Research paper

A note on the determination of non-inferiority margins with application in oncology clinical trials☆

Binbing Yu[a,*], Harry Yang[b], Antony Sabin[c]

[a] *Oncology Biometrics, AstraZeneca, One MedImmune Way, Gaithersburg, MD, 20878, USA*
[b] *AI & Analytics Research, Biopharmaceutical Unit, AstraZeneca, One MedImmune Way, Gaithersburg, MD 20878, USA*
[c] *Oncology Biometrics, AstraZeneca Ltd., Central Cambridge, UK*

ARTICLE INFO

ABSTRACT

The goal of a non-inferiority trial is to evaluate whether the effect of an experimental treatment is not inferior to that of the active control. Determination of an appropriate non-inferiority margin is critical to the demonstration of non-inferiority. A commonly used method is called the fixed-margin approach recommended by the FDA. The fixed-margin approach consists of two steps: first the lower limit of the $1 - \alpha^*$ two-sided confidence interval (CI) of the active-control effect versus placebo is calculated from relevant historical trials or meta-analysis; second, the non-inferiority margin is obtained as a fraction of the lower confidence limit of the control effect to preserve partial control effect. An alternative method is to use the point estimate, instead of the lower confidence limit, of the active-control effect. The fixed-margin approach based on the lower limit may be ultra-conservative with unconditional Type 1 error rate much smaller than target $\alpha/2$ level, while the margin based on the point estimate is liberal. We derive the Type 1 error rate as a function of variances of the effect estimates in the historical and the current non-inferiority trials. We also propose an alternative approach for the non-inferiority margin that maintains the target Type 1 error rate. For the endpoint of landmark survival, we conduct simulations to compare the fixed-margin methods and the proposed method. For illustration, we apply the proposed method to an oncology non-inferiority clinical trial to determine an alternative non-inferiority margin.

1. Introduction

If the absence of any treatment may lead to death or serious irreversible morbidity to patients in clinical trials for indications where effective treatments exist, a placebo-controlled trial is considered unethical [1,2]. In additional to the superiority trials that demonstrate the efficacy of a new treatment over an existing therapy, non-inferiority (NI) trials are often used to evaluate whether the experimental treatment is not unacceptably less efficacious than the current active-control treatment already and as the same time is more effective than the placebo. The maximum acceptable amount of loss of clinical efficacy is represented by the NI margin [3].

It is critical to set the appropriate NI margin for a NI trial as the choice of the NI margin dictates the conclusions of the trial and clinical decision-making. The choice of NI margin must satisfy the requirement that the difference in clinical benefits between the two treatments is negligible. In practice, the methods of determining the NI margin vary

considerably [4] and often times the rationales for defining the margins are not reported [5]. There are several guidance documents available for the design, conduct and analysis of NI clinical trials [6,7]. While 'clinical judgement' was mentioned in the guidance, there was rarely a clinical trial that the NI margin was solely based on clinical judgement. In practice, the determination of NI margin should also be a statistical issue that summarizes information of treatment effects from historical data [8]. Therefore, the NI margin must be pre-specified based on both clinical and statistical reasoning.

There has been extensive research on the statistical determination of the NI margin. The most commonly used method is the fixed-margin approach, which is a two-step process. Here we consider oncology clinical trials where the primary endpoint is the objective response rate or landmark survival rate. In this scenario, the higher response rate indicates a better clinical efficacy. The treatment effect is measured as the difference of response rates between the treatment and placebo groups. The first step of the fixed margin approach is to calculate the

---

lower limit of the two-sided $1 - \alpha^*$ confidence interval (CI) of the clinical effect of active-control versus placebo. Typically, $\alpha^* = 0.05$ is used to obtain the 95% CI. The second step is to multiply the lower confidence limit by a factor of λ to preserve $(1 - \lambda) \times 100\%$ of the active-control effect. The fixed-margin approach has been shown to be conservative, where the unconditional Type 1 error rate may be substantially lower than the target level in some cases [9,10]. The ultra-conservative NI margin may lead to very high or even infeasible sample sizes for the NI studies. Sankoh [10] demonstrated the conservativeness of the fixed-margin approach for continuous endpoints with simulations.

In the article, we derive the formula for unconditional Type 1 error rate for the NI margin based on the CI and propose a simple alternative for determining the NI margin, which is still conservative but with Type 1 error rate closer to the target level. For survival endpoints, e.g., landmark survival, we use simulation to demonstrate the conservativeness of the fixed-margin approach and to evaluate the performance of the proposed method. The rest of the article is organized as follows. In Section 2, we describe the NI test and two methods for determining the NI margin. We derive the formula for the unconditional Type 1 error rate and propose the modified CI approach. We evaluate the performances of various methods using simulations. The proposed method is illustrated using a hypothetical NI trial for Nivolumab in Section 4. The article ends with discussions in Section 5.

2. Methods for determining the non-inferiority margin

We consider the oncology clinical trials where the endpoint is the objective response rate or survival rate. Let μ_P, μ_E and μ_C be the response rates for the placebo (P), experimental treatment (E) and active-control (C), respectively. Without loss of generality, we assume that a higher response rate indicates a better clinical effect and $\mu_C \geq \mu_E \geq \mu_P$. Let $\mu_{C-P} = \mu_C - \mu_P$ be the true effect of active control. If the true effect is known, the NI margin is often defined as $\Delta_\lambda = \lambda \mu_{C-P}$ to preserve $(1 - \lambda) \times 100\%$ of the control effect.

Let $\mu_{C-E} = \mu_C - \mu_E$ denote the difference between the control and the experimental treatments. The hypothesis to be tested in a NI trial is

$$H_0: \mu_{C-E} \geq \Delta_\lambda \text{ versus } H_1: \mu_{C-E} < \Delta_\lambda. \tag{1}$$

Let $\hat{\mu}_{C-E}$ and $\hat{\sigma}_{C-E}$ be the point estimate of the difference μ_{C-E} and the associated standard error. If the upper confidence limit $\hat{\mu}_{C-E} + z_{1-\alpha/2}\hat{\sigma}_{C-E} < \Delta_\lambda$, the null hypothesis is rejected and the NI conclusion is reached. Because the upper limit of a two-sided $(1 - \alpha)$ CI is used, the Type 1 error rate is $\alpha/2$. Note that under the null hypothesis with the known effect of μ_{C-P} and the constancy assumption that the active-control effect in the current NI study is similar to that observed in historical data, the asymptotic estimates should satisfy that

$$\hat{\mu}_{C-E} \sim N(\lambda \mu_{C-P}, \sigma_{C-E}^2), \tag{2}$$

$$\hat{\mu}_{C-P} \sim N(\mu_{C-P}, \sigma_{C-P}^2). \tag{3}$$

Therefore, the asymptotic distribution of the difference is

$$\hat{\delta} = \hat{\mu}_{C-E} - \lambda \hat{\mu}_{C-P} \sim N(0, \sigma_{C-E}^2 + \lambda^2 \sigma_{C-P}^2). \tag{4}$$

Because the true active-control effect is often unknown and should be estimated from historical data. The estimate may be based on one relevant study with the same patients population and trial design. If there are more than one study, one may estimate μ_{C-P} using meta-analysis by combining the control effect estimates from individual studies. Here, we consider the situation where one historical trial is available for the estimation of placebo and active-control effects and assume the constancy assumption. Let and $\hat{\mu}_{C0}$ be the estimates of mean response rates μ_P and μ_C from the historical trial, respectively. Let $\hat{\mu}_{C-P} = \hat{\mu}_{C0} - \hat{\mu}_{P0}$ be the point estimate of difference μ_{C-P} and let $\hat{\sigma}_{C-P}$ be the standard error. We start by discussing two common approaches to the determination of NI margin. The first one is the point-estimate

(PE) approach by setting $M_1 = \hat{\mu}_{C-P}$ and the second one is to use the lower limit of a two-sided CI

$$M_1 = \hat{\mu}_{C-P} - z_{1-\alpha^*/2}\hat{\sigma}_{C-P}, \tag{5}$$

where z_p is the $p \times 100\%$ percentile of a standard normal distribution. The CI approach is also called the fixed-margin approach when $\alpha^* = \alpha$. Particularly when $\alpha = \alpha^* = 0.05$, the CI approach is sometimes called the 95%–95% method [7]. The type 1 error rate of an NI test with NI margin $\hat{\Delta}_\lambda = \lambda M_1$ is calculated as

$$\Pr(\text{Rejecting } H_0 | H_0) = \Pr(\hat{\mu}_{C-E} + z_{1-\alpha/2}\hat{\sigma}_{C-E} < \Delta_\lambda | H_0), \tag{6}$$

where M_1 is defined by the PE or CI approaches. The type 1 error probability in (6) is an unconditional error because the probability incorporates the statistical uncertainty of the estimate of NI margin $\hat{\Delta}_\lambda$ [11,12]. If the margin is treated as a fixed known constant after it is estimated, the conditional error probability is controlled at level $\alpha/2$. However, when M_1 is estimated from the PE and CI approaches, the unconditional error probability may be different from target $\alpha/2$ level because of the uncertainty associated with the estimation of the NI margin. The uncertainty may arise from the sampling variability within a trial or between-trial variability.

The Type 1 error rate under consideration is the unconditional probability thereafter. We have the following theorem for the Type 1 error rates for the PE and CI approaches.

Theorem 1. Let α_{PE} and α_{CI} be the Type 1 error rates for the PE and fixed margin CI approaches, respectively. If $\hat{\mu}_{C-E}$ and $\hat{\mu}_{C-P}$ are asymptotic normal as specified in (2) and (3) and $\hat{\sigma}_{C-E}$ and $\hat{\sigma}_{C-P}$ are consistent estimates of σ_{C-E} and σ_{C-P}, the asymptotic Type 1 error rates $\alpha_{PE} > \alpha/2 > \alpha_{CI}$.

PROOF. Note that $\hat{\sigma}_{C-E} < (\hat{\sigma}_{C-E}^2 + \lambda^2 \hat{\sigma}_{C-P}^2)^{\frac{1}{2}} < \hat{\sigma}_{C-E} + \lambda \hat{\sigma}_{C-P}$. When the PE approach is used, $M_1 = \hat{\mu}_{C-P}$ and the Type 1 error rate is

$$
\begin{aligned}
\alpha_{PE} &= \Pr(\hat{\mu}_{C-E} + z_{1-\alpha/2}\hat{\sigma}_{C-E} < \lambda \hat{\mu}_{C-P}) \\
&= \Pr(\hat{\mu}_{C-E} - \lambda \hat{\mu}_{C-P} + z_{1-\alpha/2}\hat{\sigma}_{C-E} < 0) \\
&> \Pr(\hat{\mu}_{C-E} - \lambda \hat{\mu}_{C-P} + z_{1-\alpha/2}(\hat{\sigma}_{C-E}^2 + \lambda^2 \hat{\sigma}_{C-P}^2)^{\frac{1}{2}} < 0).
\end{aligned}
$$

Because $\hat{\sigma}_{C-E}$ and $\hat{\sigma}_{C-P}$ are consistent estimates of σ_{C-E} and σ_{C-P} and $\hat{\delta} = \hat{\mu}_{C-E} - \lambda \hat{\mu}_{C-P}$ is asymptotic normal as equation (4),

$$\Pr\left(\hat{\delta} + z_{1-\alpha/2}(\hat{\sigma}_{C-E}^2 + \lambda^2 \hat{\sigma}_{C-P}^2)^{\frac{1}{2}} < 0\right) \to \Pr\left(-\hat{\delta} > z_{1-\alpha/2}\left(\sigma_{C-E}^2 + \lambda^2 \sigma_{C-P}^2\right)^{\frac{1}{2}}\right) = \alpha/2.$$

Therefore, the Type 1 error for the PE approach is liberal with $\alpha_{PE} > \alpha/2$.

When the fixed-margin or CI approach is used as defined in Equation (5), $\alpha = \alpha^*$, the Type 1 error rate is

$$
\begin{aligned}
\alpha_{CI} &= \Pr(\hat{\mu}_{C-E} + z_{1-\alpha/2}\hat{\sigma}_{C-E} < \lambda(\hat{\mu}_{C-P} - z_{1-\alpha^*/2}\hat{\sigma}_{C-P})) \\
&= \Pr(\hat{\mu}_{C-E} - \lambda \hat{\mu}_{C-P} + z_{1-\alpha/2}\hat{\sigma}_{C-E} + \lambda z_{1-\alpha^*/2}\hat{\sigma}_{C-P} < 0) \\
&< \Pr(\hat{\mu}_{C-E} - \lambda \hat{\mu}_{C-P} + z_{1-\alpha/2}(\hat{\sigma}_{C-E}^2 + \hat{\sigma}_{C-P}^2)^{\frac{1}{2}} < 0) \to \alpha/2.
\end{aligned}
$$

This indicates the fixed margin CI method is conservative with $\alpha_{CI} < \alpha/2$.

Corollary 1. Let $k = \hat{\sigma}_{C-E}/\hat{\sigma}_{C-P}$. Under the assumptions for Theorem 1, the actual Type 1 error rates for the PE and fixed margin CI approaches are

$$\alpha_{PE} = 1 - \Phi\left(\frac{z_{1-\alpha/2}k}{(k^2 + \lambda^2)^{\frac{1}{2}}}\right), \tag{7}$$

$$\alpha_{CI} = 1 - \Phi\left(\frac{z_{1-\alpha/2}k + z_{1-\alpha^*/2}\lambda}{(k^2 + \lambda^2)^{\frac{1}{2}}}\right). \tag{8}$$

PROOF. The actual Type 1 error rate for the fixed-margin CI

approach is

$$\alpha_{CI} = \Pr\left(\hat{\mu}_{C-E} - \lambda\hat{\mu}_{C-P} + z_{1-\alpha/2}\hat{\sigma}_{C-E} + \lambda z_{1-\alpha^*/2}\hat{\sigma}_{C-P} < 0\right)$$

$$= \Pr\left(-\frac{\hat{\mu}_{C-E} - \lambda\hat{\mu}_{C-P}}{(\hat{\sigma}_{C-E}^2 + \lambda^2\hat{\sigma}_{C-P}^2)^{\frac{1}{2}}} > \frac{z_{1-\alpha/2}\hat{\sigma}_{C-E} + \lambda z_{1-\alpha^*/2}\hat{\sigma}_{C-P}}{(\hat{\sigma}_{C-E}^2 + \lambda^2\hat{\sigma}_{C-P}^2)^{\frac{1}{2}}}\right).$$

Note that $-\frac{\hat{\mu}_{C-E} - \lambda\hat{\mu}_{C-P}}{(\hat{\sigma}_{C-E}^2 + \lambda^2\hat{\sigma}_{C-P}^2)^{\frac{1}{2}}}$ asymptotic follows a standard normal distribution.

$$\alpha_{CI} = 1 - \Phi\left(\frac{z_{1-\alpha/2}\hat{\sigma}_{C-E} + \lambda z_{1-\alpha^*/2}\hat{\sigma}_{C-P}}{(\hat{\sigma}_{C-E}^2 + \lambda^2\hat{\sigma}_{C-P}^2)^{\frac{1}{2}}}\right) = 1 - \Phi\left(\frac{z_{1-\alpha/2}k + z_{1-\alpha^*/2}\lambda}{(k^2 + \lambda^2)^{\frac{1}{2}}}\right).$$

The Type 1 error rate for the PE approach α_{PE} is obtained by removing the term $z_{1-\alpha/2}\lambda$ in Equation (7).

We have the following remarks about the Type 1 error rates of the NI test.

Remark 1. When $\lambda = 0$, i.e., 100% of the control effect is retained, both the PE and the fixed-margin CI approaches maintain the α level $\alpha_{CI} = \alpha_{PE} = \alpha/2$.

Remark 2. If the value of λ is fixed and $z_{1-\alpha/2} > 1$, α_{CI} is an increasing function of k.

Remark 3. If the effect of control effect is very precise with a small standard error, i.e., $\hat{\sigma}_{C-P} \approx 0$, then $k \approx \infty$. Then α_{CI} and α_{PE} are both close to $\alpha/2$.

Remark 4. If the NI trial involves a larger sample size than the historical trials or the control effect is estimated from meta-analysis of heterogeneous studies, we expect that $\hat{\sigma}_{C-E} \leq \hat{\sigma}_{C-P}$. The upper bound of α_{CI} can be obtained by setting $k = 1$ is Equation (8).

In order to maintain the target Type 1 error rate for the fixed-margin CI approach, one has to adjust the α^* value so that

$$z_{1-\alpha/2} = \frac{z_{1-\alpha/2}k + z_{1-\alpha^*/2}\lambda}{(k^2 + \lambda^2)^{\frac{1}{2}}}. \tag{9}$$

The correct α^* value that is needed for maintaining the target Type 1 *error rate is*

$$\alpha^* = 2\left\{1 - \Phi\left(z_{1-\alpha/2}\left[(k^2 + \lambda^2)^{\frac{1}{2}} - k\right]/\lambda\right)\right\}. \tag{10}$$

The CI approach with the correct α^* level as in (10) is referred to as the the modified CI approach, in comparison to the fixed-margin CI approach.

The conservativeness of the fixed-margin CI approach has been noticed by several authors [11,13]. Rothmann [14] proposed a two-CI procedure based on the CI for $\hat{\mu}_{C-E}$ and $\lambda\hat{\mu}_{C-P}$, where the NI will be inferred when

$$\hat{\mu}_{C-E} + z_{1-\alpha/2}\hat{\sigma}_{C-E} < \lambda\{\hat{\mu}_{C-P} - z_{1-\alpha^*/2}\hat{\sigma}_{C-P}\}, \tag{11}$$

where α^* is defined as equation (10). Rothmann's procedure relaxes the conservatism of 95%-95% fixed margin method and maintain the correct Type 1 error. This procedure belongs to the category of synthesis method that combined data from historical trials and the current NI trial into one analysis [7]. The Rothmann's two-CI procedure and the modified CI approach are equivalent analytically and both control the Type 1 error. The two-CI procedure can only be conducted when both historical data and NI trial data are available. The proposed modified CI approach is more suitable for the conduct of clinical trials, where the NI margin is pre-defined in advance of the NI trial. It also allows input of clinical judgement of the appropriateness of the NI margin and planning of the NI trial [7].

In practice, it is difficult to determine the exact value of k, which is the ratio of the standard deviation of the NI trial to that of the historical trial. As shown in Fig. 1 in the simulation, the half-width of the
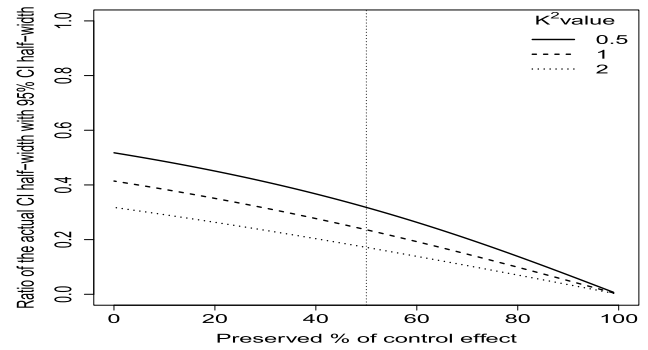

Fig. 1. The ratio of the half-width with the confidence level $1 - \alpha^*$ to $z_{0.975}$.

corrected $1 - \alpha^*$ CI is an increasing function of k, one may use the lower bound of k to set the correct α level. Another choice is to use set $k = \sqrt{n_H/n_{NI}}$, where n_H and n_{NI} are the number of subjects in the historical and the current NI trials. This is because the ratio of variance is inversely related to the number of subjects under the hypothesis of equivalent survival. In many situations, it is difficult to project the sample size in the NI trial, one may set a reasonable range for k to obtain the bounds of Type 1 error. Based on equation (10), the modified NI margin based on α^* is a decreasing function of k. For example, if we anticipate that the original NI trial size n_{NI} needs to be doubled, we may set the NI margin to be a value corresponding to $k = \sqrt{0.5}\sqrt{n_H/n_{NI}}$. This adjustment may still control the Type 1 error below $\alpha/2$, but not overly conservative.

3. Numerical study

Sankoh [10] conducted numeric studies to show the inflated Type 1 error rates for the PE approach and the deflated Type 1 error rates for the CI approach. Here we focus on the CI approach, which is recommended by the regulatory agencies for setting the NI margin. First, we calculate the correct α^* value that is needed to maintain the correct Type 1 error rate. Second, we run simulations to examine the performance of the two CI approaches and the impact of using an adjusted α^* value for testing the NI of landmark survival rates at a fixed time point.

3.1. Ratio of the NI margins from the 95%–95% fixed margin approach and the modified CI approach

To compare with the 95%–95% fixed margin approach, we examine the reduction of half-width of the CI for constructing the NI margin in Equation (5). Let $R = z_{1-\alpha^*}/z_{0.975}$, where $z_{1-\alpha^*}$ is half-width of the CI that maintains the correct Type 1 error rate in Equation (10) and $z_{0.975} = 1.96$ for the 95%–95% fixed margin approach. The plot of R with respect to different values of k and preservation fraction is shown in Fig. 1. We see clearly that the actual half-width of the $1 - \alpha^*$ CI that leads to the correct Type 1 error rate is much smaller than the full 95% CI. In a typical scenario where $k = 1$ and 50% control effect is preserved, $R = 0.236$. Therefore, the NI margin can be constructed as $M_1 = \hat{\mu}_{C-P} - 0.236 \times 1.96 \times \hat{\sigma}_{C-P}$.

3.2. Type 1 error rate of the modified CI approach

To evaluate the performance of the NI margins based on the fixed-margin CI approach and the modified CI approach in oncology NI trials, we run some additional simulations. We assume a mixture cure model with Weibull model for the uncured patients for the historical trial. The overall survival function is

$$S_x(t) = \pi_x + (1 - \pi_x)\exp(-(t/b_x)^a),$$

where $x = P$ for the placebo group and $x = C$ for the control group, π_x is
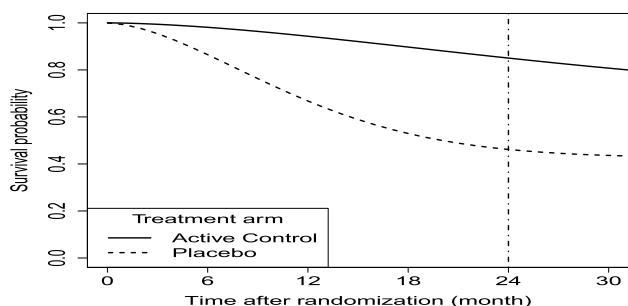
Fig. 2. Survival curves for the historical data in the simulation.

Fig. 3. Type 1 error rate for the 95%–95% fixed-margin CI approach.

the cure or long-term survival rate, a is the shape parameter and b_x is the scale parameter. For the simulation, we set $\pi_P = 0.43$, $\pi_C = 0.68$, $a = 1.69$, $b_P = 13.1$ and $b_C = 31.7$. The survival curves for the control and placebo groups are shown in Fig. 2. The primary endpoint is the landmark survival rate at month 24. The true difference in the 24-month survival rates is $\mu_{C-P} = S_C(24) - S_P(24) = 0.13$. The NI margin is $\Delta_\lambda = \lambda\mu_{C-P}$ to preserve $(1-\lambda)\%$ of the control effect.

Suppose that an experimental treatment is under development with the goal to show that the survival rate at month 24 is not inferior to that for the active control by a margin of Δ_λ. The hypothesis to be tested is

H_0: $S_C(24) - S_E(24) \geq \Delta_\lambda$ versus H_1: $S_C(24) - S_E(24) < \Delta_\lambda$.

The simulation is conducted as follows:

1. Historical data for the active control and placebo are generated from the mixture cure models with sample sizes 100, 150, 200 and 300 per treatment arm.
2. The NI margin is determined as the lower 95% confidence limit per FDA guidance or by the modified CI approach using Equation (9) with $k = \sqrt{n_H/n_{NI}}$. In the situation that we don't know the actual sample size n_{NI}, but we expect that $n_H \leq n_{NI} \leq 4n_H$, so that $0.5 \leq k \leq 1$. The NI margin is set at the minimum value of the lower 95% confidence limit with $k = 0.5$.
3. The data from the NI trials are generated from the mixture cure model
   (a) Under the constancy assumption, the survival data for the active control are generated from the survival function $S_1(t)$.
   (b) The survival data for the experimental treatment group are generated from the mixture cure model such that $S_C(24) - S_E(24) = \Delta_\lambda$.
4. Perform the NI test by comparing the upper bound of the survival difference and the NI margin.

The point estimate and the CI of the landmark survival difference are calculated using the non-parametric Beta product method [15]. The Type 1 error rates are calculated as the proportion of rejecting the null hypothesis when the null hypothesis is true. We set $\alpha = 0.05$ so the target Type 1 error rate is 0.025. Note that the a new set of historical data are generated in each simulation, the resulting Type 1 error rate is unconditional on the historical data. The simulation results are based on 5,000 simulations with sample size $n_{NI} = 200$ and 400. Here we only present the results for $n_{NI} = 200$ as the results for $n_{NI} = 400$ are similar.

In Fig. 3, we see that the Type 1 error rates based on the 95%–95% fixed margin are substantially lower than 0.025. When the percent of retained control effect decreases, the Type 1 error rates may drop to below 0.005. This indicates that the fixed-margin approach may be ultra-conservative for the NI test for landmark survival rates. The Type 1 error rates for the modified CI approach with adjusted α^* value are shown in Fig. 4. The adjust α^* value is based on equation (10) with $k = \sqrt{n_H/n_{NI}}$. We see that the Type 1 error rates increases remarkably to around 0.02, but still are all below the target level 0.025.
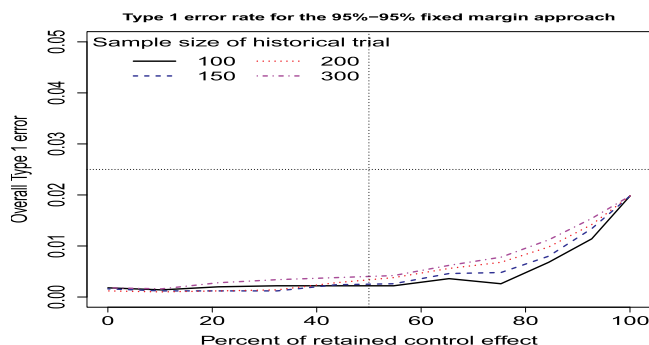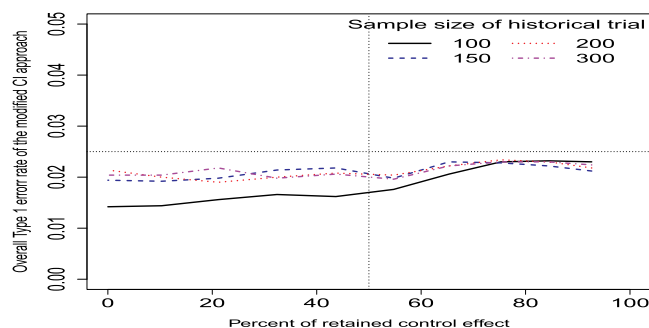

Fig. 4. Type 1 error rate for the CI approach with the adjusted α^* value.

In the situation when the sample size for the NI trial may be increased in an adaptive clinical trial, one may set the NI margin as if the NI trial sample size would be doubled, i.e., $k = \sqrt{n_H/(2n_{NI})}$. The resulting Type 1 error rate is shown in Fig. 5. We see that the Type 1 error rates is about 0.01, which is still conservative. In a special case where $n_H = n_{NI}$, $k^2 = 0.5$. As shown in Fig. 1, the resulting NI margin is about double the width of the 95%–95% fixed NI margin. This shows that the NI margin based on the modified CI approach is still conservative, while is much wider than the 95%–95% fixed margin.

4. Application

For illustration, we apply the proposed method to the overall survival data from the ATTRACTION-2 clinical trial. This is a randomised, double-blind, placebo-controlled, phase 3 trial to evaluate the effect of nivolumab with placebo for patients with advanced gastric or gastro-oesophageal junction cancer refractory to, or intolerant of, at least two previous chemotherapy regimens [16]. The survival data are reconstructed from the plots of Kaplan-Meier curves of overall survival
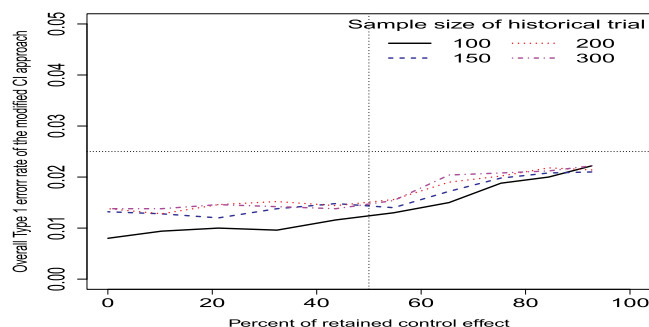

Fig. 5. Type 1 error rate for the CI approach with the NI trial sample size doubled.
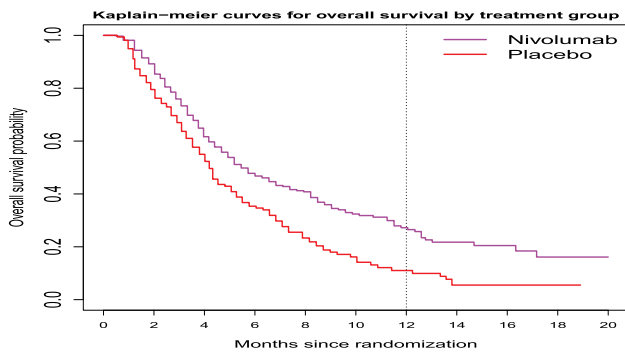
Fig. 6. Kaplan-Meier curves by treatment for the reconstructed overall survival data for the ATTRACTION-2 trial.

using the algorithm described by Ref. [17]. The digitized data for the PFS curves are extracted using WebPlotDigitizer [18]. The Kaplan-Meier curves of the reconstructed data for all patients in the study are shown in Fig. 6.

Suppose that a competitor company is developing an experimental treatment that has non-inferior survival but may reduce the side effects and improve the quality of life substantially. For simplicity, we assume that the effect of nivolumab is unbiased in ATTRACTION-2 trial and there is a small varability across different trials of nivolumab for the indication under study. We use the landmark survival rate at month 12 as the endpoint. The difference of the 1-year survival probabilities between the two groups is 0.162 with standard error 0.046 and 95% CI (0.068, 0.246). What would be the appropriate NI margin for non-inferiority test? The 95%–95% fixed NI margin is 0.034, while the NI margins based on the modified CI approach are 0.140 for $k = 1$ and 0.123 for $k = 0.5$. The 95%–95% NI margin of 0.034 is smaller than the standard error of the 1-year survival differences. Using quality-adjusted time without symptoms of toxicity (Q-TWiST) as the outcome in cancer clinical trials, Revicki et al. [19] recommended that a difference of 10% in overall survvial is clinically important and meaningful. If less is known about a specific treatment and/or disease area, a clinically meaningful improvement in survival should be greater than 5%, but not more than 10%. Furthermore, using the NI margin 0.034 may lead to a sample size greater than 1,000 for the NI trial if indeed there is no difference between the experimental and the control treatments. Therefore, the 95%–95% fixed margin is too tight and not feasible. The alternative NI margin of 0.123 is slightly more conservative than 0.140 and falls within the range of 0.10–0.15 for the NI margins for the difference in probability of survival (DPS) at a particular time, which is the landmark survival, in Table 3 of the summary by Tanaka et al. [20]. Ideally, the NI margin should be based on both clinical reasoning and historical data. Appropriate NI margin is necessary to maintain the correct Type 1 error rate.

5. Discussion

For the determination of NI margin, we prove that the 95%–95% fixed margin CI approach is ultra-conservative while the PE approach is liberal. For the non-inferiority test of landmark survival in oncology clinical trials, we confirm this conclusion with simulations. The proposed modified CI approach yields the unconditional Type 1 error rates closer to the target level, thus avoiding many false negative results. The NI margin based on the modified CI approach is still conservative, but is typically wider than the 95%–95% margin. The statistically determined NI margin should be evaluated by clinical judgement to ensure that the

margin is clinically meaningful.

In this article, we also assume the constancy assumption, where the active-control effect in the current NI study is similar to that observed in historical data. It is of interest to assess the robustness of the proposed method to unknown bias [21] and non-constancy [22].

Acknowledgement

References

[1] J. Schumi, J.T. Wittes, Through the looking glass: understanding non-inferiority, Trials 12 (1) (2011) 106.

[2] S. Hahn, Understanding noninferiority trials, Korean J. Pediatr. 55 (11) (2012) 403.

[3] FDA, Guidance for Industry Non-inferiority Clinical Trials, Center for Biologics Evaluation and Research (CBER).

[4] G. Wangge, K.C. Roes, A. de Boer, A.W. Hoes, M.J. Knol, The challenges of determining noninferiority margins: a case study of noninferiority randomized controlled trials of novel oral anticoagulants, CMAJ (Can. Med. Assoc. J.) 185 (3) (2013) 222–227.

[5] T.A. Althunian, A. de Boer, R.H.H. Groenwold, O.H. Klungel, Defining the non-inferiority margin and analysing noninferiority: an overview, Br. J. Clin. Pharmacol. 83 (8) (2017) 1636–1642, https://doi.org/10.1111/bcp.13280.

[6] EMA, Guideline on the Choice of the Non-inferiority Margin, (2005).

[7] FDA, Non-Inferiority Clinical Trials to Establish Effectiveness, Guidance for Industry, (November 2016) URL https://www.fda.gov/media/78504/download.

[8] R.B. D'Agostino Sr., J.M. Massaro, L.M. Sullivan, Non-inferiority trials: design concepts and issues–the encounters of academic consultants in statistics, Stat. Med. 22 (2) (2003) 169–186.

[9] H.J. Hung, S.-J. Wang, R. O'Neill, A regulatory perspective on choice of margin and statistical inference issue in non-inferiority trials, Biom. J.: J. Math. Methods Biosci. 47 (1) (2005) 28–36.

[10] A.J. Sankoh, A note on the conservativeness of the confidence interval approach for the selection of non-inferiority margin in the two-arm active-control trial, Stat. Med. 27 (19) (2008) 3732–3742.

[11] H. James Hung, S.-J. Wang, Y. Tsong, J. Lawrence, R.T. O'Neil, Some fundamental issues with non-inferiority testing in active controlled trials, Stat. Med. 22 (2) (2003) 213–225.

[12] J. Lawrence, Some remarks about the analysis of active control studies, Biom. J. 47 (5) (2005) 616–622.

[13] E.B. Holmgren, Establishing equivalence by showing that a specified percentage of the effect of the active control over placebo is maintained, J. Biopharm. Stat. 9 (4) (1999) 651–659.

[14] M. Rothmann, N. Li, G. Chen, G.Y. Chi, R. Temple, H.-H. Tsou, Design and analysis of non-inferiority mortality trials in oncology, Stat. Med. 22 (2) (2003) 239–264.

[15] M.P. Fay, E.H. Brittain, M.A. Proschan, Pointwise confidence intervals for a survival distribution with small samples or heavy censoring, Biostatistics 14 (4) (2013) 723–736, https://doi.org/10.1093/biostatistics/kxt016.

[16] Y.-K. Kang, N. Boku, T. Satoh, M.-H. Ryu, Y. Chao, K. Kato, H.C. Chung, J.-S. Chen, K. Muro, W.K. Kang, K.-H. Yeh, T. Yoshikawa, S.C. Oh, L.-Y. Bai, T. Tamura, K.-W. Lee, Y. Hamamoto, J.G. Kim, K. Chin, D.-Y. Oh, K. Minashi, J.Y. Cho, M. Tsuda, L.-T. Chen, Nivolumab in patients with advanced gastric or gastro-oesophageal junction cancer refractory to, or intolerant of, at least two previous chemotherapy regimens (ONO-4538-12, ATTRACTION-2): a randomised, double-blind, placebo-controlled, phase 3 trial, The Lancet 390 (10111) (2017) 2461–2471, https://doi.org/10.1016/s0140-6736(17)31827-5.

[17] P. Guyot, A. Ades, M. J. Ouwens, N. J. Welton, Enhanced secondary analysis of survival data: reconstructing the data from published kaplan-meier survival curves, BMC Med. Res. Methodol. 12 (1). doi:10.1186/1471-2288-12-9.

[18] A. Rohatgi, Webplotdigitizer, version 4.2 (Apr. 2019). URL https://automeris.io/WebPlotDigitizer.

[19] D.A. Revicki, D. Feeny, T.L. Hunt, B.F. Cole, Analyzing oncology clinical trial data using the Q-TWiST method: clinical importance and sources for health state preference data, Qual. Life Res. 15 (3) (2006) 411–423.

[20] S. Tanaka, Y. Kinjo, Y. Kataoka, K. Yoshimura, S. Teramukai, Statistical issues and recommendations for noninferiority trials in oncology: a systematic review, Clin. Cancer Res. 18 (7) (2012) 1837–1847.

[21] K. Odem-Davis, T.R. Fleming, Adjusting for unknown bias in noninferiority clinical trials, Stat. Biopharm. Res. 5 (3) (2013) 248–258.

[22] Q. Liu, Y. Li, K. Odem-Davis, On robustness of noninferiority clinical trial designs against bias, variability, and nonconstancy, J. Biopharm. Stat. 25 (1) (2015) 206–225.