Journal of Biometrics & Biostatistics
ISSN: 2155-6180

Chow and Song, J Biom Biostat 2016, 7:3
DOI: 10.4172/2155-6180.1000301

Research Article Open Access

On Selection of Margin in Non-Inferiority Trials

Shein-Chung Chow[1]* and Fuyu Song[2]

[1]Duke University School of Medicine, Durham, North Carolina, USA
[2]Peking University Clinical Research Institute, Peking University Health Science Center, Beijing, China

Abstract

In clinical trials, it is unethical to use a placebo control in treating patients with severe or life-threatening diseases such as cancer when approved and effective therapies (e.g., standard of care or active control agents) are available. Alternatively, an active control or non-inferiority trial is often considered. In practice, one of the key issues for a non-inferiority trial is the determination of non-inferiority margin which has an impact on the power analysis for sample size calculation. In its 2010 draft guidance, the United States Food and Drug Administration (FDA) recommend a couple of margins for testing non-inferiority of a test treatment as compared to an active control agent or a standard of care treatment. In this article, several margins, which not only adjust for variability associated with the observed data but also take into consideration of the retention rate of the treatment effect, are proposed.

Keywords: Active control trials; Non-inferiority margin; Non-inferiority hypothesis; Confidence lower bound; Retention rate.

Introduction

In clinical trials, it is unethical to treat patients with critical/severe and/or life-threatening diseases such as cancer when approved and effective therapies such as standard of care or active control agents are available. In this case, an active control or non-inferiority trial comparing a test treatment with an active control agent or a standard of care treatment is often considered. The ultimate goal of a non-inferiority trial is to establishment non-inferiority of the test treatment by demonstrating that the test treatment is not inferior to (or at least as effective as) the active control agent or the standard of care treatment.

In practice, there may be a need to develop a new treatment or therapy that is non-inferior (but not necessarily superior) to an established efficacious treatment due to the following reasons: (i) the test treatment is less toxic, (ii) the test treatment has a better safety profile, (iii) test treatment is easy to administer, (iv) the test treatment is less expensive, (v) the test treatment has better quality of life, (vi) test treatment provides alternative treatment with some clinical benefits, e.g., generics or biosimilars. Clinical trials of this kind are referred to as non-inferiority trials. A comprehensive overview of design concepts and important issues that are commonly encountered in active control or non-inferiority trials can be found in D'Agostino et al. [1].

For testing non-inferiority, we typically reject the null hypothesis of inferiority that the difference between the test treatment and the active control agent or the standard of care treatment is greater than a pre-specified non-inferiority margin (a clinically meaningful difference) and then conclude non-inferiority of the test treatment as compared to the active control agent or the standard of care treatment. The test treatment can then serve as an alternative to the active control agent or the standard of care treatment. In practice, it, however, should be noted that unlike equivalence testing, non-inferiority testing is a one-sided equivalence testing which consists of the concepts of equivalence and superiority. In other words, superiority may be tested after the non-inferiority has been established. We conclude equivalence if fail to reject the null hypothesis of non-superiority. On the other hand, superiority may be concluded if the null hypothesis of non-superiority is rejected.

In practice, one of the key issues in a non-inferiority trial is the selection of an appropriate non-inferiority margin. The selected non-inferiority margin is very sensitive to power calculation for sample size. Different non-inferiority margins may lead to very different sample sizes required for achieving a desired power for establishing non-inferiority of the test treatment. In practice, despite the existence of some studies [2-5], there is no established rule or gold standard for determination of non-inferiority margins in active control trials until early 2000. In 2000, the International Conference on Harmonization (ICH) published a guideline on *Choice of Control Group and Related Design and Conduct Issues in Clinical Trials* to assist the sponsors for selection of an appropriate non-inferiority margin [6]. As indicated in the ICH E10 guideline, non-inferiority margins may be selected based on past experience in placebo control trials under similar conditions to the new trial. ICH E10 also pointed out that the selection of a non-inferiority should be suitably conservative and reflect uncertainties in the evidence on which the choice is based on. In 2010, the United States Food and Drug Administration (FDA) also published Draft Guidance on *Non-inferiority Clinical Trials* [7]. The 2010 FDA guidance recommends a couple of approaches for selection of non-inferiority margins, namely, M_1 and M_2.

In addition to the FDA's recommended approaches, the purpose of this article is to propose alternative methods for selection of non-inferiority margin in non-inferiority trials. In the next section, the relationship between non-inferiority testing and equivalence testing in active control trials is briefly described. Also included in this section is the impact on sample size requirement for achieving a desired power for establishment of non-inferiority. Section 3 discusses regulatory requirements and non-inferiority hypothesis for a pre-specified non-inferiority margin. Various methods for selection of non-inferiority margins are reviewed in Section 4. An example is given in Section 5 to illustrate various methods for determination of non-inferiority margin. Brief concluding remarks are given in the last section.

***Corresponding author:** Shein-Chung Chow, Professor, Biostatistics and Bioinformatics, Duke University School of Medicine, 2424 Erwin Road, Hock Suite 1102, Room 11068, Durham, North Carolina, USA, Tel: (919) 668-7523; E-mail: sheinchung.chow@duke.edu

Received April 14, 2016; **Accepted** April 28, 2016; **Published** May 05, 2016

Citation: Chow SC, Song F (2016) On Selection of Margin in Non-Inferiority Trials. J Biom Biostat 7: 301. doi:10.4172/2155-6180.1000301

Non-inferiority versus equivalence

In clinical trials, some investigators often mix up with the concepts of non-inferiority testing and equivalence testing. Thus, the question whether an equivalence testing can be replaced with a non-inferiority (especially for equivalence testing for bio similar products) has been raised and discussed. In this section, we will explore the relationship among non-inferiority, equivalence, and superiority, their corresponding testing hypotheses, and power calculations for sample size.

Relationship among non-inferiority, equivalence, and superiority

To study the relationship among non-inferiority, equivalence and superiority, we first assume that the non-inferiority margin, equivalence limit, and superiority margin are the same. Let M denote the non-inferiority margin (also equivalence limit and superiority margin). Also, let μ_T and μ_s be the mean responses of the test treatment and standard therapy (active control agent), respectively. If we assume that an observed mean response on the right hand side of μ_s is an indication of improvement, then the relationship among non-inferiority, equivalence and superiority is illustrated in Figure 1.

From Figure 1, if μ_T falls within the equivalence limit of, (μ_s – M, μ_s + M), we conclude that the test treatment is equivalent to the active control agent or the standard of care treatment. Consider the left hand side, i.e., $\mu_s < \mu_s$ – M. In this case, the test treatment is considered inferior to the active control agent or the standard of care treatment. Thus, μ_s – M ≤ μ_T is an indication that the test treatment is not inferior to the active control agent. When μ_s – M ≤ μ_T (i.e., μ_T is on the right side of μ_s – M. In this case, the test treatment could be either equivalent to the active control agent if $\mu_T < \mu_s$ + M or superior to the active control agent if μ_s + M < μ_T. Thus, we could test for superiority once the non-inferiority has been established without paying any statistical penalty because it is a closed testing procedure. Thus, non-inferiority consists of the concepts of equivalence and superiority and equivalence can be established through testing for non-inferiority and testing for non-superiority. Both non-inferiority testing and superiority testing are considered one-sided equivalence testing.

To provide a better understanding of the relationship among non-inferiority, equivalence, and superiority, their corresponding hypotheses are given in Figure 2.

It, however, should be noted that if an observed response less than μ_s is considered improvement, the hypotheses for testing non-inferiority and superiority need to be modified.

Impact on sample size requirement

As indicated earlier, one of the major issues in non-inferiority trials is the selection of an appropriate non-inferiority margin for achieving a desired power for establishing non-inferiority. Let μ_T and μ_s be the mean responses for the test treatment and the standard of care treatment or the active control agent, respectively. Also let M be the non-inferiority margin. For illustration purpose, we will focus on the study endpoint of binary responses. Based on formulas provided in Chow, Shao, and Wang [8], sample size requirement for testing non-inferiority or equivalence for achieving an 80% power at the 5% level of significance for various combinations of μ_T and μ_s are summarized in Table 1.

As it can be seen from Table 1, testing non-inferiority (which is one-sided equivalence testing) requires less subjects. It, however, should be noted FDA requires non-inferiority testing be performed based on one side of a two-sided test at the 5% level of significance, which is equivalent to a one-sided test at the 5% level of significance. In other words, FDA requires the significance level of 2.5% should be used when performing a one-sided non-inferiority testing. In this case, we need to increase sample size in order for achieving the same level of power for establishing non-inferiority. Also, Table 1 indicates that a narrower margin requires a much larger sample size for achieving the desired power for establishing non-inferiority. As a result, the selection of non-inferiority margin in very critical in non-inferiority trials.

Non-Inferiority Hypothesis

Regulatory requirements

International conference on harmonization (ICH) guideline: For selection of non-inferiority margin, the ICH E10 Guideline suggests that the non-inferiority margin M should be chosen to satisfy at least the following two criteria:

(1). The ability to claim that the test treatment is not inferior to the active control agent and, at the same time, is superior to the placebo (even though the placebo is not included in the non-inferiority trial).

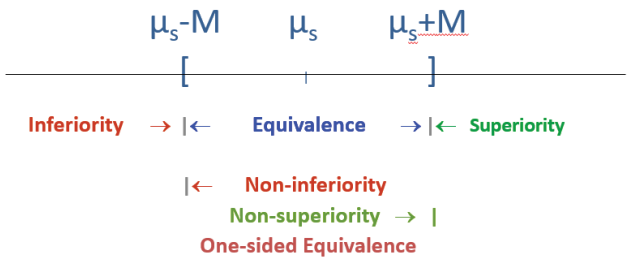

Figure 1: Relationship among non-inferiority, superiority, and equivalence.

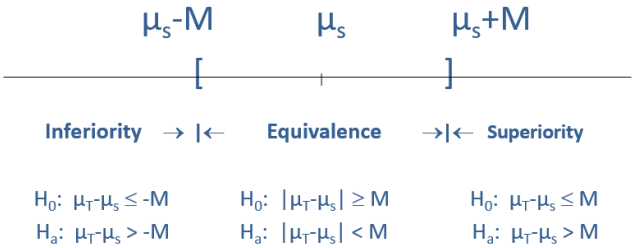

Figure 2: Hypotheses for testing non-inferiority, equivalence, and superiority.

$\mu_T=\mu_s$[2]	NI margin or Equivalence limit	Non-inferiority Testing[3]	Equivalence Testing
$\mu_T=\mu_S \geq 90\%$	8%	174 (348)	241 (482)
	10%	112 (224)	155 (310)
$80\% \leq \mu_T=\mu_S<90\%$	12%	138 (276)	191 (382)
	15%	88 (176)	122 (244)
$70\% \leq \mu_T=\mu_S<80\%$	15%	116 (232)	160 (320)
	20%	65 (130)	90 (180)

Note: [1]Power calculation was performed for achieving an 80% at the 5% level of significance

[2]$\mu_T=\mu_s$=90%, 80%, and 70% were considered in this illustration.

[3]Non-inferiority testing was performed based on one-sided test at the 5% level of Significance.

Table 1: Sample size[1] requirement for binary responses.

(2). The non-inferiority margin should be suitably conservative. That is, the margin should account for the variability associated with the response.

Note that a fixed margin, i.e., it does not depend on any parameters, is rarely suitable under (1). The ICH E10 also indicated that the selected margin should not be greater than the smallest effect size that the active drug would be reliably expected to have compared with placebo in the setting of a placebo-controlled trial.

United states food and drug administration (FDA) guidance: The 2010 FDA draft guidance basically consists of four parts: (1) a general discussion of regulatory, study design, scientific, and statistical issues associated with the use of non-inferiority studies when these are used to establish the effectiveness of a new drug, (2) details of some of the issues such as the quantitative analytical and statistical approaches used to determine the non-inferiority margin for use in non-inferiority studies, (3) Q&A of some commonly asked questions, and (4) five examples of successful and unsuccessful efforts for determining non-inferiority margins and the conduct of non-inferiority studies [7]. In principle, the 2010 FDA draft guidance is very similar to the ICH E10 guideline. However, the 2010 FDA draft guidance provides more details regarding study design and statistical issues. The 2010 FDA draft guidance recommends two approaches be considered for determination of non-inferiority margin based on historical data of the approved active control agent. These approaches will be discussed in Section 5.3.

Hypothesis setting and clinically meaningful margin

Let T, C and P denote the new or test treatment, the active protocol agent which has been demonstrated to be superior to a placebo, and the placebo, respectively. Thus, the relationship among T, C, P, and M (clinically meaningful margin) is illustrated in Figure 3a-3c.

As it can be seen from Figure 3a, if T falls within (C-M,C+M), we consider T and C are therapeutically equivalent assuming that the right side of C is improving and the left side of C is worsening. Thus, if T falls on the left hand side of C-M, i.e., T<C-M or C-T>M, we claim that T is inferior to C or C is superior to T. On the other hand, T is considered non-inferior to C if it falls on the right side of C-M, i.e., C-M<T or C-T<M. In this case, hypotheses for testing non-inferiority between T and C can be described as follows.

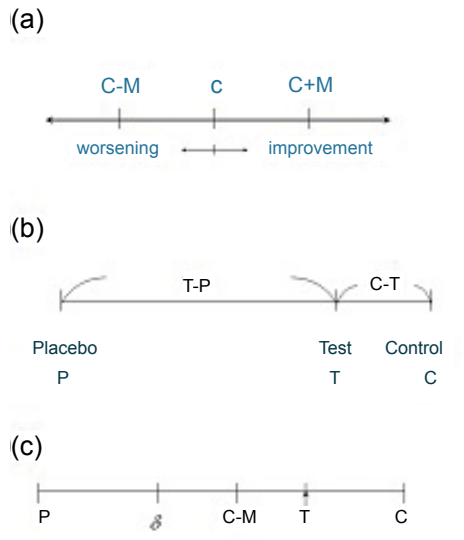

Figure 3: Relationship among T, C, P, and M.

H_0: C-T> M (or C-M>T, T is inferior to C);

H_a: C-T< M (or C-M<T, T is not inferior to C).

Thus, we would reject the null hypothesis that T is inferior to C and conclude that the difference between T and C is less than a clinically meaningful non-inferiority margin (M) and hence T (test treatment) is at least as effective as (or not worsen than) C, the active control agent.

Figure 3b and 3c describes relationship among T, C, P, and M. If T is not inferior to C and is superior to P, then (i) T > C-M or T-C > -M and (ii) T-P > δ, where M ≥ δ.

Retention of treatment effect in the absence of placebo

According to Figure 3b, Hung et al. [3] proposed the concept of retention ratio, denoted by r, of the effect of the test treatment (i.e., T-P) and the effect of the active control agent (i.e., C-P) as compared to a placebo control regardless the presence of the placebo in the study. That is,

$$r = \frac{T-P}{C-P},$$

where r is a fixed constant between 0 and 1. Chow and Shao [9] introduced the parameter of δ, which is the superiority margin as compared to the placebo. The relationship among P, T, C, δ, and M is illustrated in Figure 3c. At the worst possible scenario, we may select M=δ=T−P.

In this case, the retention rate becomes

$$r = \frac{T-P}{C-P} = \frac{\delta}{C-P} = \frac{M}{C-P},$$

This leads

$M = r\,(C-P)$.

Jones et al. [10] suggests that r=0.5 be chosen, while r=0.2 is probably the most commonly employed for selection of non-inferiority margin without any clinical judgment or statistical reasoning. Thus, the selection of non-inferiority margin depends upon the estimation of the retention rate of the effect of the test treatment relative to the effect of the active control agent.

Methods for Selection of Non-Inferiority Margin

Classical method

In clinical trials, equivalent limits for therapeutic equivalence generally depend on the nature of the drug, targeted patient population, and clinical endpoints (efficacy and safety parameters) for the assessment of therapeutic effect. For example, for some drugs, such as topical antifungals or vaginal antifungals, that may not be absorbed in blood, the FDA proposed some equivalent limits for some clinical endpoints such as binary response [11]. As an example, for the study endpoint of cure rate, if the cure rate for the reference drug is greater than 95%, then a difference in cure rate within 5% is not considered a clinically important difference (Table 2).

Table 2 indicates that if the response rate is between 80% and 90%, then a non-inferiority margin or equivalence limit of 15% should be chosen for non-inferiority or equivalence trials.

FDA's recommendations

The 2010 FDA draft guidance recommends two non-inferiority margins, namely M_1 and M_2 should be considered. The 2010 FDA draft

Equivalence limits (%)	Response rate for the reference drug (%)
±20	50-80
±15	80-90
±10	90-95
±5	>95

Table 2: Equivalence limits for binary responses.

guidance indicated that M_1 can be derived based on (1) the treatment effect estimated from the historical experience with the active control drug, (2) assessment of the likelihood that the current effect of the active control is similar to the past effect (the constancy assumption), and (3) assessment of the quality of the non-inferiority trial, particularly looking for defects that could reduce a difference between the active control and the new drug. Thus, M_1 is defined as the entire effect of the active control assumed to be present in the non-inferiority study

$$M_1=C-P. \quad (1)$$

In the 2010 FDA draft guidance, FDA also indicates that M_2 can be selected based on a clinical judgment which is never be greater than M_1 even if for active control drugs with small effects. In practice, a clinical judgment might argue that a larger difference is not clinically important. Thus, ruling out that a difference between the active control and test treatment that is larger than M_1 is a critical finding that supports the conclusion of effectiveness. Thus, M_2 can be obtained as follows

$$M_2=(1-\delta_0)\,M_1=(1-\delta_0)\,(C-P), \quad (2)$$

where

$$\delta_0=1-r=1-\frac{T-P}{C-P}=\frac{C-T}{C-P}.$$

Note that δ_0 is usually referred to as the ratio of the effect of the active control agent as compared to the test treatment and the effect of the active control agent as compared to the placebo. Thus, δ_0 becomes smaller if the difference between C and T decreases, i.e., T is close to C (the retention rate of T is close to 1). In this case, the FDA suggests a wider margin for the non-inferiority testing.

Chow and shao's method

Following the idea that the selected margin should not be greater than the smallest effect size that the active control has [6], Chow and Shao [9] introduced another parameter δ which is a superiority margin if the placebo ($\delta_0 > 0$)and assumed that the non-inferiority margin *M* is proportional to δ, i.e., M=λδ. Then, Under the worst scenario, i.e., T-C achieves its lower bound –M, then the largest possible M is given by, M=C – P –δ which leads to

$$M=\frac{\lambda}{1+\lambda}(C-P),$$

where

$$\lambda=\frac{r}{1-r}.$$

It can be seen that if $0 < r \le 1$, then $0 < \lambda \le 1/2$.

To account for the variability of C-P, Chow and Shao suggested the non-inferiority margins, $M_1$and M_2 be modified as follows, respectively,

$$M_3=M_1-(z_{1-\alpha}+z_\beta)SE_{C-T}=C-P-(z_{1-\alpha}+z_\beta)SE_{C-T}, \quad (3)$$

where $SE_{(C-T)}$ is the standard error of $\hat{C-T}$ and $z_a=\Phi^{-1}$ (a) assuming that

$$SE_{C-P} \approx SE_{T-P} \approx SE_{C-T}.$$

Similarly, M_2 can be modified as follows

$$M_4=rM_3=r\{C-P-(z_{1-\alpha}+z_\beta)\,SE_{(C-T)}\} \quad (4)$$

$$=\frac{\lambda}{1+\lambda}\left\{C-P-(z_{1-\alpha}+z_\beta)SE_{C-T}\right\},$$

$$=\left(1-\frac{1}{1+\lambda}\right)M_3,$$

Where δ_0 is chosen to be $\frac{1}{1+\lambda}$ as suggested by Chow and Shao [9]

Alternative methods

Let C_L and C_U be the minimum and maximum effect of C when comparing with P. If the effect of the test treatment falls within the range of (C_L, C_U) we consider T is equivalent to C and superior to P. Consider the worst possible scenario that the effect of the active control falls on C_U, while the effect of the test treatment T falls on C_L. In this case, we may consider the difference between C_L and C_U and the non-inferiority margin. That is

$$M_5=\hat{C}_U-\hat{C}_L. \quad (5)$$

In addition, since the selection of M depends upon the choice of δ_0, in practice, δ_0 is often chosen as either δ_0=0.5 (r=0.5) or δ_0=0.8 (r=0.2). The non-inferiority margin becomes narrower when δ_0 closes to 1. Based on the above argument, at the worst possible scenario, δ_0 can be estimated by

$$\hat{\delta}_0=\frac{\bar{C}-\bar{T}}{\bar{C}-\bar{P}}=1-\frac{\hat{T}-\hat{P}}{\hat{C}-\hat{P}}=1-\frac{\hat{C}_L}{\hat{C}_U}.$$

Thus,

$$M_6=rM_1=(1-\frac{\hat{C}_L}{\hat{C}_U})\,(\hat{C}-\hat{P}) \quad (6)$$

Remarks

It should be noted that the above methods (except for the classical method) for determination of non-inferiority margin M is based on data observed from previous superiority studies comparing the active control agent and a placebo and data collected from superiority studies comparing the test treatment and the placebo if available. Thus, the selected margin is in fact an estimate rather than a fixed margin. In other words, the selected margin is a random variable whose statistical properties are unknown. In addition, since the selected non-inferiority margin has significant impact on power calculation for sample size, it is suggested that a sensitivity analysis be performed to carefully evaluate the potential impact of the selected margin on non-inferiority testing.

As indicated by the ICH guideline, the selection of a non-inferiority margin should take both clinical judgment and statistical reasoning into consideration. The 2010 FDA draft guidance, however, however, emphasizes on statistical reasoning based on historical data from previous superiority studies comparing the active control agent and the placebo. In practice, there is always discrepancy between the margin suggested by the investigator and the margin recommended by the FDA. In this case, it is suggested that medical/statistical reviewers be consulted/communicated and hopefully to reach an agreement on the selection of the non-inferiority margin following the general principles as described in the FDA draft guidance.

An Example

Suppose that a pharmaceutical company is interested in conducting a non-inferiority trial comparing a test treatment with a standard of care treatment in terms of safety and efficacy. The test treatment is intended for treating patients with certain diseases. The primary efficacy endpoint is the cure rate. At the planning stage of the non-inferiority trial, the question regarding the selection of an appropriate non-inferiority margin and power calculation for sample size based on the selected margin is raised. It is recognized that a narrower margin will require a much larger sample size for achieving the desired power for establishing non-inferiority of the test treatment. The pharmaceutical company then follows both ICH guideline and FDA guidance for selection of an appropriate non-inferiority margin based on available historical data of the active control agent as compared to placebo.

Historical data for comparing the active control agent with a placebo are summarized in Table 3. Since the response rate for the active control is C=61.4%, the classical method suggests a non-inferiority margin of 20% be considered. Also, from Table 3, the placebo effect is P=14.3%. Thus,

M_1=C– P=61.4% – 14.3%=47.1%.

The range of C – P is given by (39.7%, 56.7%). If we assume that the retention rate is 70% (i.e., δ_0=1−r=0.3), then r=1 – δ_0=0.3. This gives

$M_2=(1-\delta_0)\, M_{1=}0.3 \times 47.1\%=14.1\%$.

Assuming that $SE_{C\text{-}P} \approx SE_{T\text{-}P} \approx SE_{C\text{-}T,}$ we have $SE_{C\text{-}T}$=6.7%. This leads to

$M_3=M_1-(z_{(1-\alpha)}+z_\beta)\, SE_{C\text{-}T}$

=47.1% – (1.96+0.84)) × 6.7%

=27.7%,

Consequently,

$$M_4=\left(1-\frac{1}{1+\lambda}\right)M_3=0.76\times 27.7\%=21\%.$$

For the proposed margin M_5, since the minimum effect and maximum effect of C-P are given by $C_L=39.7\%$ and $C_U=56.7\%$, we have

$$M_5=\hat{C}_U-\hat{C}_L=56.7\%-39.7\%=17\%.$$

Also, since $\hat{\delta}_0=\frac{39.7\%}{56.7\%}=0.68$, $r=1-\hat{\delta}_0=1-0.68=0.32$. This leads to

$$M_6=rM_1=\left(1-\frac{\hat{C}_L}{\hat{C}_U}\right)\left(\hat{C}-\hat{P}\right)=0.32\times 47.1\%=15.1\%.$$

To provide a better understanding, these margins are summarized in Table 4. As it can be seen from Table 4, the margin ranges from 14.1% to 47.1% (the entire effect of the active control agent) with a median of 21% which is close to the classical method. It should be noted that, prior to the publication of the 2010 draft guidance, FDA recommends a non-inferiority margin of 15%, while the sponsor is requesting a non-inferiority margin of 20% (Table 4).

Note that considering M=0.5(C-P), a conservative estimate of C effect is obtained using the lower 95% confidence limit of 53.4%. Assuming a 14% therapeutic cure rate of placebo, a 34% therapeutic cure rate from T will maintain the retention ratio of (T-P)/(C-P) for 50%.

Concluding Remarks

In this article, following similar ideas described in the 2010 FDA draft guidance on non-inferiority clinical trials, several alternative methods for selection of an appropriate non-inferiority margin are discussed. These methods were derived by taking into consideration of (1) the variability of the observed mean difference between the active control agent (C) and the placebo (P), the test treatment (T) and the active control agent, and the test treatment and the placebo (if available) and (2) the retention rate between the effect of test treatment as compared to the placebo (T-P) and the effect of the active control agent as compared to the placebo (C-P). The proposed methods utilize median of estimates of the retention rates based on the historical data observed from superiority studies of the active control agent as compared to the placebo.

In practice, for selection of an appropriate margin for establishing of non-inferiority of a test treatment as compared to a standard of care treatment, FDA suggests that the 2010 FDA draft guidance be consulted. In addition, communications with medical and/or statistical reviewers are encouraged especially when there is disagreement on the selected margin. It, however, should be noted that in some cases, power calculation for sample size based on binary response (e.g., incidence rate of adverse events and cure rate) may not be feasible for clinical studies with extremely low incidence rates.

These methods described in this article show the non-inferiority in efficacy of the test treatment to the active control agent, but do not have the evidence of the superiority of the test treatment to the active

Active control Agent	Year of Submission	N	Active Control (C)	Placebo Cure Rate (P)	Difference in Cure Rate
C_1	1984 1985	279 209	63.1% 60.2%	7.3% 4.0%	55.8% 56.2%
C_2	1986	101	60.0%	14.0%	46.0%
C_3	1986 1986 1986 1988	100 108 90 137	70.0% 55.1% 66.0% 58.7%	13.3% 13.6% 18.6% 17.6%	56.7% 41.5% 47.4% 41.1%
C_4	1982 1986 1988	203 88 97	60.2% 60.0% 60.9%	20.5% 16.7% 17.6%	39.7% 43.3% 43.3%
Mean SD Minimum Maximum	–	–	61.4% 4.1% 55.1% 70.0%	14.3% 5.2% 4.0% 20.5%	47.1% 6.7% 39.7% 56.7%

Table 3: Summary statistics of historical data.

Method	Suggested non-inferiority margin
Classical Methods	20.0%
Hung et al.'s suggestion with r=0.5	23.1%
FDA's M_1 approach	47.1%
FDA's M_2 approach with r=0.3*	14.1%
Chow and Shao's M_3 margin	27.7%
Chow and Shao's M_4 margin	21.0%
Proposed M_5 margin	17.0%
Proposed M_6 margin	15.1%
*Retention rate of 70%	

Table 4: Non-inferiority margins suggested by various methods.

control agent in safety. Tsou et al. [12] proposed a non-inferiority test statistic for testing the mixed hypothesis based on treatment difference and relative risk for active control trial. One benefit of the mixed test is that we do not need to choose between difference test and ratio test in advance. In particular, this mixed null hypothesis consists of a margin based on treatment difference and a margin based on relative risk. Tsou et al. [12] proposed mixed non-inferiority test not only preserves the type I error rate at desired level but also gives the similar power as that from the difference test or as that from the ratio test.

References

1. D'Agostino RB Sr, Massaro JM, Sullivan LM (2003) Non-inferiority trials: design concepts and issues - the encounters of academic consultants in statistics. Stat Med 22: 169-186.
2. Tsong Y, Higgins K, Wang SJ, Hung HMJ (1999) An overview of equivalence testing - CDER reviewers perspective. Proceedings of the Biopharmaceutical Section of American Statistical Association.
3. James Hung HM, Wang SJ, Tsong Y, Lawrence J, O'Neil RT (2003) Some fundamental issues with non-inferiority testing in active controlled trials. Stat Med 22: 213-225.
4. Laster LL, Johnson MF (2003) Non-inferiority trials: the 'at least as good as' criterion. Stat Med 22: 187-200.
5. Phillips KF (2003) A new test of non-inferiority for anti-infective trials. Stat Med 22: 201-212.
6. ICH E10 (2000) International Conference on Harmonization Guideline on Choice of Control Group and Related Design and Conduct Issues in Clinical Trials, USA.
7. FDA (2010) Guidance for Industry – Non-inferiority Clinical Trials. The United States Food and Drug Administration, Rockville, Maryland, USA.
8. Chow SC, Shao J, Wang H (2007) Sample Size Calculation in Clinical Research, Second Edition. Chapman and Hall/CRC Press, New York.
9. Chow SC, Shao J (2006) On non-inferiority margin and statistical tests in active control trials. Stat Med 25: 1101-1113.
10. Jones B, Jarvis P, Lewis JA, Ebbutt AF (1996) Trials to assess equivalence: the importance of rigorous methods. BMJ 313: 36-39.
11. Huque M, Dubey SD (1990) A three arm design and analysis for clinical trials in establishing therapeutic equivalence with clinical endpoints. Proceedings of the Biopharmaceutical Section of the American Statistical Association.
12. Tsou HH, Hsiao CF, Chow SC, Yue L, Xu Y, et al. (2007) Mixed noninferiority margin and statistical tests in active controlled trials. J Biopharm Stat 17: 339-357.