *invariant* under parallel location and scale transformations, i.e., if a and $b > 0$ are constants, then $\rho\left(F_{\frac{X_T-a}{b}}, F_{\frac{X_C-a}{b}}\right) = \rho(F_{X_T}, F_{X_C})$.

13.3.1 The Standard Index and Margin Functions Under Normal Distributions

First consider the inferiority index under normal distributions. Let $X_T \sim N(\mu_T, \sigma_T^2)$ and $X_C \sim N(\mu_C, \sigma_C^2)$ be normally distributed with μ_T, μ_C and σ_T^2, σ_C^2 denoting the respective means and variances of their distributions F_{X_T} and F_{X_C}. Let $X_T^* = (X_T - \mu_C)/\sigma_C$ and $X_C^* = (X_C - \mu_C)/\sigma_C$ denote the parallel location and scale transformation of X_T and X_C relative to X_C, respectively. Then, $X_T^* \sim N(\delta_{SRD}, \sigma^2)$ and $X_C^* \sim N(0,1)$, where $\delta_{SRD} = (\mu_T - \mu_C)/\sigma_C$ is the scaled difference measure and $\sigma^2 = \sigma_T^2/\sigma_C^2$ is their variance ratio. It then follows from the invariance property that

$$\rho = \rho(F_{X_T}, F_{X_C}) = \rho(F_{X_T^*}, F_{X_C^*}) = Sup_{-\infty<t<\infty} [\Phi((t - \delta_{SRD})/\sigma) - \Phi(t)], \tag{13.10}$$

where Φ denotes the standard normal distribution. In light of Eq. 13.10, the inferiority index between two normal distributions will be called the *standard inferiority index* and denoted by $\rho_S = \rho(F_{X_T}, F_{X_C})$ for short. From Eq. 13.10, one

can see that ρ_S is linked naturally to the scaled difference measure δ_{SRD} and the variance ratio σ^2 by the function $g_S(\delta_{SRD}, \sigma)$ as defined by Eq. 13.11.

$$\rho_S = g_S(\delta_{SRD}, \sigma) = \begin{cases} [2\Phi(-\delta_{SRD}/2) - 1], & -\infty < \delta_{SRD} \le 0, \text{if } \sigma^2 = 1 \\ \Phi((t^* - \delta_{SRD})/\sigma) - \Phi(t^*), & -\infty < \delta_{SRD} \le 0, \text{if } \sigma^2 \ne 1 \end{cases}, \tag{13.11}$$

where $t^* = \frac{-\delta_{SRD}\,\sigma^{-2} - \sqrt{\delta_{SRD}^2\,\sigma^{-2} + (1-\sigma^{-2})\log\sigma^2}}{(\sigma^{-2}-1)}$ denote the point at which the supremum in Eq. 13.10 is attained. The function $\rho_S = g_S(\delta_{SRD}, \sigma)$ is called *the standard inferiority index function,* or simply the *standard index function* for short. For any value $\delta_{SRD,\,o}$ of the scaled difference measure δ_{SRD} and any value σ_o^2 of the variance ratio σ^2, the standard inferiority index function $g_S(\delta_{SRD,\,o}, \sigma_o)$ assigns an inferiority index value $\rho_{S,o}$ indicating the degree of stringency of $\delta_{SRD,\,o}$ at the given variance ratio σ_o^2.

Conversely, for a specified level of the standard inferiority index $\rho_S = \rho_{S,o}$, there is a *standard margin function* $\delta_{SRD}(\sigma|\rho_{S,o})$, which is defined by Eq. 13.12.

$$\delta_{SRD}(\sigma|\rho_{S,o}) = \begin{cases} -2\Phi^{-1}\left(\frac{\rho_{s,o}+1}{2}\right) < 0, & \text{for } \sigma = 1, \quad 0 \le \rho_{S,o} < 1 \\ g_S^{-1}(\rho_{S,o}, \sigma), & \sigma \in (\sigma_1(\rho_{S,o}),\ \sigma_2(\rho_{S,o})) \ \&\ \sigma \ne 1, \quad 0 \le \rho_{S,o} < 1 \end{cases}. \tag{13.12}$$

For a given inferiority index value of $\rho_{S,o}$, the interval $(\sigma_1(\rho_{S,o}),\ \sigma_2(\rho_{S,o}))$ in Eq. 13.12 is determined by setting $\delta_{SRD} = 0$ under the second alternative in Eq. 13.11 when $\sigma^2 \ne 1$ as shown by Eq. 13.13.

$$\begin{aligned} \rho_s &= \Phi\left(\frac{t_{max}\,(0, \sigma\,(\rho_s))}{\sigma}\right) - \Phi(t_{max}\,(0, \sigma\,(\rho_s)) \\ &= \Phi(-\sqrt{(1-\sigma^{-2})\log\sigma^2}/(1-\sigma^{-2})\sigma) - \Phi(-\sqrt{(1-\sigma^{-2})\log\sigma^2}/(1-\sigma^{-2})). \end{aligned} \tag{13.13}$$

In Eq. 13.12, when the variance ratio $\sigma = 1$, the margin function $\delta_{SRD}(\sigma|\rho_{S,o})$ is given by $-2\Phi^{-1}\left(\frac{\rho_{s,o}+1}{2}\right)$, which is derived from the first alternative in Eq. 13.11. For variance ratio $\sigma \in (\sigma_1(\rho_{S,o}),\ \sigma_2(\rho_{S,o}))$ and $\sigma \ne 1$, the inverse function $g_S^{-1}(\rho_{S,o}, \sigma)$ is solved implicitly from the second alternative in Eq. 13.11.

For a specified value of the inferiority index $\rho_S = \rho_{S,o}$, the standard margin function $\delta_{SRD}(\sigma|\rho_{S,o})$ in Eq. 13.12 has the same degree of stringency given by $\rho_{S,o}$ throughout the interval $(\sigma_1(\rho_{S,o}),\ \sigma_2(\rho_{S,o}))$. Then, for any given variance ratio σ_o^2, the margin function defines a fixed margin $\delta_{SRD,\,o} = \delta_{SRD}(\sigma_o|\rho_{S,o})$ that can be used to define a fixed margin NI hypothesis for δ_{SRD} as given in Eq. 13.14 with the degree of stringency $\rho_{S,o}$.

$$\begin{aligned} H_{SRD,\,o}&: \delta_{SRD} \le \delta_{SRD,\,o} = \delta_{SRD}(\sigma_o|\rho_{S,o}) \quad \text{vs.} \\ H_{SRD,a}&: \delta_{SRD} > \delta_{SRD,\,o}(\sigma) = \delta_{SRD}(\sigma_o|\rho_{S,o}). \end{aligned} \tag{13.14}$$

Therefore, if the fixed NI margin $\delta_{SRD,\,o} = \delta_{SRD}(\sigma_o|\rho_{S,o})$ happens to be derived from the margin function defined by Eq. 13.12 at a specified standard inferiority index level $\rho_{S,o}$ and a given variance ratio σ_o^2, then Corollary 1 would be applicable and the NI hypothesis in Eq. 13.14 may be rejected at the $\alpha = 0.025$ level of significance if the test statistic $\widehat{T}_{SRD,\,o} = \sqrt{n}\,(\widehat{\delta}_{SRD} - \delta_{SRD}(p_{C,o}|\rho_{S,\,o}))/\Sigma_{SRD,\,o} > 1.96$.

13.3.2 The Standard Index and Margin Functions Under the Bernoulli Distributions

Now let $X_T \sim Bernoulli\,(p_T)$ and $X_C \sim Bernoulli\,(p_C)$ be two independent Bernoulli random variables with distributions $F_{X_T}(t) = 1 - p_T$, at $t = 0$ and $= p_T$, at $t = 1$, and $F_{X_C}(t) = 1 - p_C$, at $t = 0$ and $= p_C$, at $t = 1$. Assuming $p_T < p_C$, then from the definition of the inferiority index given in Eq. 13.9, it follows that $\rho(F_{X_T}, F_{X_C}) = [F_{X_T}(0) - F_{X_C}(0)] = -(p_T - p_C) = -\delta_{RD}$. Thus, based on the definition given by Eq. 13.9, the inferiority index between two Bernoulli distributions is simply equal to the negative of the RD measure δ_{RD} and is not a function of the variance ratio σ^2. What this implies is that the index $\rho(F_{X_T}, F_{X_C}) = -\delta_{RD}$ cannot account for any potential difference in the variance between the treatment and control. This is important because as discussed in Chi and Koch (2012), for Bernoulli distributions, the slope of the variance function changes dramatically outside the range (0.3, 0.7) when the response rate moves towards 1 (or 0). This is the precise reason why one needs to adjust the margin for the RD measure by σ_C if one wants to be able to define a margin function that properly accounts for the anticipated differences in the rate of change of σ_T and σ_C at the boundary of the inferiority null hypothesis. This is especially relevant when the control response rate is outside the range of (0.30, 0.70). This is consistent with the intuition that as the control response rate p_C moves closer to 1 (or 0), then the NI margin should become tighter and tighter. Therefore, the inferiority index as defined in Eq. 13.9 would not be useful under Bernoulli distributions and a different strategy is needed. The alternative strategy is to use the standard index function ρ_S given in Eq. 13.11 under the normal distributions for the Bernoulli distributions. This strategy is possible on account of Theorem 3.

13.3.2.1 Linking the Standard Inferiority Index to the Scaled Rate Difference Measure δ_{SRD}

Theorem 3: 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 $\widehat{p}_T = \sum_{i=1}^n X_{T,i}/n$ and $\widehat{p}_C = \sum_{j=1}^n X_{C,j}/n$ denote the sample means of X_T and X_C, respectively, and $\widehat{p}^*_{T,n} = \sqrt{n}[(\widehat{p}_T - p_C)/\sigma_C]$ and $\widehat{p}^*_{C,n} = \sqrt{n}[(\widehat{p}_C - p_C)/\sigma_C]$ denote their parallel location and scale transforms. Let $\rho_n = \rho(F_{\widehat{p}^*_{T,n}}, F_{\widehat{p}^*_{C,n}})$ denote

the inferiority index between the distributions of the two transformed statistics. Then,

$$\lim_{n\to\infty} \rho(F_{\widehat{p}^*_{T,n}},\ F_{\widehat{p}^*_{C,n}}) = \rho_S(\Phi(\delta_{SRD},\ \sigma^2),\ \Phi) \tag{13.15}$$

where $\rho_S(\Phi(\delta_{SRD},\ \sigma^2),\ \Phi)$ is the standard inferiority index between the cumulative normal distribution $\Phi(\delta_{SRD},\ \sigma^2)$ and the standard normal distribution Φ, where $\delta_{SRD} = (p_T - p_C)/\sigma_C$ with $\sigma_C = \sqrt{p_C(1-p_C)}$ and $\sigma^2 = \sigma_T^2/\sigma_C^2$ with $\sigma_T = \sqrt{p_T(1-p_T)}$.

Proof: It follows from the central limit theorem that $\widehat{p}_T \sim N(p_T,\ \sigma_T^2/n)$ and $\widehat{p}_C \sim N(p_C,\ \sigma_C^2/n)$, where $\sigma_T^2 = p_T(1-p_T)$ and $\sigma_C^2 = p_C(1-p_C)$. Then, one has $\widehat{p}^*_{T,n} = \sqrt{n}[(\widehat{p}_T - p_C)/\sigma_C] \sim N(\delta_{SRD},\ \sigma^2)$ and $\widehat{p}^*_{C,n} = \sqrt{n}[(\widehat{p}_C - p_C)/\sigma_C] \sim N(0,\ 1)$, where $\delta_{SRD} = (p_T - p_C)/\sigma_C$ and $\sigma^2 = \sigma_T^2/\sigma_C^2$. Then, Eq. 13.15 follows from the definition of inferiority index, its invariance property under parallel location and scale transformation and an application of the Berry–Esseen theorem (Berry 1941, Esseen 1942) on the uniform convergence of the central limit theorem. Details are omitted. ■

Then, from Eqs. 13.11 and 13.15, one has

$$\rho_S = \rho_S(\Phi(\delta_{SRD},\ \sigma^2),\ \Phi) = g_S(\delta_{SRD},\ \sigma). \tag{13.16}$$

Therefore, Theorem 3 and Eq. 13.16 show that the SRD measure $\delta_{SRD} = (p_T - p_C)/\sigma_C$ and the variance ratio $\sigma^2 = \sigma_T^2/\sigma_C^2$ under the Bernoulli distributions are asymptotically linked to the standard inferiority index ρ_S by the standard index function g_S. Now, by substituting the functional relationship between p_T, p_C, and δ_{SRD} as given by $\pi_T(p_C, \delta_{SRD})$ in Eq. 13.17,

$$p_T = \pi_T(p_C, \delta_{SRD}) = p_C + \sigma_C \delta_{SRD} = p_C + \sqrt{p_C(1-p_C)}\,\delta_{SRD}, \tag{13.17}$$

into the variance ratio $\sigma^2 = p_T(1-p_T)/p_C(1-p_C)$ in Eq. 13.16, one derives the index function g^*_{SRD},

$$\begin{aligned}\rho_S &= g^*_{SRD}\,(\delta_{SRD}, p_C)\\ &= g_S\left(\delta_{SRD}, \sqrt{\frac{\pi_T\,(p_C, \delta_{SRD})\,(1-\pi_T\,(p_C, \delta_{SRD}))}{p_C\,(1-p_C)}}\right), \text{for } \delta_{SRD} < 0 \text{ and } 0 < p_C < 1.\end{aligned} \tag{13.18}$$

Equation 13.18 shows that the standard inferiority index ρ_S is now asymptotically linked to SRD measure δ_{SRD} and the control response rate p_C by the function g^*_{SRD} which is defined through the composition of the standard index function g_S and the variance ratio as a function of δ_{SRD} and p_C given by $\sigma^2 = \gamma(\delta_{SRD},\ p_C) = \frac{\pi_T(p_C,\ \delta_{SRD})(1-\pi_T(p_C,\delta_{SRD}))}{p_C(1-p_C)}$, where π_T is defined in Eq. 13.17. The key point here is that the index function g^*_{SRD} has now incorporated the variance ratio σ^2 into its relationship, even though it now appears to be only a function of δ_{SRD} and p_C. This index function g^*_{SRD} then allows one to use the standard inferiority index ρ_S as an objective measure for assessing the degree of stringency for any value of the SRD

measure $\delta_{SRD} = \delta_{SRD,\,o}$ at any control response rate $p_C = p_{C,o}$. Conversely, upon setting the standard inferiority index ρ_S at a specific level $\rho_{S,o}$ in its inverse function g^{*-1}_{SRD}, which is derived from the inverse function g_S^{-1} through Eq. 13.18, one derives a margin function $\delta_{SRD}\,(p_C|\rho_{S,\,o})$ for the SRD measure δ_{SRD} as given by Eq. 13.19,

$$\delta_{SRD}\,(p_C|\rho_{S,\,o}) = g^{*-1}_{SRD}(\rho_{S,\,o}, p_C). \tag{13.19}$$

This specific indexed margin function corresponds to a level curve of the surface of the index function g^*_{SRD} given in Eq. 13.18 by setting the index level $\rho_S = \rho_{S,\,o}$. Thus, in a given application, if the control response rate p_C is thought to be equal to $p_{C,o}$, then Eq. 13.20

$$\delta_{SRD,\,o} = \delta_{SRD}(p_{C,o}|\rho_{S,\,o}) = g^{*-1}_{SRD}(\rho_{S,\,o}, p_{C,o}) \tag{13.20}$$

defines a fixed margin at the control response rate $p_{C,o}$ with the degree of stringency $\rho_{S,\,o}$. With this fixed margin $\delta_{SRD,\,o}$, the NI hypothesis for δ_{SRD} can then be stated as

$$\begin{aligned} H_{SRD,\,o}\colon\ \delta_{SRD} \le \delta_{SRD,\,o} = \delta_{SRD}(p_{C,o}|\rho_{S,\,o}) \quad \text{vs.} \\ H_{SRD,a}\colon\ \delta_{SRD} > \delta_{SRD,\,o} = \delta_{SRD}(p_{C,o}|\rho_{S,\,o}) \end{aligned} \tag{13.21}$$

and Corollary 2 would be applicable. It shows that the test statistic $\widehat{T}_{SRD,\,o}$ at the boundary of the inferiority null of Eq. 13.21 for the SRD measure δ_{SRD} converges asymptotically to a normal distribution. The inferiority null hypothesis in Eq. 13.21 may be rejected at the $\alpha = 0.025$ significance level if the test statistic $\widehat{T}_{SRD,\,o} = \sqrt{n}\,(\widehat{\delta}_{SRD} - \delta_{SRD}(p_{C,o}|\rho_{S,\,o}))/\Sigma_{SRD,\,o} > 1.96$.

13.3.2.2 Linking the Standard Inferiority Index to the Rate Difference Measure δ_{RD}

The relationship between δ_{SRD} and δ_{RD} is given by $\delta_{SRD} = f_{RD}(\delta_{RD}, p_C) = \delta_{RD}/\sigma_C$. Upon substituting this relationship into Eq. 13.18, one obtains the index function g^*_{RD} given in Eq. 13.22,

$$\rho_S = g^*_{RD}(\delta_{RD},\ p_C) = g^*_{SRD}(f_{RD}(\delta_{RD}, p_C),\ p_C)\,, \quad \text{for } 0 < p_C < 1, \tag{13.22}$$

which links the standard inferiority index ρ_S to δ_{RD} and p_C. From Eq. 13.22, one can derive the margin function given by Eq. 13.23 that links ρ_S and p_C to δ_{RD} given by

$$\delta_{RD} = g^{*-1}_{RD}(\rho_S, p_C), \text{ for } 0 < \rho_S < 1 \text{ and } 0 < p_C < 1. \tag{13.23}$$

Analogous to the case for the SRD measure δ_{SRD}, one can use the index function g^*_{RD} defined by Eq. 13.22 to assess the degree of stringency of any value of the RD measure $\delta_{RD} = \delta_{RD,o}$ at any given control response rate $p_C = p_{C,o}$. Similarly, by setting the standard index $\rho_S = \rho_{S,o}$ in Eq. 13.23, one can define a specific indexed margin function for δ_{RD} given by Eq. 13.24

$$\delta_{RD}(p_C|\rho_{S,o}) = g^{*-1}_{RD}(\rho_S, o, p_C), \text{ for } 0 < p_C < 1 \tag{13.24}$$

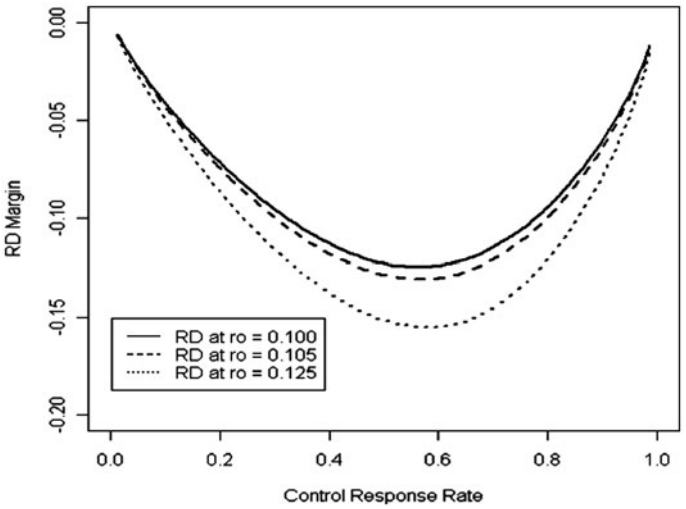

Fig. 13.1 Margin functions $\delta_{RD}(p_C|\rho_S)$ for inferiority index $\rho_S = 0.10, 0.105$, and 0.125. *RD* rate difference

with a degree of stringency given by $\rho_{S,o}$. Now, for a given $p_C = p_{C,o}$, the indexed margin function Eq. 13.24 defines a fixed margin $\delta_{RD,o} = \delta_{RD}(p_{C,o}|\rho_{S,o}) = g^{*-1}_{RD}(\rho_{S,o}, p_{C,o})$ for the RD measure δ_{RD}. Note that the fixed margin $\delta_{RD,o}$ has now been adjusted for σ_C via δ_{SRD} through f_{RD} in Eq. 13.22. One can then use this fixed margin $\delta_{RD,o}$ to define a NI hypothesis relative to δ_{RD} with the degree of stringency $\rho_{S,\ o}$ as given by Eq. 13.25,

$$H_{RD,\ o}:\ \delta_{RD} \le \delta_{RD,o} = \delta_{RD}(p_{C,o}|\rho_{S,o}) \quad \text{vs.}$$
$$H_{RD,a}:\ \delta_{RD} > \delta_{RD,o} = \delta_{RD}(p_{C,o}|\rho_{S,o}). \tag{13.25}$$

Theorem 2 shows that the test statistic $\widehat{T}_{RD,o}$ at the boundary of the inferiority null of Eq. 13.25 for the RD measure δ_{RD} converges asymptotically to a normal distribution. The inferiority null hypothesis in Eq. 13.25 may be rejected at the $\alpha = 0.025$ significance level if the test statistic $\widehat{T}_{RD,o} = \sqrt{n}\,(\widehat{\delta}_{RD} - \delta_{RD}(p_{C,o}|\rho_{S,\ o}))/\Sigma_{RD,o} > 1.96$.

It is of interest to note that the inverse function defined by Eq. 13.23 defines a family of margin functions given by Eq. 13.26

$$\{\delta_{RD}(p_C|\rho_S) = g^{*-1}_{RD}(\rho_S,\ p_C), \quad \text{for } 0 < \rho_S < 1 \text{ and } 0 < p_C < 1\} \tag{13.26}$$

as illustrated in Fig. 13.1.

By setting the standard index ρ_S equal to a specific value $\rho_{S,o}$, then $\delta_{RD}(p_C|\rho_{S,o})$ defines an indexed margin function with a stringency level of $\rho_{S,o}$ for δ_{RD} as a function of the control response rate p_C. This entire margin function $\delta_{RD}(p_C|\rho_{S,o})$ will have the same degree of stringency $\rho_{S,o}$ at every control response rate p_C, for $0 < p_C < 1$. Furthermore, at the given index level $\rho_{S,\ o}$, the margin function $\delta_{SRD}\ (p_C|\rho_{S,\ o})$ defined by Eq. 13.19 and the margin function $\delta_{RD}\ (p_C|\rho_{S,\ o})$

defined by Eq. 13.24 are equally stringent with the same degree of stringency $\rho_{S,o}$. At a given control response rate $p_{C,o}$, these margin functions define equally stringent NI hypotheses as given by Eqs. 13.21 and 13.25. However, the performance of the test statistics $\hat{T}_{SRD,\,o}$and $\hat{T}_{RD,o}$for their respective NI hypotheses, Eqs. 13.21 and 13.25, may differ. Similar derivations can be done for other binary effect measures by using their corresponding functional relationship with the SRD measure δ_{SRD} or the RD measure δ_{RD} to arrive at equally stringent margin functions for these binary effect measures. These equally stringent margin functions can then be used to define equally stringent NI hypotheses. The relative performance of the test statistics for these equally stringent NI hypotheses can then be investigated. One may refer to Chi and Koch (2012) for a discussion of such an investigation comparing the RD measure and the log odds ratio measure.

13.4 A Hybrid Design for the Rate Difference Measure

It has been shown in Sect. 13.2 how to improve the efficiency of the test statistic for testing the fixed margin NI hypothesis in Eq. 13.7 by incorporating the information on the variance ratio at the boundary of the inferiority null of Eq. 13.7 into its asymptotic variance. In Sect. 13.3, an index function g^{*}_{RD} has been derived in Eq. 13.22 that links the RD measure δ_{RD} and control response rate p_C to the standard inferiority index ρ_S. Furthermore, its inverse function g^{*-1}_{RD} in Eq. 13.23 links the standard inferiority index ρ_S and control response rate p_C to the RD measure δ_{RD} so that by setting the standard inferiority index ρ_S at a specified level $\rho_{S,o}$, the inverse function then defines a margin function $\delta_{RD}(p_C|\rho_{S,o})$ given by Eq. 13.24 which has the degree of stringency $\rho_{S,o}$.

In this section, these results are combined to produce a hybrid design for NI trials with binary outcomes intended to address the question of how to set a margin and what margin to use in the event the true control response rate appears to deviate from the assumed control response rate.

13.4.1 An Empirically Based Margin Function for the Rate Difference Measure

How to set the NI margin is a problem that has been around for quite a while. The FDA's proposed two-step empirically based fixed margin approach is really a very good approach. However, it needs to be supplemented by an objective measure of the degree of stringency of the empirically derived fixed margin, and in addition, the fixed margin design needs to be modified to be able to accommodate variability in the margin in the event the true control response rate actually deviates from the assumed rate. It is the purpose of this section to show how the index function g^{*}_{RD} defined by Eq. 13.22 and the margin function g^{*-1}_{RD} defined by Eq. 13.23 can be used in tandem to address both issues in a hybrid design that preserves the empirical nature of FDA's fixed margin approach.

Now consider the problem of designing an NI trial with binary outcomes using the RD measure δ_{RD}. Assume that relevant historical studies involving the active control and placebo are available. Using the FDA's two-step fixed margin approach described above, one can derive an estimate $p_{C,o}$ for the control response rate p_C and a conservative estimate of the CE. Furthermore, with input from clinical experts, an NI margin $\delta_{RD,o}$ is derived which represents the maximum amount of loss of the CE that can be tolerated.

Then, from the pair of empirically based estimates $(\delta_{RD,o}, p_{C,o})$, one can derive the degree of stringency of the margin $\delta_{RD,o}$ at $p_{C,o}$ from the index function g^*_{RD} defined by Eq. 13.22, which is given by

$$\rho_{S,o} = g^*_{RD}(\delta_{RD,o}, p_{C,o}). \tag{13.27}$$

Now, the index $\rho_{S,o}$ given by Eq. 13.27 is *empirically based* because it is derived from the empirically based estimates $(\delta_{RD,o}, p_{C,o})$ through the index function g^*_{RD} as defined by Eq. 13.27.

Using this empirically based index $\rho_{S,o}$, one can define an empirically based margin function through the inverse function g^{*-1}_{RD} given by Eq. 13.24, or Eq. 13.28,

$$\delta_{RD}(p_C|\rho_{S,o}) = g^{*-1}_{RD}(\rho_{S,o}, p_C), \text{ for } 0 < p_C < 1. \tag{13.28}$$

It is obvious that when this margin function is evaluated at the estimate $p_{C,o}$, it should yield the empirically based margin $\delta_{RD,o}$, i.e., one has

$$\delta_{RD}(p_{C,o}|\rho_{S,o}) = \delta_{RD,o}. \tag{13.29}$$

Thus, from the empirically based pair of estimates $(\delta_{RD,o}, p_{C,o})$, one is able to derive the corresponding empirically based standard inferiority index $\rho_{S,o}$ through the index function g^*_{RD} as given by Eq. 13.27. Then, using empirically based index $\rho_{S,o}$, one can define an empirically based margin function $\delta_{RD}(p_C|\rho_{S,o})$ given by Eq. 13.28 that has the degree of stringency given by $\rho_{S,o}$. Thus, out of the family of possible margin functions defined by Eq. 13.26, one identifies a special indexed margin function $\delta_{RD}(p_C|\rho_{S,o})$ that is based on the empirically based pair of estimates $(\delta_{RD,o}, p_{C,o})$.

Hence, it has now been shown that from the empirically based pair of estimates $\delta_{RD}(p_C|\rho_{S,o})$, one can derive its *implicit* degree of stringency $\rho_{S,o}$ through the index function $g^*_{RD}(\delta_{RD}, p_C)$ given by Eq. 13.22. In actual practice, based on the degree of stringency $\rho_{S,o}$, one may opt to further tighten or relax the empirically derived fixed margin $\delta_{RD,o}$ as deemed appropriate. Now assume that such adjustment has been done if needed. Then, one can simply define the NI hypothesis in Eq. 13.25 using this empirically based margin $\delta_{RD,o}$ and test the inferiority null hypothesis using Theorem 2. This approach without the link to the standard index function ρ_S is essentially what has routinely been done. But as noted earlier, the FDA Anti-Infective Division has posed the question as to what margin to use in the event the control response rate from the current NI trial appears to deviate from the estimated control response rate $p_{C,o}$? Obviously, by simply defining an NI hypothesis (Eq. 13.25) based on an empirically derived fixed margin, one will not be able to address this question. So, further work is needed and is discussed in the next section.

13.4.2 A Hybrid Design with a Linear Margin

Now consider the NI hypothesis in Eq. 13.30

$$H_{RD,o} : \delta_{RD} \leq \delta_{RD}(p_C|\rho_{S,o}) \quad \text{vs} \quad H_{RD,a} : \delta_{RD} > \delta_{RD}(p_C|\rho_{S,o}), 0 < p_C < 1. \tag{13.30}$$

Unlike the NI hypotheses in Eq. 13.25, the NI hypothesis in Eq. 13.30 is actually defined by a margin function, and not by a fixed margin. But it is not just any margin function. It is a natural and empirically derived margin function with the empirically determined degree of stringency $\rho_{S,o}$. In Zhang (2006), the author starts off with a given margin function and develops his method for a general variable margin. In Sect. 13.4.1, a natural and special indexed margin function is derived with the empirically determined degree of stringency. In this section, this empirically based margin function will be used to integrate the fixed margin into a linear margin for the hybrid design to be proposed.

Figure 13.1 displays the graphs of margin function $\delta_{RD}(p_C|\rho_S)$ at three selected degrees of stringency. From these graphs, one can see that the power for testing the NI hypothesis in Eq. 13.30 will be low if the true control response rate p_C is considerably larger (or smaller) than the empirically based estimate $p_{C,o}$ because the margin is getting tighter as p_C approaches 1 (or 0).

Since the index function g^*_{RD} is continuously differentiable, it follows from the implicit function theorem that the margin function $\delta_{RD}(p_C|\rho_{S,o})$ given by Eq. 13.28 is continuously differentiable and its derivative is given by

$$\frac{\partial \delta_{RD}(p_C|\rho_{S,o})}{\partial p_C} = \frac{\partial g^{*-1}_{RD}(\rho_{S,o}, p_C)}{\partial p_C} = -\frac{\frac{\partial g^*_{RD}}{\partial p_C}}{\frac{\partial g^*_{RD}}{\partial \delta_{RD}}}. \tag{13.31}$$

Now consider the first-order Taylor approximation of the margin function $\partial \delta_{RD}(p_C|\rho_{S,o})$ expanded around the point $p_{C,o}$ given in Eq. 13.32 as illustrated by Fig. 13.2:

$$\begin{aligned} L(p_C|\rho_{S,o}, p_{C,o}) &= g^{*-1}_{RD}(\rho_{S,o}, p_C)|_{p_C=p_{C,o}} + \frac{\partial g^{*-1}_{RD}(\rho_{S,o}, p_C)}{\partial p_C}|_{p_C=p_{C,o}}(p_C - p_{C,o}) \\ &= \delta_{RD}(p_{C,o}|\rho_{S,o}) + \frac{\partial g^{*-1}_{RD}(\rho_{S,o}, p_{C,o})}{\partial p_C}(p_C - p_{C,o}) = \delta_{RD,o} \\ &\quad + \frac{\partial g^{*-1}_{RD}(\rho_{S,o}, p_{C,o})}{\partial p_C}(p_C - p_{C,o}). \end{aligned} \tag{13.32}$$

The expression in the linear approximation $L(p_C|\rho_{S,o}, p_{C,o})$ in Eq. 13.32 is equal to the fixed margin $\delta_{RD,o}$ plus the linear term $\frac{\partial g^{*-1}_{RD}(\rho_{S,o}, p_{C,o})}{\partial p_C}(p_C - p_{C,o})$. If the true control response rate p_C from the NI trial turns out to be equal to $p_{C,o}$, then $L(p_{C,o}|\rho_{S,o}, p_{C,o}) = \delta_{RD,o}$. But if p_C deviates from $p_{C,o}$, then the margin is equal to the given fixed margin $L(p_C|\rho_{S,o}, p_{C,o}) = \delta_{RD,o}$ plus the deviation term

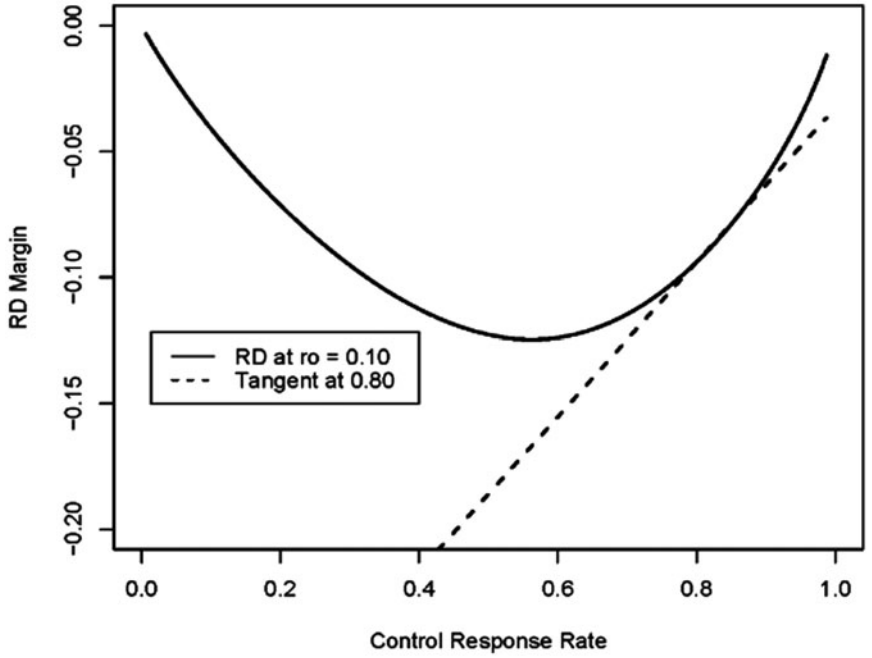

Fig. 13.2 First-order Taylor approximation to the margin function $\delta_{RD}(p_C|\rho_{S,o} = 0.10)$ at $p_{C,o} = 0.80$. RD rate difference

$\frac{\partial g_{RD}^{*-1}(\rho_{S,o}, p_{C,o})}{\partial p_C}(p_C - p_{C,o})$, which represents a first-order adjustment to the margin $\delta_{RD,o}$ for the deviation.

The linear margin $L(p_C|\rho_{S,o}, p_{C,o})$ is called a hybrid margin because it *explicitly integrates* the given empirically derived pair $(\delta_{RD,o}, p_{C,o})$ based on FDA's two-step fixed margin approach with a variable term $\frac{\partial g_{RD}^{*-1}(\rho_{S,o}, p_{C,o})}{\partial p_C}(p_C - p_{C,o})$ that can accommodate the possibility that the true control response rate p_C may deviate somewhat from the best empirically based estimate of the control response rate $p_{C,o}$.

Now a natural question to ask is how stringent is the linear margin $L(p_C|\rho_{S,o}, p_{C,o})$? The empirically based margin function $\delta_{RD}(p_C|\rho_{S,o})$ has the stringency $\rho_{S,o}$, so the linear margin function $L(p_C|\rho_{S,o}, p_{C,o})$ cannot be at this same 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 example, with $(\delta_{RD,o}, p_{C,o}) = (-0.10, 0.80)$, this interval is approximately (0.75, 0.90) as shown in Table 13.1.

Therefore, now one may consider the following hybrid NI hypothesis as approximately equivalent to the NI hypothesis in Eq. 13.30 within a certain interval of p_C:

$$H_{RD,o} : \delta_{RD} \leq L(p_C|\rho_{S,o}, p_{C,o}) \quad \text{vs}$$
$$H_{RD,a} : \delta_{RD} > L(p_C|\rho_{S,o}, p_{C,o}), p_{o,L} < p_C < p_{o,R}. \tag{13.33}$$

Table 13.1 Comparing the margin functions $\delta_{RD}(p_C|\rho_{S,o})$ and $L(p_C|\rho_{S,o}, p_{C,o})$ with Taylor expansion at $p_{C,o} = 0.80$

True control	Margin function	
Response ate p_C	$\delta_{RD}(p_C\|\rho_{S,o})$	$L(p_C\|\rho_{S,o}, p_{C,o})$
0.50	− 0.1228	− 0.1859
0.55	− 0.1246	− 0.1706
0.60	− 0.1239	− 0.1553
0.65	− 0.1207	− 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	− 0.0601	− 0.0614
0.95	− 0.0360	− 0.0460

The hybrid NI hypothesis (Eq. 13.32) can be equivalently written as the NI hypothesis in Eq. 13.33,

$$\begin{aligned} H_{RD,o} &: \delta_{RD} - L(p_C|\rho_{S,o}, p_{C,o}) \leq 0 \quad \text{vs} \\ H_{RD,a} &: \delta_{RD} - L(p_C|\rho_{S,o}, p_{C,o}) > 0, p_{o,L} < p_C < p_{o,R}. \end{aligned} \tag{13.34}$$

13.4.3 The Test Statistic for the Hybrid Design NI Hypothesis

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

$$\hat{\Delta}_{RD} = [\hat{\delta}_{RD} - L(\hat{p}_C|\rho_{S,o}, p_{C,o})]. \tag{13.35}$$

Then,

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

Let

$$\hat{\Delta}_{RD,o} = [\hat{\Delta}_{RD} - E(\hat{\Delta}_{RD}|H_o)]. \tag{13.36}$$

The asymptotic normality of the test statistic $\hat{\Delta}_{RD,o}$ at the inferiority null of the hybrid NI hypothesis (Eq. 13.33 or Eq. 13.34) is established in Theorem 4.

Theorem 4 The statistic $\sqrt{n}\hat{\Delta}_{RD,o}$ is asymptotically normal $N(o, \sum(p_C|H_o))$ at the boundary of the inferiority null of Eq. 13.33 or Eq. 13.34, where

$$\sum{}^2(p_C|H_o) = \sum{}^2_{RD,o}(p_C|H_o) + \left[2\left(\frac{\partial g^{*-1}_{RD}(\rho_{S,o}, p_{C,o})}{\partial p_C}\right) + \left(\frac{\partial g^{*-1}_{RD}(\rho_{S,o}, p_{C,o})}{\partial p_C}\right)^2\right] p_C(1-p_C), \tag{13.37}$$

and

$$\Sigma^2_{RD,o}\left(p_{C,o}|H_o\right) = \left[\left(\sigma^2_{C,o}+\sigma^2_{T,o}\right)\left(1+\frac{\delta^2_{RD,o}}{16\sigma^2_{C,o}\sigma^2_{T,o}}\right)+\frac{\delta^2_{RD,o}}{2}\right] - \left[\left(1-2p_{C,o}\right)\delta_{RD,o}\right]$$

is the variance of the statistic under the fixed margin NI hypothesis (Eq. 13.25) 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 $\sigma^2_{C,o} = p_{C,o}(1-p_{C,o})$, $\sigma^2_{T,o} = p_{T,o}(1-p_{T,o})$, and $p_{T,o} = p_{C,o} + \delta_{RD,o}$.

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

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

by applying Eq. 13.8 of Theorem 2. ■

Hence, the hybrid test statistic,

$$\widehat{T}^{HB}_{RD,o} = \frac{\sqrt{n}\,\widehat{\Delta}_{RD,\,o}}{\sqrt{\Sigma(p_C|H_o)}} \sim N(0,1), \tag{13.38}$$

where the unknown true p_C may be substituted by the sample proportion $\widehat{p}_C$. The hybrid inferiority hypothesis in Eq. 13.33 or Eq. 13.34 may be rejected at the $\alpha = 0.025$ significance level if $\widehat{T}^{HB}_{RD,o} = \sqrt{n}\frac{[\widehat{\delta}_{RD} - \mathcal{L}(\widehat{p}_C|\rho_{S,o}, p_{C,o})]}{\Sigma(\widehat{p}_C|H_o)} > 1.96$.

13.4.4 The Performance of the Hybrid Test Statistic $\widehat{T}^{HB}_{RD,o}$

It should be pointed out that the focus of the hybrid NI design is still on the fixed 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 may deviate somewhat from $p_{C,o}$. Therefore, it would be of interest to investigate the performance of the test $\widehat{T}^{HB}_{RD,o}$ at $p_{C,o}$.

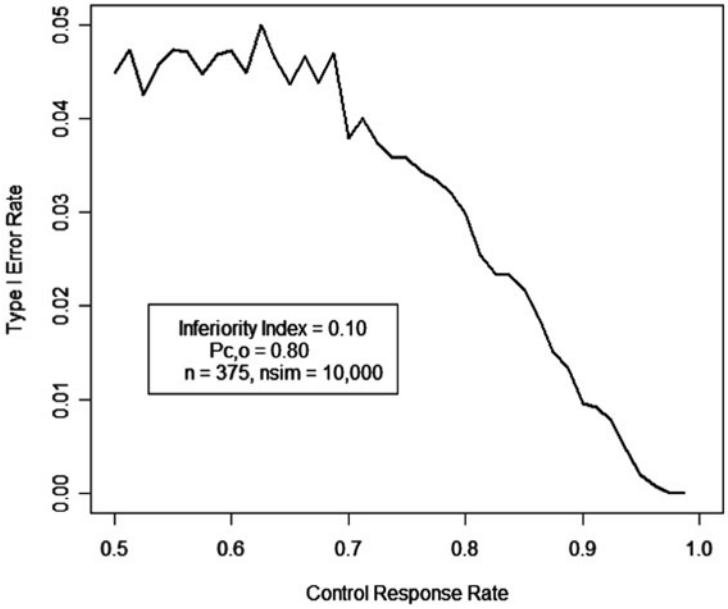

Fig. 13.3 Simulated overall type I error rate for hybrid design with the linear margin $\mathcal{L}(p_C|\rho_{S,o}, p_{C,o})$, where $\rho_{S,o} = 0.10$, $P_{c,o} = 0.80$, and $\alpha = 0.025$

13.4.4.1 Simulation of the Type I Error Rate

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

$$\alpha(p_C) = 1 - \Phi\left(\frac{\sqrt{n}\widehat{\Delta}_{RD,o}}{\sqrt{\Sigma\,(p_C|H_o)}}\right). \tag{13.39}$$

Figure 13.3 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 $\widehat{p}_C$. Furthermore, for $p_C < p_{C,o}$, the margin becomes more liberal, whereas for $p_C > p_{C,o}$, the margin becomes tighter. Therefore, by using a piecewise linear margin as discussed in Remark 3 should improve the type I error control substantially for $p_C < p_{C,o}$.

In light of the type I error rate inflation when $p_C = p_{C,o}$, one may wish to control this by lowering the significance level α. Table 13.2 and Fig. 13.4 show that if the overall significance level is lowered to approximately $\alpha = 0.020$, then the simulated type I error rate when $p_C = p_{C,o}$ is roughly controlled at 0.025.

However, instead of lowering the significance level $\alpha = 0.025$ to 0.020, it might be more preferable to consider replacing the linear margin $\mathcal{L}(p_C|\rho_{S,o}, p_{C,o})$ by a piecewise linear margin constructed by joining together two or more (if necessary) linear

Table 13.2 Simulated unadjusted and adjusted type I error rates for hybrid design with Taylor expansion at the point $p_{C,o} = 0.80$

True control	Significance level	
Response rate p_C	0.025	0.020
0.50	0.0448	0.0398
0.55	0.0473	0.0383
0.60	0.0472	0.0415
0.65	0.0436	0.0370
0.70	0.0378	0.0376
0.75	0.0358	0.0302
0.80	0.0298	0.0248
0.85	0.0217	0.0167
0.90	0.0096	0.0082
0.95	0.0020	0.0011

margins $\mathcal{L}(p_C|\rho_{S,o}, p_{C,o,1})$ and $\mathcal{L}(p_C|\rho_{S,o}, p_{C,o,2})$ at their point of intersection. For example, by piecing together $\mathcal{L}(p_C|\rho_{S,o}, p_{C,o,1})$ with $p_{C,o,1} = 0.65$ and $\mathcal{L}(p_C|\rho_{S,o}, p_{C,o})$ with $p_{C,o} = 0.80$ would improve substantially the approximation by the linear margin $\mathcal{L}(p_C|\rho_{S,o}, p_{C,o})$ alone as illustrated in Fig. 13.5. This would further improve the type I error control for $p_C < p_{C,o}$.

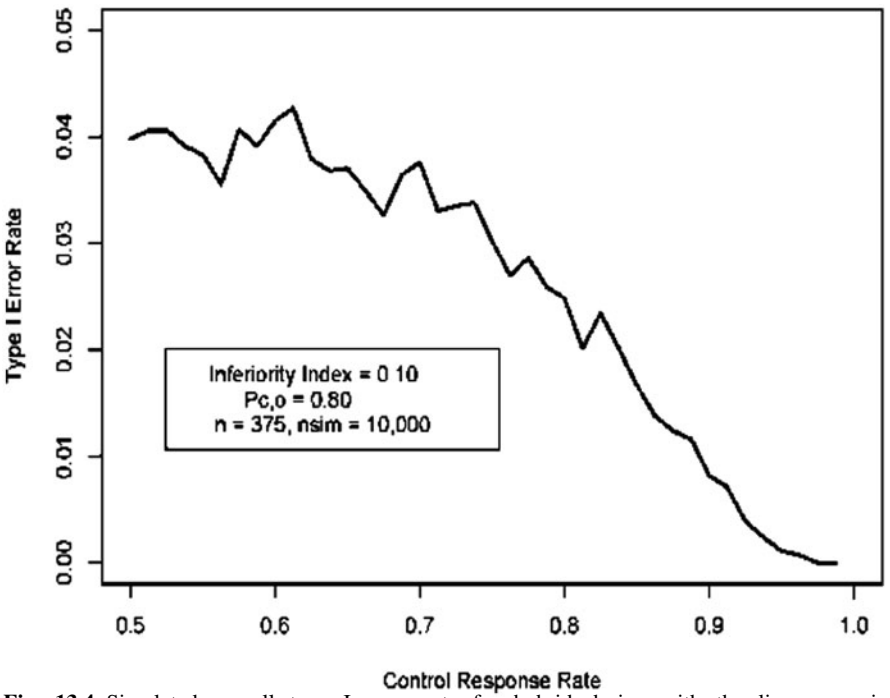

Fig. 13.4 Simulated overall type I error rate for hybrid design with the linear margin $\mathcal{L}(p_C|\rho_{S,o}, p_{C,o})$, where $\rho_{S,o} = 0.10$, $P_{c,o} = 0.80$, and $\alpha = 0.020$

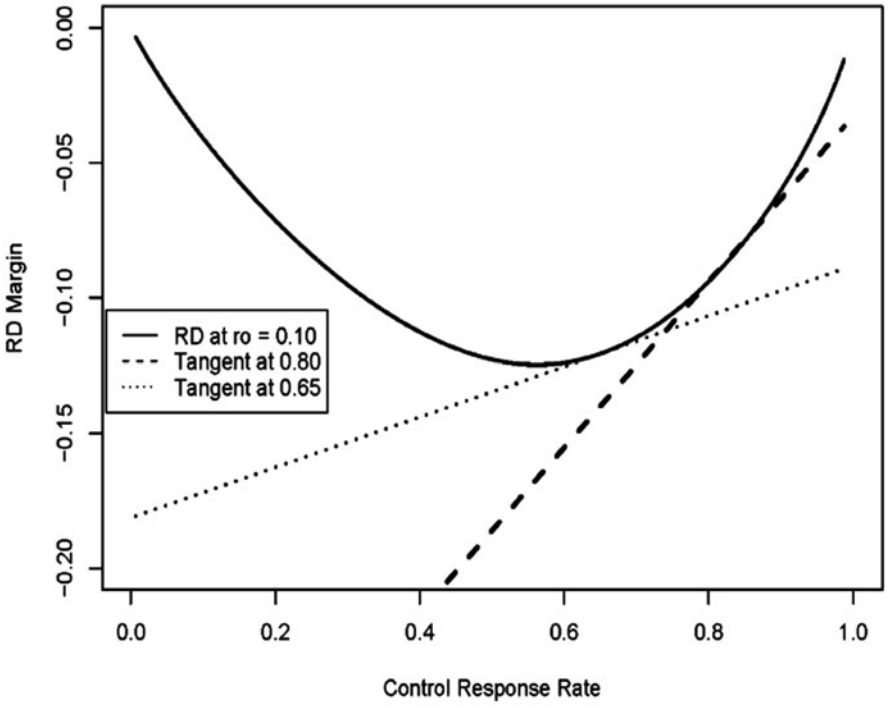

Fig. 13.5 Constructing piecewise linear margin for hybrid design with linear margin $\mathcal{L}(p_C|\rho_{S,o}, p_{C,o})$ expanded around $p_{C,o} = 0.65$ and 0.80. RD rate difference

13.4.4.2 Power Function

To derive the power function for the test statistic $\widehat{T}_{RD,o}$, one notes that under the specific alternative hypothesis $H_{sa}: \delta_{RD}(p_C) \equiv 0$, it follows from Eq. 13.35 that

$$E(\widehat{\Delta}_{RD}|H_{sa}) = E[\widehat{\delta}_{RD} - \mathcal{L}(\widehat{p}_C|\rho_{S,o}, p_{C,o})|H_{sa}] = -\mathcal{L}(p_C|\rho_{S,o}, p_{C,o}). \quad (13.40)$$

Now, let

$$\begin{aligned}\widehat{\Delta}_{RD,a} &= \widehat{\Delta}_{RD} - E\left(\widehat{\Delta}_{RD}|H_{sa}\right) = \widehat{\Delta}_{RD} + \mathcal{L}\left(p_C|\rho_{S,o}, p_{C,o}\right) \\ &= \left[\widehat{\delta}_{RD} - \frac{\partial g_{RD}^{*-1}(\rho_{S,o}, p_{C,o})}{\partial p_C}(\widehat{p}_C - p_C)\right]. \end{aligned} \quad (13.41)$$

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

$$\Sigma(p_C|H_{sa}) = \left[2 + 2\left(\frac{\partial g_{RD}^{*-1}(\rho_{S,o}, p_{C,o})}{\partial p_C}\right) + \left(\frac{\partial g_{RD}^{*-1}(\rho_{S,o}, p_{C,o})}{\partial p_C}\right)^2\right] p_C(1 - p_C). \quad (13.42)$$

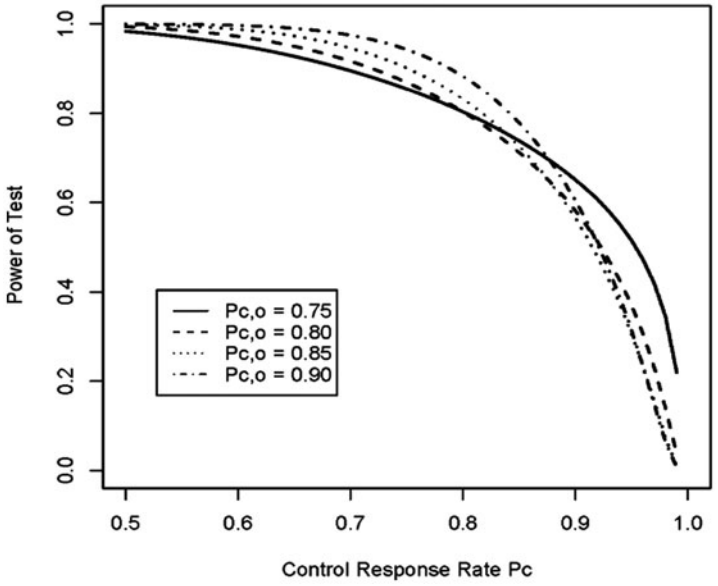

Fig. 13.6 Power functions for hybrid design with linear margin $\mathcal{L}(p_C|\rho_{S,o}, p_{C,o})$ at expansion points $p_{C,o} = 0.75$, 0.80, 0.85, and 0.90 and $n = 386$, $\alpha = 0.025$

Thus,

$$\widehat{T}^{HB}_{RD,sa} = \frac{\sqrt{n}\,\widehat{\Delta}_{RD,\,a}}{\sqrt{\Sigma(p_C|H_{sa})}} \sim N(0,1). \tag{13.43}$$

Now, from Eqs. 13.32 and 13.41, one has

$$\widehat{T}^{HB}_{RD,o} = \widehat{T}^{HB}_{RD,sa}\frac{\sqrt{\Sigma(p_C|H_{sa})}}{\sqrt{\Sigma(p_C|H_o)}} - \frac{\sqrt{n}[\mathcal{L}(p_C|\rho_{S,o}, p_{C,o})]}{\sqrt{\Sigma(p_C|H_o)}}. \tag{13.44}$$

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

$$1 - \beta = 1 - \Phi\left(\frac{1.96\sqrt{\Sigma\,(p_C|H_o)} + \sqrt{n}\mathcal{L}(p_C|\rho_{S,o}, p_{C,o})}{\sqrt{\Sigma\,(p_C|H_{sa})}}\right). \tag{13.45}$$

Now for the power function plot in Fig. 13.6, $n = 386$ was selected because it corresponds to an 80 % power for the hybrid design with a linear margin $\mathcal{L}(p_C|\rho_{S,o}, p_{C,o})$, where $\rho_{S,o} = 0.10$ and $p_{C,o} = 0.80$. Similarly, for Fig. 13.7, $n = 516$ corresponds to a 90 % power for the hybrid design with a linear margin $\mathcal{L}(p_C|\rho_{S,o}, p_{C,o})$, where $\rho_{S,o} = 0.10$ and $p_{C,o} = 0.80$. Both power plots in Figs. 13.6 and 13.7 show that the power drops off quickly when $p_C > p_{C,o}$ due to the dramatic change in variance as $p_C \to 1$. The deflation in type I error rate for $p_C > p_{C,o}$ might be a desirable feature since it raises a natural barrier to prevent ejection of the inferiority null of Eq. 13.33 or 13.34 when the true control response rate p_C is much greater than the assumed control response rate $p_{C,o}$.

The powers for selected true control response rate p_C in the plots in Figs. 13.6 and 13.7 are given in Table 13.3.

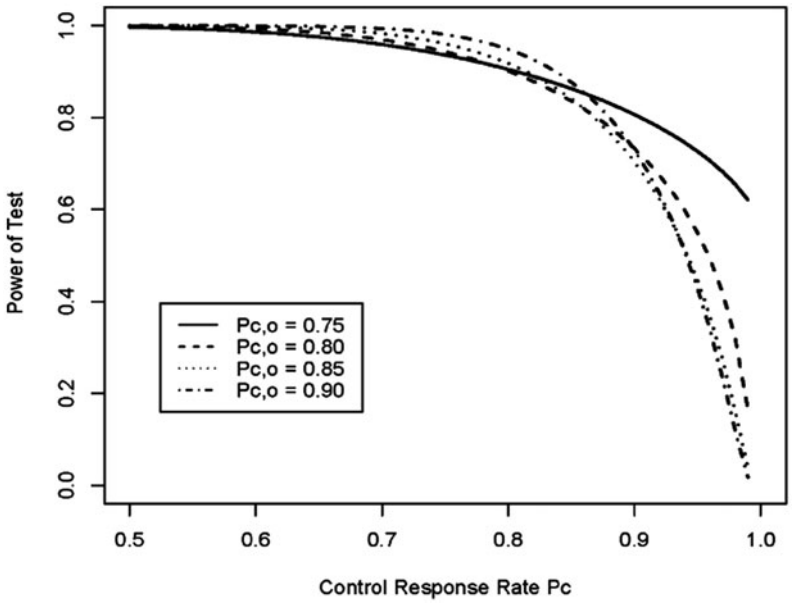

Fig. 13.7 Power functions for hybrid design with linear margin $\mathcal{L}(p_C|\rho_{S,o}, p_{C,o})$ at expansion points $p_{C,o} = 0.75$, 0.80, 0.85, and 0.90 and $n = 516$, $\alpha = 0.025$

Table 13.3 Pointwise power across p_C at a significance level of $\alpha = 0.025$ and different expansion point $p_{C,o}$

		Expansion point $p_{C,o}$			
Sample size for hybrid design p_C	p_C	0.75	0.80	0.85	0.90
375	0.70	0.895	0.917	0.946	0.975
(80 % Power for $p_{C,o} = 0.80$)	0.75	0.855	0.869	0.902	0.943
	0.80	0.803	0.803	0.832	0.883
	0.85	0.738	0.712	0.725	0.779
	0.90	0.651	0.581	0.564	0.604
	0.95	0.515	0.369	0.305	0.312
500	0.70	0.959	0.969	0.983	0.994
(90 % Power for $p_{C,o} = 0.80$)	0.75	0.936	0.943	0.961	0.981
	0.80	0.905	0.901	0.918	0.949
	0.85	0.863	0.836	0.840	0.877
	0.90	0.807	0.733	0.703	0.729
	0.95	0.726	0.549	0.446	0.428

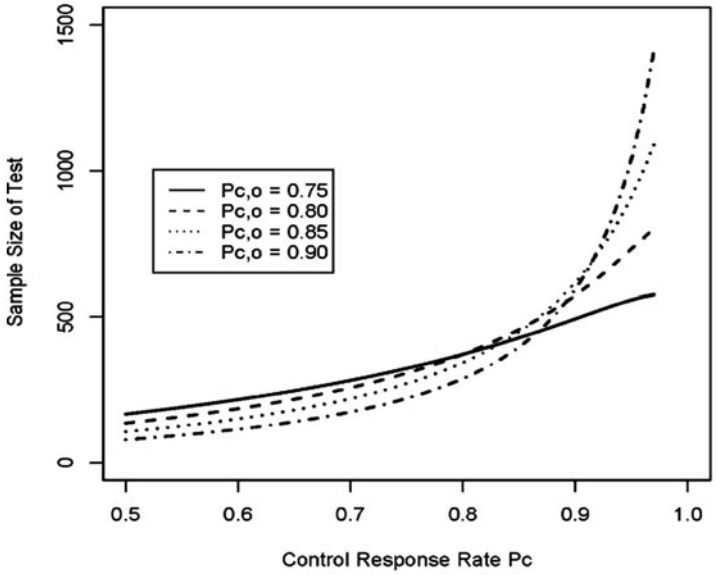

Fig. 13.8 Plots of sample size per group at $\alpha = 0.025$ and 80 % power for hybrid design with linear margin $\mathcal{L}(p_C|\rho_{S,o}, p_{C,o})$ at $\rho_{S,o} = 0.10$ and expansion points $p_{C,o} = 0.75$, 0.80, 0.85, and 0.90

13.4.4.3 Sample Size Calculation

From Eq. 13.44, the sample size formula is derived by setting

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

Solving for n, one obtains,

$$n = \frac{(1.96\sqrt{\Sigma(p_C|H_o)} + z_{1-\beta}\sqrt{\Sigma(p_C|H_{sa})})^2}{\mathcal{L}^2(p_C|\rho_{S,o}, p_{C,o})}. \tag{13.46}$$

Figures 13.8 and 13.9 display the sample size plots for the hybrid design with linear margin $\mathcal{L}(p_C|\rho_{S,o}, p_{C,o})$ at the expansion points $p_{C,o} = 0.75$, 0.80, 0.85, and 0.90 (Table 13.4).

As discussed in Sect. 13.4.3.1, if one also wishes to control the type I error rate at $p_C = p_{C,o}$ at $\alpha = 0.025$, then one needs to increase the sample size accordingly. Table 13.5 shows the sample size needed for such adjustment.

13.4.5 An Application to the Design of HABP/VABP Trials

The FDA Anti-infective Advisory Committee convened in November 2011 to discuss issues related to the design of NI trials for HABP and VABP [US FDA (2011)]. In

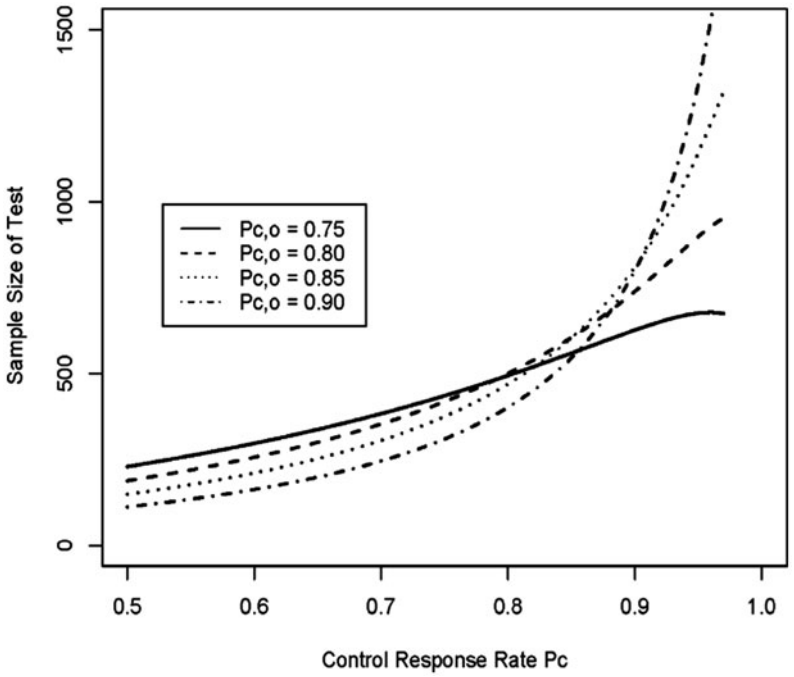

Fig. 13.9 Plots of sample size per group at $\alpha = 0.025$ and 90 % power for hybrid design with linear margin $\mathcal{L}(p_C|\rho_{S,o}, p_{C,o})$ at $\rho_{S,o} = 0.10$ and expansion points $p_{C,o} = 0.75$, 0.80, 0.85, and 0.90

Table 13.4 Selected sample size per group at $\alpha = 0.025$ for hybrid design with linear margin $\mathcal{L}(p_C|\rho_{S,o}, p_{C,o})$ at $\rho_{S,o} = 0.10$ and expansion points $p_{C,o} = 0.75$, 0.80, 0.85, and 0.90

	Power	
Taylor expansion point $P_{C,o}$	80 %	90 %
0.75	323	435
0.80	386	516
0.85	450	605
0.90	593	798

Table 13.5 Unadjusted and adjusted sample size per group for hybrid design with Taylor expansion at the point $p_{C,o} = 0.80$

	Significance level	
Power	0.025	0.020
0.80	386	400
0.90	516	529

the briefing book provided to the Committee members, FDA presented the following data based on two historical placebo-controlled studies and five recent active control studies (Table 13.6).

The estimated control survival rate is equal to 80 %. An estimate of the CE is given by $\text{CE} = [(0.52 - 0.23) - 0.09] = 0.20$, which is obtained by taking the difference

Table 13.6 Estimated mortality rates and 95 % confidence intervals

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

Table 13.7 Sample size calculated by FDA for the rate difference measure δ_{RD}

Power	HABP	VABP
80 %	834	714
90 %	1114	894

HABP hospital-acquired bacterial pneumonia, *VABP* ventilation-associated bacterial pneumonia

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 was then set at $\delta_{RD,\,o} = -\text{CE} \times \frac{1}{2} = -0.20 \times \frac{1}{2} = -0.10$, where the fraction of one half is based on clinical judgment regarding the size of the margin. FDA posed to the Committee several questions, including the following: What margin should one use in the event the control survival rate from the NI trial appears to deviate from the estimated control survival rate of 80 %?

For the RD measure δ_{RD}, FDA calculated the sample sizes required for 80 and 90 % power at a significance level of 0.025 after an adjustment of 60 %/70 % microbiologic evaluability rate for HABP/VABP trials, respectively, as given in Table 13.7.

Thus, with the given fixed margin of $\delta_{RD,o} = -0.10$ at an estimated survival rate of $p_{C,o} = 0.80$ (equivalent to a 20 % mortality rate), the degree of stringency for the empirically derived pair $(\delta_{RD,o},\ p_{C,o}) = (-0.10, 0.80)$ can be assessed using the standard index function in Eq. 13.22 and is equal to $\rho_{S,o} = g^*_{RD}(\delta_{RD,o}, p_{C,o}) = g^*_{RD}(-0.10,\ 0.80) = 0.1057$. Now, for simplicity of discussion, consider rounding it to 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 $\delta_{RD,o} = -0.0939$ instead of the margin of $\delta_{RD,o} = -0.10$, at $p_{C,o} = 0.80$. Now upon setting the inferiority index level to $\rho_{S,o} = 0.10$ in the margin function given by Eq. 13.24, one obtains the special indexed margin function $\delta_{RD}(p_C|0.10) = g^{*-1}_{RD}(0.10,\ p_C)$ with the degree of stringency specified by $\rho_{S,o} = 0.10$. After applying the Taylor expansion around the point $p_{C,o} = 0.80$, one finds the linear margin function is equal to

$$\begin{aligned}\mathcal{L}(p_C|\rho_{S,o},\ p_{C,o}) &= \mathcal{L}(p_C|0.10,\ 0.80) = \delta_{RD}(p_C|0.10)\\ &\quad + \frac{\partial g^{*-1}_{RD}}{\partial p_C}(0.10,\ 0.80)(p_C - 0.80)\\ &= -0.0939 + \ 0.3066\,(p_C - 0.80).\end{aligned}$$

The hybrid NI hypothesis is then defined by

$$H_o: \ \delta_{RD} - [-0.0939 \ + \ 0.3066\,(p_C - 0.80)] \leq 0$$
$$vs.$$
$$H_a: \ \delta_{RD} - [-0.0939 \ + \ 0.3066\,(p_C - 0.80)] > 0. \qquad (13.47)$$

Based on the hybrid design that has just been discussed in Sect. 13.4, to test the hybrid NI hypothesis (Eq. 13.47) at the expansion point $p_{C,o} = 0.80$ with a significance level of $\alpha = 0.025$ and a power of 80 %, a sample size of $n = 386$ subjects per group would be needed (see Table 13.3, Table 13.4 or Table 13.5). Now the sample size per group needed for an HABP/VABP trial is given by 643/551, reflecting an adjustment for a 60 %/70 % microbiologic evaluability rate, or for a total sample size of 1286/1102. On the other hand, for the fixed margin NI hypothesis, the sample size per group is $n = 283$. After adjusting for 60 %/70 % microbiologic evaluability rate, this gives rise to a sample size per group of 472/404 or a total sample size of 944/809 for HABP/VABP trials (see Chi and Koch 2012), reflecting a 36.2 %/36.4 % increase.

Thus, one can see that at a significance level of $\alpha = 0.025$ and a power of 80 %, the flexibility realized in a hybrid NI design with a linear margin $\mathcal{L}(p_C|\rho_{S,o},\ p_{C,o})$ derived at the empirically based inferiority index value of $\rho_{S,o} = 0.10$ and the expansion point $p_{C,o} = 0.80$, which is the estimated control response rate, is gained at the cost of about a 36 % increase in the size over that required for a corresponding fixed margin NI design.

Now the hybrid design with its NI hypothesis given by Eq. 13.33 or 13.34 has a linear 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 Eq. 13.33 or Eq. 13.34 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 $\delta_{RD}(0.90|\rho_{S,o}) = g_{RD}^{*-1}(0.10,\ 0.90) = -0.0601$ based on the margin function in Eq. 13.24 as shown in Table 13.1. 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 inflation in the type I error rate despite the adjustment. Without adjustment by lowering the nominal significance level from $\alpha = 0.025$, one may consider a better alternative discussed earlier by 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 indexed 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 $\mathcal{L}(p_C|\rho_{S,o}, p_{C,o})$. Phillips (2003) has actually constructed piecewise linear margin based on consensus opinions of clinical experts. It is not linked to any index function and is unrelated to the piecewise linear margin as discussed in this chapter. The theoretical properties of an NI design with a piecewise linear margin has been investigated by Zhang (2006) for the likelihood ratio test. The method developed there may be applicable to the hybrid

design with a piecewise linear margin. It should be of practical interest to investigate this matter further for the RD measure along the line as suggested in Sect. 13.5.

13.5 Summary Discussion

At the November 2011 FDA Anti-Infective Advisory Committee (US FDA 2011) meeting discussing the design of HABP or VABP trials, the agency posed several questions to the Committee. This chapter attempts to address two of the questions. The first question concerns the appropriateness of the empirically derived fixed margin associated with an estimated control response rate using FDA's two-step procedure. The second question pertains to what margin one should use when the expected control response rate p_C from the NI trial appears to deviate from the estimated 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 chapter is defined by a special linear margin, $\mathcal{L}(p_C|\rho_{S,o}, p_{C,o})$, which is the first-order Taylor expansion of a specific indexed margin function around the estimated control response rate $p_{C,o}$. The specific indexed margin function is defined as follows. First, derive the index value $\rho_{S,o} = g^*_{RD}(\delta_{RD,o}, p_{C,o})$ from the index function given in Eq. 13.22 at the empirically derived pair $(\delta_{RD,o}, p_{C,o})$. Therefore, $\rho_{S,o}$ is an empirically derived inferiority index value. Next, set the index ρ_S in the margin function $\delta_{RD} = g^{*-1}_{RD}(\rho_S, p_C)$ given in Eq. 13.23 equal to this empirically derived index value $\rho_{S,o}$ which defines the specific margin function $\delta_{RD}(p_C|\rho_{S,o}) = g^{*-1}_{RD}(\rho_{S,o}, p_C)$ given by Eq. 13.24. Now define the linear margin given by $\mathcal{L}(p_C|\rho_{S,o}, p_{C,o}) = \delta_{RD}(p_{C,o}|\rho_{S,o}) + \frac{\partial g^{*-1}_{RD}(\rho_{S,o},p_C)}{\partial p_C}(p_C - p_{C,o})$ as the first-order Taylor approximation to the margin function $\delta_{RD}(p_C|\rho_{S,o})$ expanded around $p_C = p_{C,o}$. Clearly, $\mathcal{L}(p_{C,o}|\rho_{S,o}, p_{C,o}) = \delta_{RD}(p_{C,o}|\rho_{S,o}) = \delta_{RD,o}$ when $p_C = p_{C,o}$. 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 $\delta_{RD}(p_{C,o}|\rho_{S,o}) = \delta_{RD,o}$ by the quantity $\frac{\partial g^{*-1}_{RD}(\rho_{S,o},p_C)}{\partial p_C}(p_C - p_{C,o})$. Hence, the linear margin integrates the empirically derived pair $\delta_{RD,o}, p_{C,o})$ with a variable component that adjusts for the deviation $(p_C - p_{C,o})$. Thus, the NI hypothesis defined by such a linear margin is called a hybrid design. Such a hybrid design conveys the stringency of the margin through the empirically derived index value $\rho_{S,o}$ and at the same time also has the flexibility to adjust for the margin in the event the control response rate from the trial deviates from the estimated control response rate $p_{C,o}$. Of course, this flexibility of a hybrid design is gained at the cost of a 33 % increase in sample size compared to that required for a fixed margin design for the example considered.

The linear margin $\mathcal{L}(p_C|\rho_{S,o}, p_{C,o})$ tends to be more liberal if the true control response rate is in the range of $0.50 < p_C < p_{C,o}$. For example, with $(\delta_{RD,o}, p_{C,o}) = (-0.10, 0.80)$, the linear margin closely approximates the margin function $\delta_{RD}(p_C|\rho_{S,o})$ only for p_C over the range (0.75, 0.90). One may try to

minimize this type I error rate inflation by lowering the overall significance level. But this approach would be too drastic and still would not fully resolve the problem. An alternative strategy is to construct a piecewise linear margin by joining two linear margins $\mathcal{L}(p_C|\rho_{S,o}, p_{C,1})$ and $\mathcal{L}(p_C|\rho_{S,o}, p_{C,o})$ at their point of intersection, where $0.50 < p_{C,1} < p_{C,o}$ along the line of Phillips (2003) and Zhang (2006) who demonstrated the asymptotic convergence of the likelihood ratio test statistic for the NI hypothesis defined by a piecewise linear margin. However, an even better strategy is to define a spline function that joins the two linear margins by smoothing out the corner where they intersect [Reinsch (1967), Byrne and Chi (1972), De Boor (2001)]. Such a spline margin would have the regularity property required for the estimate as well as for the asymptotic convergence of the test statistic associated with such hybrid NI hypothesis. The idea of a hybrid design with an empirically determined spline margin based on the special indexed margin function $\delta_{RD}(p_C|\rho_{S,o})$ deserves further investigation because it can provide margins closely matching those from the margin function $\delta_{RD}(p_C|\rho_{S,o})$ throughout the interval (0.50, 1) and has sufficient regularity properties required for the convergence theorem to hold. All of these hybrid designs are of special appeal because they integrate the FDA's two-step fixed margin approach with the flexibility of a variable margin and their associated test statistics have reasonable performance characteristics by taking advantage of the improvement made by the convergence theorem (Theorem 2) under fixed margin for the RD measure. However, these positive attributes come with a 36 % increase in sample size over those required under a fixed margin NI hypothesis.

Looking beyond binary outcome trials, some of the ideas used in this chapter can be extended to normal distributions to provide a natural framework for handling problems involving heterogeneity of variance, such as in establishing bioequivalence of highly variable drugs. Unlike the case under Bernoulli distributions, where for a given index value, the margin is simply a function of the control response rate p_C, under normal distributions, for a given index value, the margin would be a function of the variance of the control σ_C and the variance ratio σ^2 when not assumed to be known.

Post Note: In this chapter, the author has corrected an error that appeared in the original paper by Li and Chi (2011). Specifically, in Eq. 13.7 on page 293 of the Li and Chi (2011) paper, the number "4" appearing in the denominator of the third term should be replaced by the number "2" as shown in Eq. 13.2 in the present chapter. This correction has no impact under normal distributions. But under Bernoulli distributions, the impact of this correction is to increase the variance S in Corollary 2 of Li and Chi (2011) on page 298 by an amount $\frac{\delta^2(\rho,\sigma)}{1+\sigma^2}$ and thus the variance there should be $S = \left\{1 + \frac{\delta^2(\rho,\sigma)}{16\sigma_C^2\sigma^2}\right\} + \frac{\delta^2(\rho,\sigma)}{1+\sigma^2}$. It should also be pointed out that at the end of this same corollary, the variance estimate $\widehat{\sigma}_C^2$ is missing by a factor of ½ and it should be given by $\widehat{\sigma}_C^2 = \frac{1}{2}\left[\frac{\widehat{p}_T(1-\widehat{p}_T)}{\sigma^2} + \widehat{p}_C(1-\widehat{p}_C)\right]$.

This same error also appears in Chi and Koch (2012). Specifically, at the end of Theorem 2 of Chi and Koch (2012), the variance $\Sigma^2_{SRD,\,o}$ should be given by $\Sigma^2_{SRD,\,o} = \left(1+\sigma_o^2\right)\left(1 + \frac{\delta^2_{SRD,\,o}}{16\sigma_C^2\sigma_o^2}\right) + \frac{\delta^2_{SRD,\,o}}{2}$ as given in Eq. 13.4 of the present chapter.

Hence, it follows that Eq. 13.24 in Theorem 4 of Chi and Koch (2012) should be replaced by $\Sigma^2_{RD,o} = \left[\left(\sigma^2_{C,o} + \sigma^2_{T,o}\right)\left(1 + \frac{\delta^2_{RD,o}}{16\sigma^2_{C,o}\sigma^2_{T,o}}\right) + \frac{\delta^2_{RD,o}}{2}\right] - (1 - 2p_{C,o})\delta_{RD,o}$ which is given by Eq. 13.8 of the present chapter. In addition, in Theorem 4 of Chi and Koch (2012), in the expression for the variance $\Sigma^2_{LOR,o}$ given by Eq. 13.26, the variance $\Sigma^2_{SRD,\,o}$ in the first term should be as given above which is given by Eq. 13.4 of the present chapter.

Similarly, the same error appears in Chi (2013). In Theorem 1 of Chi (2012), the variance term $\Sigma^2_{RD,o}(p_{C,o}|H_o)$ in Eq. 13.15 should be replaced by $\Sigma^2_{RD,o}\left(p_{C,o}|H_o\right) = \left[\left(\sigma^2_{C,o} + \sigma^2_{T,o}\right)\left(1 + \frac{\delta^2_{RD,o}}{16\sigma^2_{C,o}\sigma^2_{T,o}}\right) + \frac{\delta^2_{RD,o}}{2}\right] - (1 - 2p_{C,o})\delta_{RD,o}$ which is given by Eq. 13.8 of this chapter.

References

Berry AC (1941) The accuracy of the Gaussian approximation to the sum of independent variates. Trans Am Math Soc 49:122–136

Bristol DR (1996) Determining equivalence and the impact of sample size in anti-infective studies: a point to consider. J Biopharm Stat 6(3):319–326

Byrne GD, Chi DNH (1972) Linear multistep formulas based on g-splines. SIAM J Numerical Anal 9:316–324

Chi GYH (2012) Inferiority index and the Behrens-Fisher problem for non-inferiority trials. In: 2012 Proceedings of the American Statistical Association, Biopharmaceutical Section, p 776–784

Chi GYH (2013) A hybrid design for non-inferiority trials with binary outcomes. In: 2013 Proceedings of the American Statistical Association, Biopharmaceutical Section, p 3728–3742

Chi GYH, Koch GG (2012) Inferiority index and margin function for non-inferiority trials with binary outcomes. In: 2013 Proceedings of the American Statistical Association, September 2012 FDA-Industry Workshop, p 4451–4465

de Boor C (2001) A practical guide to splines. Applied mathematical sciences series, vol. 27. New York, Springer-Verlag

Esseen CG (1942) On the Liapunoff limit of error in the theory of probability. Ark Mat Astron Fys A28:1–19

Garrett AD (2003) Therapeutic equivalence: fallacies and falsification. Stat Med 22:741–762

Kim MY, 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. J Clin Epidemiol 57:1253–1261

Li G, Chi GYH (2011) Inferiority index and margin in non-inferiority trials. Stat Biopharm Res 3(2):288–301

Munk A, Skipka B, Stratmann B (2005) Testing general hypotheses under binomial sampling: the two sample case—asymptotic theory and exact procedures. Comput Stat Data Anal 49:723–739

Phillips KF (2003) A new test of non-inferiority for anti-infective trials. Stat Med 22:201–212

Reinsch CH (1967) Smoothing by spline functions. Numerische Mathematik 10:177–183

Röhmel J (1998) Therapeutic equivalence investigations: statistical considerations. Stat Med 17:1703–1714

Röhmel J (2001) Statistical considerations of FDA and CPMP rules for the investigation of new anti-bacterial products. Stat Med 20:2561–2571

Senn S (2000) Consensus and controversy in pharmaceutical statistics (with discussion). J Roy Stat Soc Ser. 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. J Biopharm Stat 8(2):263–282

US Food and Drug Administration (1992) Anti-infective points to consider guidance, Office of Communication/Division of Drug Information/CDER/FDA, W051, Room 2001, 10903 New Hampshire Avenue, Silver Spring, MD, 20993

US Food and Drug Administration (2010) Guidance to industry: non-inferiority clinical trials, Office of Communication/Division of Drug Information/CDER/FDA, W051, Room 2001, 10903 New Hampshire Avenue, Silver Spring, MD, 20993

US Food and Drug Administration (2011) Anti-infective drugs advisory committee meeting materials, 3–4 Nov 2011, http://www.fda.gov/AdvisoryCommittees/CommitteesMeetingMaterials/Drugs/Anti-InfectiveDrugsAdvisoryCommittee/ucm242307.htm. Accessed 7 Feb 2013

U.S. Food and Drug Administration (2013) 'IND content and format', 21 CFR 312.23 (a)(6)(ii), www.accessdata.fda.gov/scripts/cdrh/cfdocs/cfcfr/cfrsearch.cfm?fr†=312.23. Accessed 3 Jan 2014

Weng CSW, Liu JP (1994) Some pitfalls in sample size estimation for an anti-infective study. In: 1994 Proceedings of the Pharmaceutical Section of the ASA, p 56–60

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