RESEARCH

Non-inferiority test for a continuous variable with a flexible margin in an active controlled trial : An application to the "Stratall ANRS 12110 / ESTHER" trial

Arsene Brunelle Sandie^{1*}
, Nicolas Molinari²
, Anthony Wanjoya³
, Charles Kouanfack⁴
, Christian Laurent⁵
and Jules Brice Tchatchueng-Mbougua^{6,7}

*Correspondence:

sandiearse@gmail.com

¹Department of

Mathematics-Statistics, Pan
African University Institute for
basic Sciences, Technology and
Innovation (PAUSTI/JKUAT),
Nairobi, Kenya

Full list of author information is
available at the end of the article

Abstract

Background: The non-inferiority trials are becoming increasingly popular in public health and clinical research. The choice of the non-inferiority margin is the cornerstone of the non-inferiority trial. When the effect of active control intervention is unknown, it can be interesting to choose the non-inferiority margin as a function of the active control intervention effect. In this case, the uncertainty surrounding non-inferiority margin should be taken into account in statistical tests.

Methods: It was proposed in this study two procedures for the non-inferiority test with a flexible margin for the continuous endpoint. The proposed test procedures are based on the asymptotic test and confidence interval approach. Simulations have been used to assess the performances and properties of the proposed test procedures. An application was done on clinical real data. The purpose of this study was to assess the effectiveness and safety of clinical monitoring alone versus laboratory and clinical monitoring in HIV infected adults patients.

Results: The two proposed test procedures have good properties for large sample sizes. It has been found that the confidence interval level determines approximately the level of significance. The 80%, 90% and 95% two-sided confidence interval levels led approximately to a type I error of 5%, 2.5% and 1% respectively. For small sample sizes, the test with 80% confidence interval level was the most powerful.

Conclusions: This study highlights the choice of non-inferiority margin as flexible and depending on the reference treatment in an active trial of non-inferiority. The results of the study show that the proposed methods are applicable in the practice.

Trial registration: The trial data used in this study was from the "Stratall ANRS 12110 / ESTHER", registered with ClinicalTrials.gov, number NCT00301561. Date : March 13, 2006, url : <https://clinicaltrials.gov/ct2/show/NCT00301561>.

Keywords: Asymptotic test; Active controlled trial; Confidence interval; Flexible margin; Non-inferiority

1 Background

After developing a new health intervention (treatment or diagnostic test), the next step is to assess its effectiveness relative to the existing reference intervention. There are several strategies to do this, such as the superiority trial which involves testing whether the new treatment is superior to another (placebo, reference or active control treatment). However, when the active control intervention achieves maximum efficiency or the use of a placebo is unethical, it becomes difficult to statistically show the superiority of the new health intervention. Studies aimed at showing that a new intervention is not worse than the active control intervention of more than a pre-specified amount of efficiency have become increasingly common in the recent decade [1]. The expression '*is not worse than the active control intervention of more than a pre-specified amount*', means it is acceptable to lose a 'little bit' of main effect (efficiency) of the active control intervention compared to new intervention's benefits (fewer side effects, costs, tolerable and safer). This loss of maximum efficiency pre-specified is called *non-inferiority margin*. The trial showing that the new intervention is '*non-inferior*' to the active control intervention is called *non-inferiority trial* [1].

The determination of non-inferiority margin is an important and contentious issue which directly influences the acceptability of the new health intervention in non-inferiority trial. [2] provides general principles for an appropriate choice of the non-inferiority margin. The method based on the difference of endpoint between the new and active control health interventions has been popularly used in non-inferiority trial. The determination of non-inferiority margin is based on a combination of statistical reasoning, clinical judgment and data from historical studies [3, 4]. However, there are no specific standard references for determining this margin. Therefore, these may be controversial issues on its choice. The non-inferiority margin remains arbitrary and in practice, is poorly reported in most studies [5].

In [2], are proposed two different methods to determine the non-inferiority margin which are based on the effect of the active control intervention from the historical data. Denoting the effect of the active control intervention by R and the placebo by P , the effect size of active control intervention is defined by $R - P$. Thus, the first method, called (fixed margin method), is based on the smallest effect size of the active control intervention ($R - P$) from the historical trials. The second method, called (synthesis method), is based on the effect size of the active control intervention from historical data and the current one, taking into account the variability in both sources of data. The general principle is to define a non-inferiority margin such that the effect of the new intervention against active control intervention preserves at least some fraction of the active control intervention efficiency relative to placebo. The non-inferiority margin can be viewed as a function $\Delta_L(.)$ of $R - P$, that is $\Delta_L(R - P)$. Since the effect of P is determined in absence of any intervention, it means its effect is fixed and constant. Then, the effect size in active control group $R - P$ depend to the effect in the active control group R . And then, the margin can be viewed as a function of the response in active control treatment group R in the placebo controlled trial, that is $\Delta_L(R)$ [6].

When the endpoint is binary, many test procedures considering a variable margin (as a function of the active control intervention effect) have been developed. [7]

considered a non-inferiority margin as a linear function, [8] considered a piecewise linear function and [9] generalized by considering the margin as continuously differentiable and piecewise continuously differentiable. However, few studies have been conducted in the framework of a continuous endpoint. One finds that most works on the non-inferiority test for a continuous endpoint with fixed and linear margin have been focused on the confidence intervals approach [6, 10, 11].

In practice, the true expression of Δ_L is unknown. Generally, in most studies, Δ_L is a fixed value. However, it is known that the non-inferiority margin is appropriately fixed if there is a plausible value for any value of the main response in the active control group. Otherwise, the non-inferiority margin should be a function of the main response in the active control group compared to placebo group. Indeed, if it is assumed that from k historical studies, one may get estimations of the main response in active control group $R, \hat{R}_1, \dots, \hat{R}_k$, one may use such data to estimate the unknown function Δ_L . Then, for each $\hat{R}_i, i = 1, \dots, k$, the methodologies proposed by [6] or [2] which take into account the constancy assumption and the sensitivity test can be used to estimate the non-inferiority margin for each historical studies denoted by $\hat{\Delta}_{L1}, \dots, \hat{\Delta}_{Lk}$ and k couples of points : $(\hat{R}_i, \hat{\Delta}_{Li}), i = 1, \dots, k$, are obtained. These data sets can be combined by smoothing to get an approximation of the flexible margin Δ_L .

This paper proposes an asymptotic and confidence interval based test to perform the non-inferiority test with a flexible margin when the endpoint is continuous, using the mean difference. The Delta method theorem is used to construct an asymptotic test. The bootstrap is used to construct an empirical confidence interval. We used simulations to assess the performances of both proposed approaches according to the sample size. Finally, an illustration of the proposed methods was done on the Stratall ANRS 12110/ESTHER trial.

2 Methods

2.1 Notations

Definition of the basic notations used.

- X_R and X_N , the random variables for continuous primary endpoint in the active control group and new intervention group (new group) respectively.
- n_R and n_N , the sample sizes for the active control group and new group respectively .
- μ_R and μ_N , the means of continuous primary endpoint for the active group and new group respectively.
- σ_R^2 and σ_N^2 , the variances of continuous primary endpoint for the active group and new group respectively.
- $\Delta_L(\mu_R)$ the non-inferiority margin, and $\Delta = \mu_N - \mu_R$ the difference of true means.
- H_0 and H_1 are the null and alternative hypotheses respectively.

2.2 Approach using a test statistic

Without loss of generality, assuming that the increase of the endpoint corresponds to more efficiency. The non-inferiority hypotheses can be formulated as follows :

$$\begin{cases} H_0 : \mu_N \leq \mu_R - \Delta_L & \text{There is no non-inferiority} \\ H_1 : \mu_N > \mu_R - \Delta_L & \text{There is non-inferiority} \end{cases} \quad (1)$$

The formulation of hypotheses test in equation 1 shows that the acceptance of the non-inferiority means that the new intervention is not worse than active control intervention with a margin Δ_L . When Δ_L is fixed, the hypotheses 1 can be viewed as a classical composite hypotheses test for mean difference [12], therefore, based on central limit theorem on the boundary of null hypothesis, the asymptotic test Z_{fixed} can be obtained by :

$$Z_{fixed} = \frac{\bar{X}_N - \bar{X}_R + \Delta_L}{\sqrt{\frac{\sigma_N^2}{n_N} + \frac{\sigma_R^2}{n_R}}} \sim N(0, 1), \quad (2)$$

because, when Δ_L is fixed we have :

$$\begin{aligned} Var(\bar{X}_N - \bar{X}_R + \Delta_L) &= Var(\bar{X}_N) + Var(\bar{X}_R) \\ &= \frac{\sigma_N^2}{n_N} + \frac{\sigma_R^2}{n_R} \end{aligned} \quad (3)$$

The null hypothesis is rejected if $Z_{fixed} > Z_{1-\alpha}$, where $Z_{1-\alpha}$ is the $(1 - \alpha)$ percentile of standard normal distribution. Using Karlin-Rubin theorem, this test is the uniformly most powerful test of level α [13].

If Δ_L is not fixed, i.e, if Δ_L is a variable function of μ_R , then, $Var\{\bar{X}_N - \bar{X}_R + \Delta_L(\bar{X}_R)\} \neq Var(\bar{X}_N) + Var(\bar{X}_R)$, and therefore $Var(\bar{X}_N) + Var(\bar{X}_R)$ is not a valid variance of $\bar{X}_N - \bar{X}_R + \Delta_L(\bar{X}_R)$. Under the assumption that Δ_L function is continuously differentiable, we used delta method to variance estimation.

2.2.1 Variance estimation using Delta method

Assuming that $\Delta_L(\cdot)$ is continuously differentiable such that $\Delta'_L(\mu_R) \neq 0$ (Δ'_L is the first derivative of Δ_L), and using the Taylor series of order 1 in a neighborhood of μ_R ,

$$\Delta_L(\bar{X}_R) = \Delta_L(\mu_R) + \Delta'_L(\mu_R)(\bar{X}_R - \mu_R) + o_p(1) \quad (4)$$

then,

$$\begin{aligned} &\{\bar{X}_N - \bar{X}_R + \Delta_L(\bar{X}_R)\} - \{\mu_N - \mu_R + \Delta_L(\mu_R)\} \\ &= (\bar{X}_N - \mu_N) - (\bar{X}_R - \mu_R) + \{\Delta_L(\bar{X}_R) - \Delta_L(\mu_R)\} \\ &= (\bar{X}_N - \mu_N) - (\bar{X}_R - \mu_R) + \Delta'_L(\mu_R)(\bar{X}_R - \mu_R) + o_p(1) \\ &= (\bar{X}_N - \mu_N) + \{\Delta'_L(\mu_R) - 1\}(\bar{X}_R - \mu_R) + o_p(1) \end{aligned}$$

Thus, the variance estimate is :

$$Var\{\bar{X}_N - \bar{X}_R + \Delta_L(\bar{X}_R)\} = \frac{\sigma_N^2}{n_N} + \frac{\{\Delta'_L(\mu_R) - 1\}^2 \sigma_R^2}{n_R} \quad (5)$$

The test statistic can be expressed as :

$$Z_{flexible} = \frac{\{\bar{X}_N - \bar{X}_R + \Delta_L(\bar{X}_R)\} - \{\mu_N - \mu_R + \Delta_L(\mu_R)\}}{\sqrt{\frac{\sigma_N^2}{n_N} + \frac{\{\Delta'_L(\mu_R) - 1\}^2 \sigma_R^2}{n_R}}}. \quad (6)$$

2.2.2 Asymptotic properties of test statistic $Z_{flexible}$.

From central limit theorem, when n_N and n_R approach infinity, the random variable $Z_{flexible} \sim N(0, 1)$, on the boundary of null hypothesis :

$$Z_{flexible} = \frac{\bar{X}_N - \bar{X}_R + \Delta_L(\bar{X}_R)}{\sqrt{\frac{\sigma_N^2}{n_N} + \frac{\{\Delta'_L(\mu_R) - 1\}^2 \sigma_R^2}{n_R}}} \sim N(0, 1). \quad (7)$$

μ_R is unknown, σ_R^2 and σ_N^2 can also be unknown and need to be estimated. We used maximum likelihood estimation method on the boundary of null hypothesis ($\mu_N = \mu_R - \Delta_L(\mu_R)$). The unknown parameters are estimated considering the cases where variances σ_R^2 and σ_N^2 are known or unknown, equal or not.

The maximum likelihood estimators $\hat{\mu}_R$, $\hat{\sigma}_R^2$ and $\hat{\sigma}_N^2$ are consistent for μ_R , σ_R^2 and σ_N^2 . Moreover, Δ'_L has been assumed continuous, thus, $\Delta'_L(\hat{\mu}_R)$ is a consistent estimator for $\Delta'_L(\mu_R)$. The estimator $\hat{Z}_{flexible}$ of the statistic $Z_{flexible}$ can be obtained, by replacing the unknowns parameters by the estimators. Therefore, the test H'_0 versus H_1 (where H'_0 is the boundary of H_0 i.e $\mu_N = \mu_R - \Delta_L(\mu_R)$), is rejected if $\hat{Z}_{flexible} > z_{1-\alpha}$, where α is the nominal type I error and $z_{1-\alpha}$ denote the $1 - \alpha$ percentile of a standard normal distribution. The significance level of this test tends to α , when n_N and n_R approach infinity.

Assuming that under alternative hypotheses H_1 , $\mu_N - \mu_R + \Delta_L(\mu_R) = v$, such that $v > 0$, denote by η the power, it follows that :

$$\begin{aligned} \eta &= \mathbf{P}\left(\frac{\bar{X}_N - \bar{X}_R + \Delta_L(\bar{X}_R)}{\sqrt{\frac{\sigma_N^2}{n_N} + \frac{(\Delta'_L(\mu_R) - 1)^2 \sigma_R^2}{n_R}}} > z_{1-\alpha}/H_1\right) \\ &= \mathbf{P}\left(\frac{\bar{X}_N - \bar{X}_R + \Delta_L(\bar{X}_R) - v}{\sqrt{\frac{\sigma_N^2}{n_N} + \frac{(\Delta'_L(\mu_R) - 1)^2 \sigma_R^2}{n_R}}} > \right. \\ &\quad \left. z_{1-\alpha} - \frac{v}{\sqrt{\frac{\sigma_N^2}{n_N} + \frac{(\Delta'_L(\mu_R) - 1)^2 \sigma_R^2}{n_R}}}\right) \end{aligned}$$

under alternative hypothesis, $\frac{\bar{X}_N - \bar{X}_R + \Delta_L(\bar{X}_R) - v}{\sqrt{\frac{\sigma_N^2}{n_N} + \frac{(\Delta'_L(\mu_R) - 1)^2 \sigma_R^2}{n_R}}} \sim N(0, 1)$. Thus, the power is given as a function of v , n_N , n_R , and α :

$$\eta(v, n_N, n_R) = \Phi\left(\frac{v}{\sqrt{\frac{\sigma_N^2}{n_N} + \frac{(\Delta'_L(\mu_R) - 1)^2 \sigma_R^2}{n_R}}} - z_{1-\alpha}\right) \quad (8)$$

Where Φ is the cumulative distribution function of standard normal distribution. For a fixed nominal type I error α and for any fixed μ_R and μ_N such that $v = \mu_N - \mu_R + \Delta_L(\mu_R) > 0$, when $n_R \rightarrow \infty$ and $n_N \rightarrow \infty$, then $\eta \rightarrow 1$, thus the test T is asymptotically convergent. From equation 8, it is possible to find the sample size that achieve the nominal fixed power. Denoting the nominal type II error by β and assuming that $n_N = rn_R$, with $r > 1$, the sample size which will allow nominal power $(1 - \beta)$ is such that:

$$n_R \geq \frac{(z_{1-\alpha} + z_{1-\beta})^2 [\sigma_N^2 + r\sigma_R^2 \{\Delta'_L(\mu_R) - 1\}^2]}{rv^2} \quad (9)$$

This formula is equivalent to the one found by [12] when the margin is fixed.

The proposed test statistic $\hat{Z}_{flexible}$ is an asymptotic test, which works well on the assumption of large sample sizes. However, in practice, one may not have data with large sample size, and therefore, it will be invalid to use the method based on the test statistic. The non-parametric test using the percentile bootstrap confidence interval test, which does not require any assumptions on sample size or distribution can be used as an alternative [14].

2.3 Approach based on confidence interval

For any test based on confidence interval, the focus is on the level of confidence interval required to achieve a desired nominal type I error. Moreover, as discussed in [12] and [15], the type I error is a controversial issue in clinical trial tests. In the framework of non-inferiority test, when the non-inferiority margin is fixed, [16] recommended using $1 - \alpha$ or $1 - \frac{\alpha}{2}$ for two sided and one sided confidence interval levels respectively. While, [10] recommended to use $1 - 2\alpha$ for two sided and $1 - \alpha$ for one sided confidence interval. In this section, we propose a non-parametric procedure for the confidence interval construction when non-inferiority margin is an unknown function of mean effect in new intervention group. The percentile bootstrap is used for the construction of confidence interval. The type I error and statistical power are investigated in the simulation study (section 3).

Due to overlapping confidence interval and statistical significance issues [17], the hypotheses tests in equation 1 are reformulated as follow :

$$\begin{cases} H_0 : \mu_N - \mu_R + \Delta_L(\mu_R) \leq 0 \\ H_1 : \mu_N - \mu_R + \Delta_L(\mu_R) > 0 \end{cases} \quad (10)$$

The procedure involves comparing the lower bound of the $\gamma\%$ confidence interval of $\mu_N - \mu_R + \Delta_L(\mu_R)$ with 0. The null hypothesis H_0 is rejected if the lower bound of confidence interval of $\mu_N - \mu_R + \Delta_L(\mu_R)$ is greater than 0.

Estimation of actual type I error is performed using simulations and non-parametric estimation of confidence interval on the boundary of null hypothesis. The detailed steps are described below.

1. From a fixed μ_R , calculate $\mu_N = \mu_R - \Delta_L(\mu_R)$ (satisfying the null hypothesis H_0). We assume that the standard deviations σ_N and σ_R are respectively known.
2. Let m denote the number of desired simulations, for $i \in \{1 \dots m\}$, simulate m pairs of samples X_N and X_R of size n_N and n_R respectively by normal distribution $\mathcal{N}(\mu_N, \sigma_N)$ and $\mathcal{N}(\mu_R, \sigma_R)$.
3. Using bootstrap, compute the empirical percentile confidence intervals $[a_i, b_i]$ of level γ for $\mu_N - \mu_R + \Delta_L(\mu_R)$, for $i \in \{1 \dots m\}$.
4. For $i \in \{1 \dots m\}$ H_0 is rejected when $a_i > 0$, thus the level of significance is estimated by : $\alpha(\gamma) = \frac{1}{m} \sum_{i=1}^m 1_{a_i > 0}$.

Like any other power estimation, the data are drawn under the alternative hypothesis that is : $\mu_N > \mu_R - \Delta_R(\mu_R)$. Since there is a wide range of possibilities on the alternative hypothesis, in the practice, one considers the equivalence point that is $\mu_R = \mu_N$. Therefore, similar to studies such as [9, 18], the equivalence point ($\mu_R = \mu_N$) will be used for drawing data for the power estimation.

1. From μ_R , simulate m pairs of samples X_N and X_R of size n_N and n_R respectively by normal distribution $\mathcal{N}(\mu_R, \sigma_N)$ and $\mathcal{N}(\mu_R, \sigma_R)$.
2. Using bootstrap, compute the empirical percentile confidence intervals $[a_i, b_i]$ of level γ for $\mu_N - \mu_R + \Delta_L(\mu_R)$, for $i \in \{1 \cdots m\}$.
3. For $i \in \{1 \cdots m\}$ H_0 is rejected when $a_i > 0$, thus the power is estimated by :

$$\eta(\gamma) = \frac{1}{m} \sum_{i=1}^m 1_{a_i > 0}.$$

2.4 Performances assessment

The simulations were done to evaluate the finite-sample performances of the asymptotic test and confidence interval based test. The performance indicators used were actual type I error and statistical power. Monte-Carlo simulation's techniques were used for the estimation of the considered indicators. The input parameters for the re-sampling of the samples included the non-inferiority margin function represented in figure 1 and the variances of both groups. Three scenarios about variances were considered:

- Equal and unknown variances $\sigma^2 = \sigma_R^2 = \sigma_N^2$. In the simulations it is considered $\sigma^2 = 190^2$,
- Unequal and unknown variances in favor of R , that is when $\sigma_R^2 > \sigma_N^2$. In the simulations it is considered $\sigma_R^2 = 2\sigma_N^2$, with $\sigma_N^2 = 190^2$,
- Unequal and unknown variances in favor of N , that is when $\sigma_N^2 > \sigma_R^2$. In the simulations it is considered $\sigma_N^2 = 2\sigma_R^2$, with $\sigma_R^2 = 190^2$.

Only the first scenario was considered for confidence interval based approach, and all the three scenarios were considered for the approach based on the asymptotic test

Both indicators were computed according to the active control treatment effect, considering the two proposed test approaches. For the actual type I error, data were drawn on the boundary of the null hypothesis: for a given μ_R , μ_N is obtained such that $\mu_N = \mu_R - \Delta_L(\mu_R)$. For the actual power, data were drawn under the alternative hypothesis: for a given μ_R , μ_N is obtained such that $\mu_N > \mu_R - \Delta_L(\mu_R)$, usually, one takes $\mu_N = \mu_R$. μ_R was taken based on the subdivision of the interval where its values were found, with an increment of 10^{-3} . For scaling purposes, we estimate performance indicators as a function of a active control intervention effect by using the scaled mean $u_R = \mu_R / \sigma_R$.

In the approach based on the asymptotic test, the nominal type I error was fixed at $\alpha = 2.5\%$. For the confidence interval based test, we considered three levels of confidence interval : 80%, 90% and 95%. The purpose was to estimate the actual type I error for the respective confidence intervals. In the simulations, we considered three cases for sample sizes : balanced sample sizes (that is when $n_N = n_R$), unbalanced sample sizes in favor of group N (that is when $n_N > n_R$) and unbalanced sample sizes in favor of group R (that is when $n_N < n_R$). For balanced sample sizes, we used $n_N = n_R = 30, 100$ and 1000 . In the case of unbalanced sample sizes, we considered $n_N = 3n_R$ and $n_R = 3n_N$ respectively for unbalanced in favor of N and R . The number of bootstrap samples with replacement was $B = 1000$ and the number of simulation replications was $m = 10000$. The **R** software programming language [19] was used to conduct the simulations and codes are accessible in a separate file.

2.5 A practical application

This study was motivated by the randomised and non-inferiority "Stratall ANRS 12110 / ESTHER" trial [20]. The main purpose was to assess an exclusively clinical monitoring strategy compared with a clinical monitoring strategy plus laboratory monitoring in terms of effectiveness and safety in HIV-infected patients in Cameroon. The idea was to achieve the scaling-up of HIV care in rural districts where most people live with HIV, but local health facilities generally have the low-grade equipment. A total of 459 HIV infected patients were included in the study and randomly allocated to two groups, one receiving exclusively clinical monitoring (Intervention group, $N=238$) and the other receiving Laboratory and clinical monitoring (active control group, $N=221$). All patients included were initiated antiretroviral treatment and were followed up for 24 months. Clinical monitoring alone was compared to laboratory and clinical monitoring in a non-inferiority design. The continuous primary endpoint was the increase in CD4 cells count from treatment initiation to the twenty-fourth month. Based on previous studies, the non-inferiority margin ($\Delta_L(R)$) was prespecified as a linear function (25 %) of mean CD4 cells increase (μ_R) after 24 months of antiretroviral treatment in laboratory and clinical monitoring group, $\Delta_L(R) = \frac{25}{100}\mu_R$. Unlike other non-inferiority studies [21, 22], non-inferiority margin in this study was variable (depends on the mean increase of CD4 in the active control group). However, the classical confidence interval based test with 95% and 90% level were used to obtain an actual type I error (α) close to 5% [20]. Indeed, the statistical test procedures that explore the non-inferiority test for continuous data with variable margin was not available. Then, the use of confidence interval based test, with 95% or 90% level, as in [20] can be misleading. Since, as discussed in [15], the relationship between the confidence interval level and the type I error can be controversial.

More details about the background of the study and the clinical trial process can be found in [20]. Two analysis were done according to the type of data :

- 1 First, the increase of CD4 count cells at 24 months from the baseline was considered, which implies the missing or lost patients before the end of follow-up period were excluded in the analysis. In that case, the total number of patient in the analysis reduced to $n = 334$, with $n_R = 169$ and $n_N = 165$. "Observed data" refers to the case where data are analyzed by excluding participants with missing observation at 24 months.
- 2 Second, an analysis was done with all participants who attended at least one follow-up visit, and the last observation carried forward (LOCF) imputation was applied for participants whose CD4 data were missing at 24 months.

The classical parametric two-sided confidence interval based test with 95% level was used by [20] to perform the non-inferiority test. The final result was that the CLIN was not non-inferior to the LAB.

3 Results

3.1 Simulations results

The outputs results for the approach based on the asymptotic test statistic are the combination of different cases of sample sizes(balanced sample sizes, and unbalanced sample sizes in favor of groups N and R) with the different variances scenarios (equal

variances, and unequal variances in favor of groups N and R). For all considered cases, results indicate that the simulated type 1 error appears to have the same trend according to μ_R . When $\mu_R < 190$, there is a slight deflation of simulated type I error of approximately 0.02, but for $\mu_R \geq 190$ the simulated type I error rate remains approximately equal to the nominal type I error 0.025 for small and average sample sizes. However, for the unbalanced sample sizes in favor of R , the simulated type 1 errors remains approximately equal to the nominal type 1 error for any value of μ_R . Results based on the power show that the power converges to 1 and remains constant when μ_R is varying for large sample sizes. Although the estimated actual type I error is more regular for unbalanced sample sizes in favor of R , the power is smaller compared to the case of sample sizes, in particular for unbalanced sample sizes in favor of N .

The results for the approach based on confidence interval are summarized in figures 11, 12 and 13 for balanced sample sizes, and unbalanced sample sizes in favor of groups N and R respectively. The goal was to determine the actual type I error when the confidence interval levels are: 80%, 90% and 95% respectively. The shape of the simulated type I error depends on the type of sample sizes and confidence interval level. Results indicate that for a confidence interval of 80%, the simulated type I error rate varies between 0.04 and 0.06. When $\mu_R < 190$, the actual type I error is centered around 0.05, and for $\mu_R \geq 190$ it is centered around 0.06. For confidence interval level of 90%, the simulated type I error rate vary between 0.02 and 0.025, when $\mu_R < 190$, the simulated type I error is centered around 0.02, and for $\mu_R \geq 190$ it is centered around 0.025. For confidence interval level of 95%, the simulated type I error rate varies between 0.01 and 0.015, when $\mu_R < 190$, the simulated type I error is centered around 0.01, and for $\mu_R \geq 190$ it is centered around 0.015. These relationships are well verified especially for balanced sample sizes and unbalanced sample sizes in favor of R . The type I error is more regular in the case of balanced sample sizes and can be approximated by 0.05, 0.025 and 0.01 respectively for confidence interval levels of 80%, 90% and 95%. These results suggest an approximate relationship of $\gamma = 1 - 4\alpha$ between confidence interval level and type I error for balanced sample sizes. However, for unbalanced sample sizes in favor of N , the simulated type I error seems to be more irregular and remains above 0.05, 0.025 and 0.01 respectively for confidence interval level of 80%, 90% and 95%. Regardless of the sample sizes (balanced or unbalanced), the 80% confidence interval level leads to a more powerful test compared to test with 90% and 95% confidence interval levels.

3.2 Practical application results

The proposed methods are applied by considering the two cases above for analysis. It is considered a linear and non-linear margin in the applications. The linear margin $\Delta_L(R) = \frac{25}{100}R$ is considered similar to [20]. The considered non-linear margin is represented in the figure 14. That has been determined by assuming that from the specific values 140, 180, 188, 190, 206 of the increase of CD4 count cells mean in the LAB group, the plausible acceptable margin could be 35, 46, 50, 53, 59. Then, the cubic spline is used to smooth and obtain the flexible margin. The results for the approach based on the test statistic are summarized in the table 1. The p-value

is calculated based on the test statistic in equation 6. The statistical power was computed using the obtained formula in equation 8 and based on the same inputs as in [20], that are $\mu_N = \mu_R = 140$, $\sigma_N = \sigma_R = 130$. For the linear margin, different results are obtained according to the type of data used (Observed and LOCF imputed data). In the case where observed data are used, with a p-value= 0.02, the null hypothesis that CLIN is not non-inferior to the LAB is rejected at a level of 0.025. While, with a p-value= 0.09, the null hypothesis that CLIN is not non-inferior to the LAB is failed to be rejected at a level of 0.025. However, for the non-linear margin, the conclusion is that CLIN is not non-inferior to the LAB.

For the confidence interval based approach, the test is performed by considering the 80%, 90% and 95% two-sided confidence interval levels, the results are presented in the table 2. Considering the test with linear margin, the null hypothesis that CLIN is not non-inferior to LAB is rejected when the observed data is used and for any confidence interval used. However, when the LOCF imputed data are used, the null hypothesis is rejected only when the test is performed with 80% confidence interval level.

Considering the test with non-linear margin, the null hypothesis is rejected when the test is performed with 80% confidence interval level. Excluding that case, any data used (observed or LOCF imputed data), any margin used (linear and non-linear) and any other confidence interval level used, the results are that the null hypothesis that CLIN is not non-inferior to LAB fails to be rejected.

The conclusion about the non-inferiority of clinical monitoring alone to laboratory plus clinical monitoring not only depends on the choice of margin but also on the level of confidence interval level used and sample sizes. As we showed in the simulations section, confidence interval based test with 80% confidence interval level was the most powerful. The results with 80% confidence interval level were discordant with 90% and 95% confidence interval level results by considering linear margin. However, due to the irregularity of the actual type I error in 80% confidence interval level case, the result can be confusing.

We can conclude on the same result as in [20] that clinical monitoring alone was not non-inferior to laboratory plus clinical monitoring, by considering the linear margin and LOCF imputed data. That is, by considering either proposed asymptotic test and confidence interval based test with levels 90% and 95%.

4 Discussions

In this study we have proposed two non-inferiority test approaches for a continuous variable, with a flexible margin, using mean differences: an asymptotic test based on the delta method and an approach based on non-parametric confidence interval. Confidence interval approach is more used in literature and recommended by the international guideline [2]. For the non-inferiority test with continuous endpoint and fixed or linear margin, some studies like [10] and [23] investigated the confidence interval approach. However, that does not allow for explicit calculation of the sample size. Comparatively, the proposed asymptotic test approach allows explicit calculation of sample size and power formulas.

As discussed in [12] and [15], the choice of type I error according to confidence interval is a controversial issue. The results of simulations for the confidence interval

based test show that the confidence interval level determines approximately the type I error rate. The test with two-sided confidence interval level of 80%, 90% and 95% leads to a type I error which can be approximated by 0.05, 0.025 and 0.01 respectively. Though the test with 80% confidence interval level is more powerful compared to the ones obtained by 90% and 95%, it is observed that the actual type I error in this case is more irregular. Therefore, in practice, the choice of the confidence interval level that should be used must be selected as a compromise between the type I error and power.

The proposed asymptotic test is asymptotically convergent. Also, for a fixed nominal type I error equal to 0.025, the actual type I error remains approximately 0.02 and 0.03. An inflection is observed when the sampling dispersion in the active control intervention group is around 1. These results are similar to those obtained by [9] for binary endpoint and highlight the relevance of variable margin as a methodological issue. As shown in [24] and [25], the roughness of the points-to-consider margin is the primary reason for the inflation of the type I error.

The proposed methods in this study can be viewed as a generalization of the case where the non-inferiority margin is constant for a continuous endpoint. However, we highlight the fact that the asymptotic test approach uses numerical techniques such as Newton-Raphson to obtain approximate solutions. The results obtained may differ according to approximation techniques, the accuracy parameters such as tolerance and number of sub-interval are considered.

The approach based on confidence interval is more suitable than the asymptotic test for small sample sizes. Indeed, with small sample sizes, by considering the 80% confidence interval level and all other things remaining equal, the test is more powerful compared to the asymptotic test. Hence, the two proposed approaches are complementary.

5 Conclusions

In an active-controlled trial of non-inferiority, the non-inferiority margin should be a function of reference treatment. This paper produced a framework on how to perform the non-inferiority hypothesis test with a variable margin. The proposed non-inferiority hypothesis test procedures got good performances and are applicable in practice, a practical application on real clinical data was illustrative.

Abbreviations

CD4: Cluster of Differentiation 4; CLIN: Clinical monitoring alone; HIV/AIDS: Human immunodeficiency virus infection and acquired immune deficiency syndrome; LAB: Laboratory and clinical monitoring; LOCF : Last Observation Carried Forward.

Acknowledgements

ABS is grateful to the African Union, he was a recipient of a full scholarship for his doctoral studies.

Authors' contributions

ABS and JBMT drafted the manuscript, proposed methods and analyzed the data. NM, CK and CL produced real clinical data, read and edited the manuscript, and provided observations. AW read, edited the manuscript and provided observations.

Funding

No funding was obtained for this study.

Availability of data and materials

The datasets used and analysed during the current study are available from the corresponding author or the or the author named Christian Laurent (christian.laurent@ird.fr) on reasonable request.

Ethics approval and consent to participate

This study involved an analysis of data that was already analyzed in a primary research work. A confidential agreement was done with the main investigators.

Consent for publication

Not applicable.

Competing interests

The authors declare that they have no competing interests.

Author details

¹Department of Mathematics-Statistics, Pan African University Institute for basic Sciences, Technology and Innovation (PAUSTI/JKUAT), Nairobi, Kenya. ²IMAG,CNRS, Univ Montpellier, CHU Montpellier, Montpellier, France. ³Department of Statistics and Actuarial Sciences, Jomo Kenyatta University of Agriculture and Technology (JKUAT), Nairobi, Kenya. ⁴Day Hospital, Yaounde Central Hospital, Yaounde, Cameroon. ⁵IRD, INSERM, Univ Montpellier, TransVIHMI, Montpellier, France. ⁶Service d'épidémiologie et de santé publique du Centre Pasteur du Cameroun, Membre du Réseau International des Instituts Pasteur, Yaoundé, Cameroun. ⁷Université de Yaoundé I - CETIC, UPMC Université Paris 06, IRD, Unité de Modélisation Mathématique et Informatique des Systèmes Complexes (UMMISCO), Bondy, France.

References

- Rothmann MD, Wiens BL, Chan IF. Design and Analysis of Non-Inferiority Trials. Boca Raton: Taylor and Francis Group; 2012.
- Food and Drug Administration. Non-inferiority clinical trials to establish effectiveness-Guidance for industry. U.S. Department of Health and Human Services; 2016.
- Committee for Medicinal Products for Human Use, Efficacy Working Party, Committee for Release for Consultation. Committee for Medicinal Products for Human Use (CHMP) guideline on the choice of the non-inferiority margin. Stat Med. 2006 may;25(10):1628–38.
- Gladstone BP, Vach W. Choice of Non-Inferiority (NI) Margins Does Not Protect against Degradation of Treatment Effects on an Average – An Observational Study of Registered and Published NI Trials. PLoS One. 2014 jul;9(7):e103616.
- Althunian TA, Boer AD, Klungel OH, Insani WN, Groenwold RHH. Methods of defining the non-inferiority margin in randomized, double-blind controlled trials: a systematic review. Trials. 2017;.
- Ng T. Noninferiority hypotheses and choice of noninferiority margin. Statistics in Medicine. 2008;27:5392–5406.
- Phillips KF. A new test of non-inferiority for anti-infective trials. Statistics in Medicine. 2003;22:201–212.
- Kim MY, Xue X. Likelihood ratio and a Bayesian approach were superior to standard noninferiority analysis when the noninferiority margin varied with the control event rate. Journal of Clinical Epidemiology. 2004;57:1253–1261.
- Zhang Z. Non-Inferiority Testing with a Variable Margin. Biometrical Journal. 2006;48:948–965.
- Elie C, Rycke YD, Jais JP, Marion-Gallois R, Landais P. Methodological and statistical aspects of equivalence and non inferiority trials. Revue d'Épidémiologie et de Santé Publique. 2008;56:267–277.
- Tsong Y, Wang SJ, Hung HM, Cui L. Statistical issues on objectives, designs and analysis of non-inferiority test active controlled clinical trials. Journal of Biopharmaceutical Statistics. 2003;13:29–41.
- Julious SA. Sample sizes for clinical trials with Normal data. Statistics in Medicine. 2004;23:1921–1986.
- Casella G, Berger RL. Statistical Inference. 2nd ed. USA: Duxbury Advanced Series; 2002.
- Good P. Permutation, Parametric and Bootstrap Tests of Hypothesis. New-York: Springer; 2005.
- Wellek S. Testing Statistical Hypotheses of Equivalence and Noninferiority. 2nd ed. Boca Raton: Taylor and Francis Group; 2010.
- Committee for Proprietary Medicinal Products. Point to Consider on switching between superiority and non-inferiority; 2000. European Medicines Agency (EMA).
- Knezevic A. Overlapping Confidence Intervals and Statistical Significance. Cornell Statistical Consulting Unit Newsletter; 2008.
- Flight L, Julious SA. Practical guide to sample size calculations: non-inferiority and equivalence trials. Pharmaceutical Statistics. 2016;(9):80–89.
- R Core Team. A language and environment for statistical computing. R Foundation for Statistical Computing; 2018.
- Laurent C, Kouanfack C, Laborde-Balen G, Aghokeng AF, j B Mbougua, Boyer S, et al. Monitoring of HIV viral loads, CD4 cell counts, and clinical assessments versus clinical monitoring alone for antiretroviral therapy in rural district hospitals in Cameroon (Stratall ANRS 12110/ESTHER): a randomised non-inferiority trial. Lancet Infect Dis. 2011;.
- Mugenyi P, Walker AS, Hakim J, Munderi P, Gibb DM, Kityo C, et al. Routine versus clinically driven laboratory monitoring of HIV antiretroviral therapy in Africa (DART): a randomised non-inferiority trial. The Lancet. 2010;375(9709):123 – 131.
- Sanne I, Orrell C, Fox MP, Conradie F, Ive P, Zeinecker J, et al. Nurse versus doctor management of HIV-infected patients receiving antiretroviral therapy (CIPRA-SA): a randomised non-inferiority trial. The Lancet. 2010;376(9734):33 – 40.
- Tchatchueng-Mbougua JB. Problématiques rencontrées dans l'étude du traitement antiretroviral des adultes infectés par le VIH en Afrique subsaharienne. Université de Montpellier I; 2012.
- Weng CSW, Liu JP. Some pitfalls in sample size estimation for an anti-infective study. Proceedings of the Pharmaceutical Section. 1994;p. 56–60.
- Rohmel J. Statistical considerations of FDA and CPMP rules for the investigation of new anti-bacterial products. Statistics in Medicine. 2001;20:2561–2571.

Figures

Figure 1 Non-inferiority margin used in the simulations.

The curve representing the non-linear non-inferiority margin used in the simulations.

Figure 2 Type I error and power estimation for balanced sample sizes and equal variances

Estimation of type I error rate (left) and power (right), as functions of active control treatment (scaled as μ_R/σ_R), with flexible margin, balanced sample sizes and equal variances, for the approach based on the asymptotic test statistic. Group sizes are $n_N = n_R = 30$ (top row), 100 (middle row), 1000 (bottom row).

Figure 3 Type I error and power estimation for balanced sample sizes and unequal variances

Estimation of type I error rate (left) and power (right), as functions of active control treatment (scaled as μ_R/σ_R), with flexible margin, balanced sample sizes, and unequal variances in favor of N , for the approach based on the asymptotic test statistic. Group sizes are $n_N = n_R = 30$ (top row), 100 (middle row), 1000 (bottom row)

Figure 4 Type I error and power estimation for balanced sample sizes and unequal variances

Estimation of type I error rate (left) and power (right), as functions of active control treatment (scaled as μ_R/σ_R), with flexible margin, balanced sample sizes, and unequal variances in favor of R , for the approach based on the asymptotic test statistic. Group sizes are $n_N = n_R = 30$ (top row), 100 (middle row), 1000 (bottom row).

Figure 5 Type I error and power estimations for unbalanced sample sizes and equal variances

Estimation of type I error rate (left) and power (right), as functions of active control treatment (scaled as μ_R/σ_R), with a flexible margin, unbalanced sample sizes in favor of group N and equal variances, for the approach based on the asymptotic test statistic. Group sizes are $n_N = 90, n_R = 30$ (top row), $n_N = 300, n_R = 100$ (middle row), $n_N = 900, n_R = 300$ (bottom row).

Figure 6 Type I error and power estimations for unbalanced sample sizes and unequal variances

Estimation of type I error rate (left) and power (right), as functions of active control treatment (scaled as μ_R/σ_R), with a flexible margin, unbalanced sample sizes in favor of group N and unequal variances in favor of R , for the approach based on the asymptotic test statistic. Group sizes are $n_N = 90, n_R = 30$ (top row), $n_N = 300, n_R = 100$ (middle row), $n_N = 900, n_R = 300$ (bottom row).

Figure 7 Type I error and power estimations for unbalanced sample sizes and unequal variances

Estimation of type I error rate (left) and power (right), as functions of active control treatment (scaled as μ_R/σ_R), with a flexible margin, unbalanced sample sizes in favor of group N and unequal variances in favor of N , for the approach based on the asymptotic test statistic. Group sizes are $n_N = 90, n_R = 30$ (top row), $n_N = 300, n_R = 100$ (middle row), $n_N = 900, n_R = 300$ (bottom row).

Tables

Figure 8 Type I error and power estimations for unbalanced sample sizes and equal variances

Estimation of type I error rate (left) and power (right), as functions of active control treatment (scaled as μ_R/σ_R), with a flexible margin, unbalanced sample sizes in favor of group R and equal variances, for the approach based on the asymptotic test statistic. Group sizes are $n_N = 30, n_R = 90$ (top row), $n_N = 100, n_R = 300$ (middle row), $n_N = 300, n_R = 900$ (bottom row).

Figure 9 Type I error and power estimations for unbalanced sample sizes and unequal variances

Estimation of type I error rate (left) and power (right), as functions of active control treatment (scaled as μ_R/σ_R), with a flexible margin, unbalanced sample sizes in favor of group R and unequal variances in favor of N , for the approach based on the asymptotic test statistic. Group sizes are $n_N = 30, n_R = 90$ (top row), $n_N = 100, n_R = 300$ (middle row), $n_N = 300, n_R = 900$ (bottom row).

Figure 10 Type I error and power estimations for unbalanced sample sizes and unequal variances

Estimation of type I error rate (left) and power (right), as functions of active control treatment (scaled as μ_R/σ_R), with a flexible margin, unbalanced sample sizes in favor of group R and unequal variances in favor of R , for the approach based on the asymptotic test statistic. Group sizes are $n_N = 30, n_R = 90$ (top row), $n_N = 100, n_R = 300$ (middle row), $n_N = 300, n_R = 900$ (bottom row).

Figure 11 Type I error and power estimations for balanced sample sizes in confidence interval based test

Estimation of type I error rate (left in each block) and power (right in each block), as function of active control treatment (scaled as μ_R/σ_R), with a flexible margin and balanced sample sizes, for the approach based on confidence interval and grouped according to the confidence interval level. Group confidence interval level are 80% (first group block), 90% (second group block), 95% (third group block). Group sizes are $n_N = n_R = 30$ (top row), 100 (middle row), 1000 (bottom row).

Figure 12 Type I error and power estimations for unbalanced sample sizes in confidence interval based test

Estimation of type I error rate (left in each block) and power (right in each block), as function of active control treatment (scaled as μ_R/σ_R), with a flexible margin and unbalanced samples sizes in favor of group N , for the approach based on confidence interval and grouped according to the confidence interval level. Group confidence interval level are 80% (first group block), 90% (second group block), 95% (third group block). Group sizes are $n_N = 90, n_R = 30$ (first row), $n_N = 300, n_R = 100$ (middle row), $n_N = 900, n_R = 300$ (bottom row).

Figure 13 Type I error and power estimations for unbalanced sample sizes in confidence interval based test

Estimation of type I error rate (left in each block) and power (right in each block), as function of active control treatment (scaled as μ_R/σ_R), with a flexible margin and unbalanced samples sizes in favor of group R , for the approach based on confidence interval and grouped according to the confidence interval level. Group confidence interval level are 80% (first group block), 90% (second group block), 95% (third group block). Group sizes are $n_N = 30, n_R = 90$ (first row), $n_N = 100, n_R = 300$ (middle row), $n_N = 300, n_R = 900$ (bottom row).

Figure 14 Non-linear margin used in the practical application

The curve of non-linear margin used in the practical application

Table 1 P-value and power determination for the approach based on the asymptotic test statistic and according to the data used

	<i>p</i> – value	Power, 0.025	Power, 0.05
Linear margin			
Case of LOCF	0.02	0.77	0.86
Case of observed data	0.11	0.82	0.89
Non-linear margin			
Linear margin	0.09	0.91	0.95
Non-linear margin	0.07	0.93	0.96

Table 2 Confidence interval calculations and decision on non-inferiority confidence interval based test, with linear margin and non-linear margin

	80% CI	90% CI	95% CI
Linear margin			
$CLIN - LAB + \Delta_L(LAB)$	1 to 41	Case of LOCF -5 to 47	-10 to 52
Decision	Non-inferiority	No non-inferiority	No non-inferiority
$CLIN - LAB + \Delta_L(LAB)$	13 to 59	Case of observed data 7 to 67	1 to 72
Decision	Non-inferiority	Non-inferiority	Non-inferiority
Non linear margin			
$CLIN - LAB + \Delta_L(LAB)$	4 to 48	Case of LOCF -3 to 53	-10 to 58
Decision	Non-inferiority	No non-inferiority	No non-inferiority
$CLIN - LAB + \Delta_L(LAB)$	-1 to 55	Case of observed data -10 to 63	-17 to 70
Decision	No non-inferiority	No non-inferiority	No non-inferiority