

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 Sandie (✉ sandiearsene@gmail.com)

Pan African University Institute of Basic Science, Technology and Innovation

Nicholas Molinari

IMAG, CNRS, Universite de Montpellier

Anthony Wanjoya

Jomo Kenyatta University of Agriculture and Technology

Charles Kouanfack

Day Hospital

Christian Laurent

IRD, Inserm

Jules Brice Tchatchueng-Mbougua

Centre Pasteur Du Cameroun

Research

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

Posted Date: December 5th, 2020

DOI: <https://doi.org/10.21203/rs.3.rs-21325/v2>

License:   This work is licensed under a Creative Commons Attribution 4.0 International License.

[Read Full License](#)

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
 Department 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 the non-inferiority margin should be accounted for in statistical tests. In this work, we explored how to perform the non-inferiority test with a flexible margin for continuous endpoint.

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 test statistic 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, which the purpose was to assess the efficacy of clinical monitoring alone versus laboratory and clinical monitoring in HIV-infected adult patients.

Results: Basically, the two proposed test procedures have good properties. In the test based on a statistic, the actual type I error rate estimate is approximately equal to the nominal value. It has been found that the confidence interval level determines approximately the level of significance. The 80%, 90%, and 95% one-sided confidence interval levels led approximately to a type I error of 10%, 5% and 2.5% respectively. The power estimate was almost 100% for two proposed tests, except for the small scale values of the reference treatment where the power was relatively low when the sample sizes were small.

Conclusions: Based on type I error rate and power estimates, the proposed non-inferiority hypothesis test procedures have good performance and are applicable in 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 efficacy 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 efficacy 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 the main effect of the active control intervention compared to new intervention's benefits (fewer

side effects, costs, tolerable and safer). This acceptable loss of efficacy refers to the *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 Food and Drug Administration(FDA)[2] provides general principles for an appropriate choice of the non-inferiority margin. The non-inferiority margin is at the upper limit of the confidence interval, so the trial is designed to show evidence of no more than this 'loss of maximum efficacy'. Basically, the non-inferiority margin should be a function of treatment effect(flexible margin), and the uncertainty surrounding the margin should be accounted for while performing non-inferiority hypothesis tests. For binary endpoint, tests that accounts for a flexible margin have been studied [3, 4, 5]. 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, 7, 8], mainly consisting of comparing the bounds of the treatments difference to the fixed margin. However, few studies have been performed for a flexible or variable margin for continuous endpoints. The aim of this study was to derive non-inferiority tests for continuous endpoints with a flexible margin in active randomized controlled trials. An application 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 efficacy. 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 [9], 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 [10].

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 the Delta method for 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 [9] 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 [11].

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 [9] and [12], 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, [13] recommended using $1 - \alpha$ or $1 - \frac{\alpha}{2}$ for two sided and one sided confidence interval levels respectively. While, [7] 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).

An intuitive procedure based on confidence interval for hypothesis test in equation 1 would be by checking the overlapping of the confidence intervals of μ_N and $\mu_R - \Delta_L(\mu_R)$: The null hypothesis would be rejected if the two confidence intervals are non-overlap, and not rejected otherwise. In such a way, as illustrated in [14], the intervals may be overlapped while the statistics would not be necessarily non significantly different, thus, the power of the test would be lower. Due to overlapping confidence interval and statistical significance issues, 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 $i \in \{1 \cdots m\}$ 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 : $\mu_N > \mu_R - \Delta_L(\mu_R)$. 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 [5, 15], 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. In the simulations, it is considered the simple linear margin $\Delta_L(\mu_R) = 0.25 * \mu_R$; equal and unknown variances $\sigma^2 = \sigma_R^2 = \sigma_N^2$, $\sigma^2 = 1$;

Both indicators were computed for both proposed tests according to the active control treatment effect. 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 $\mu_N = \mu_R$. In all cases, it is assumed that μ_R vary in $[1, 1000]$.

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 interval levels. In all the simulations, we considered balanced sample sizes (that is when $n = n_N = n_R$), $n = 20, 100$ and 500 for small medium and large sample sizes respectively. The number of bootstrap samples with replacement was $B = 1000$ and the number of simulation replications was $m = 10000$. The **R** software programming language [16] was used to conduct the simulations and codes are accessible in a separate file on request.

2.5 A practical application

This study was motivated by the randomised and non-inferiority "Stratall ANRS 12110 / ESTHER" trial [17]. 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 [18, 19], 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% [17]. Indeed, the statistical test procedures that explore the non-inferiority test for continuous data with variable margin were not available. Then, the use of confidence interval based test, with 95% or 90% level, as in [17] can be misleading. Since, as discussed in [12], 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 [17]. 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 [17] 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

3.1.1 Test statistic based test

The outputs results for the approach based on the asymptotic test statistic are summarized in figure 1 and 2 for type I error rate and power estimates respectively. For small sample size, the actual type I error rate estimate is slightly above the nominal value, the median value estimate is 0.026, with the Interquartile Range(IQR) of [0.025 – 0.027]. As the sample size increase, the actual type I error estimates get closer to the nominal value. For medium sample size $n = 100$, the type I error estimate closer to the nominal value, the median value estimate for all value of μ_R is 0.025 (IQR = [0.024 – 0.026]). For large sample sizes $n = 500$, the actual type I error estimate is more accurate and closest to the nominal value, the median estimate is 0.025 (IQR = [0.025 – 0.026]). Concerning the power estimates, Except

at the neighborhood of 1 for small and medium sample sizes 20 and 100, the power rate is almost 100% whatever the value of μ_R .

3.1.2 Confidence interval based test

The results for the approach based on confidence interval are summarized in figures 3 and 4 for type I error rate and power estimates respectively. For confidence interval levels of 95%, 95% and 80%, the actual type I error estimate tends to 0.025, 0.05, and 0.1 respectively as the sample size increase. For small and medium sample sizes 20 and 100 the actual type I error rate estimates are slightly above the respective nominal values. The power estimate has a similar behavior as in the test statistic based test.

3.2 Practical application results

The proposed methods are applied by considering Observer data and LOCF data. with linear margin $\Delta_L(R) = \frac{25}{100}R$. 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 [17], that were $\mu_N = \mu_R = 140$, $\sigma_N = \sigma_R = 130$. In the case of observed data the p-value estimate was = 0.02, the null hypothesis that CLIN is not non-inferior to the LAB is rejected at a level of 0.025. While, for the LOCF data, the p-value was = 0.09, the null hypothesis that CLIN is not non-inferior to the LAB was not rejected at a level of 0.025.

For the confidence interval based approach, the test was performed by considering the 80%, 90% and 95% one-sided confidence interval levels, the results are presented in the table 2. Considering the test with linear margin, the null hypothesis that CLIN was not non-inferior to LAB is rejected when the observed data was used and for any confidence interval used. However, when the LOCF data were used, the null hypothesis was rejected only when the test is performed with 80% confidence interval level. Basically, the obtained results are similar and consistant with [17] : the clinical monitoring alone was not non-inferior to laboratory plus clinical monitoring.

4 Discussions

In this study we have proposed two non-inferiority test approaches for a continuous endpoints with 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 margin, some studies like [7] and [20] investigated the confidence interval approach. However, that does not allow for explicit sample size calculation. Comparatively, the proposed test statistic approach allows explicit calculation of sample size and power formulas. As discussed in [9] and [12], the choice of type I error according to confidence interval is a controversial issue. The results of simulations for the confidence interval based test showed that the confidence interval level determined approximately the type I error rate. The test with one-sided confidence

interval level of 80%, 90% and 95% leads to a type I error which can be approximated by 0.1, 0.05 and 0.025 respectively. These findings are consistent with those obtained in [13].

The proposed methods in this study can be viewed as a generalization of the case where the non-inferiority margin is fixed for continuous endpoint. Even though the proposed tests would be slightly liberal for small and medium sample sizes, the actual type 1 error estimate was approximately equal to the nominal type 1 error rate. The weakness of the power observed at the neighborhood of 1 may imply some limitations of the proposed test. Therefore, the proposed methods for outcomes on small scale should be used cautiously especially for small sizes.

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. Based on type I one error rate and power estimates, the proposed non-inferiority hypothesis test procedures have 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 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.
- 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.
- Ng T. Noninferiority hypotheses and choice of noninferiority margin. *Statistics in Medicine*. 2008;27:5392–5406.
- 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, Mbougua B, 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.
- 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.

Figures

Figure 1 Type I error rate estimates according to sample sizes for test statistic based test
 Estimation of type I error rate as functions of reference treatment, for the test statistic based test from the right to the left sample sizes are $n_N = n_R = 20, 100$ and 500 respectively.

Figure 2 Power rate estimates according to sample sizes for test statistic based test
 Power rate estimates as functions of reference treatment, for the test statistic based test. From the right to the left, sample sizes are $n_N = n_R = 20, 100$ and 500 respectively.

Figures

Figure 3 Type I error rate estimates according to sample sizes for the confidence interval based test
 Estimation of type I error rate as functions of reference treatment, for the the confidence interval based test. From the right to the left sample sizes are $n_N = n_R = 20, 100$ and 500 respectively. From the top to bottom are confidence interval level : 95% 90% and 80% respectively.

Tables

Figure 4 Power estimates according to sample sizes for the confidence interval based test
 Power rate estimates as functions of reference treatment, for the the confidence interval based test. From the righth to the left sample sizes are $n_N = n_R = 20, 100$ and 500 respectively. From the top to bottom are confidence interval level : 95% 90% and 80% respectively.

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
Case of LOCF	0.02	0.77
Case of observed data	0.11	0.82

Table 2 Confidence interval calculations and decision on non-inferiority confidence interval based test

	80% CI	90% CI	95% CI
$CLIN - LAB + \Delta_L(LAB)$	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

Figures

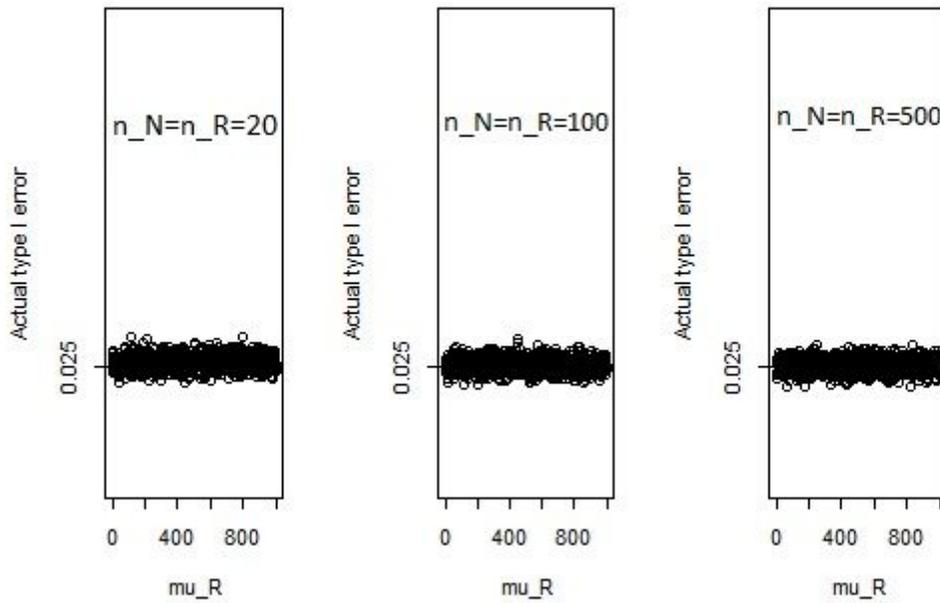


Figure 1

Type I error rate estimates according to sample sizes for test statistic based test Estimation of type I error rate as functions of reference treatment, for the test statistic based test from the right to the left sample sizes are $n_N = n_R = 20, 100$ and 500 respectively.

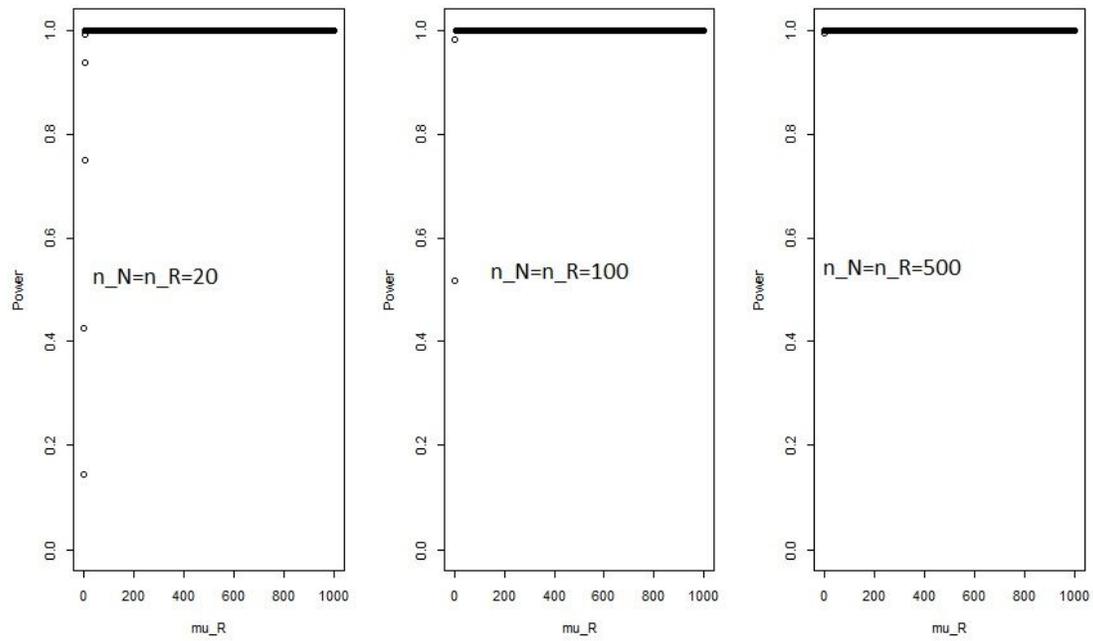


Figure 2

Power rate estimates according to sample sizes for test statistic based test Power rate estimates as functions of reference treatment, for the test statistic based test. From the right to the left, sample sizes are $n_N = n_R = 20, 100$ and 500 respectively.

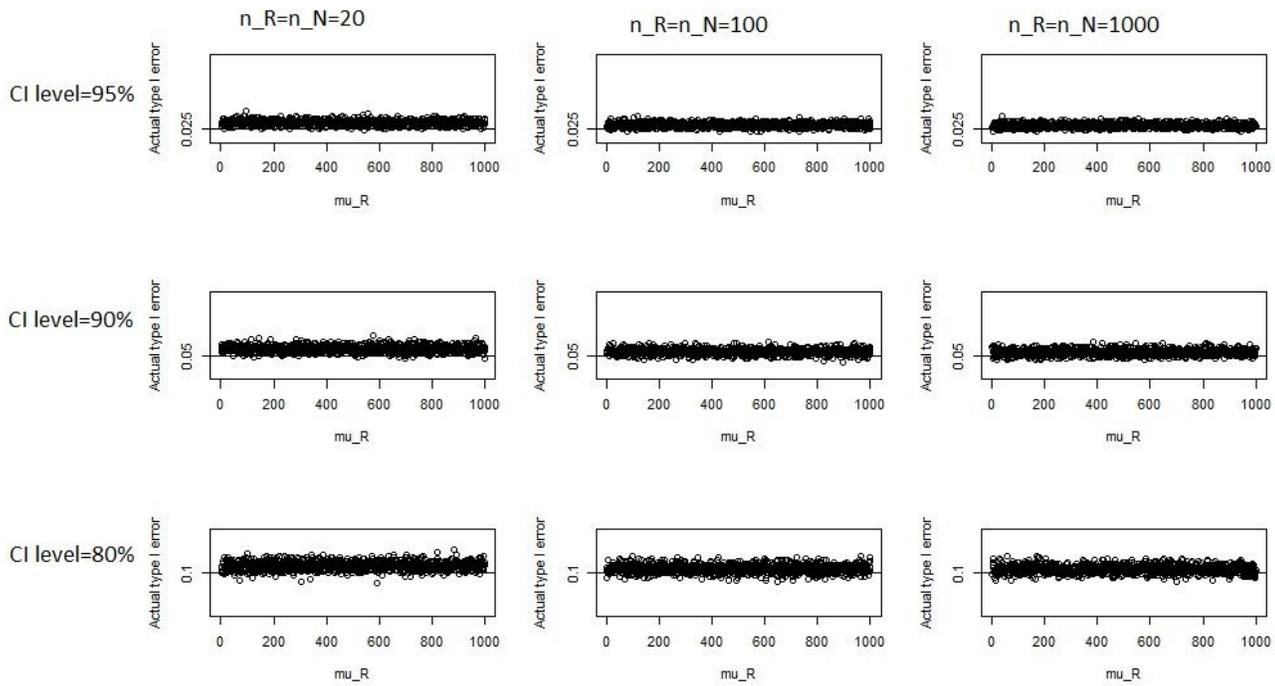


Figure 3

Type I error rate estimates according to sample sizes for the condence interval based test Estimation of type I error rate as functions of reference tretment, for the the condence interval based test. From the rigth to the left sample sizes are $n_N = n_R = 20, 100$ and 500 respectively. From the top to bottom are condence interval level : 95% 90% and 80% respectively.

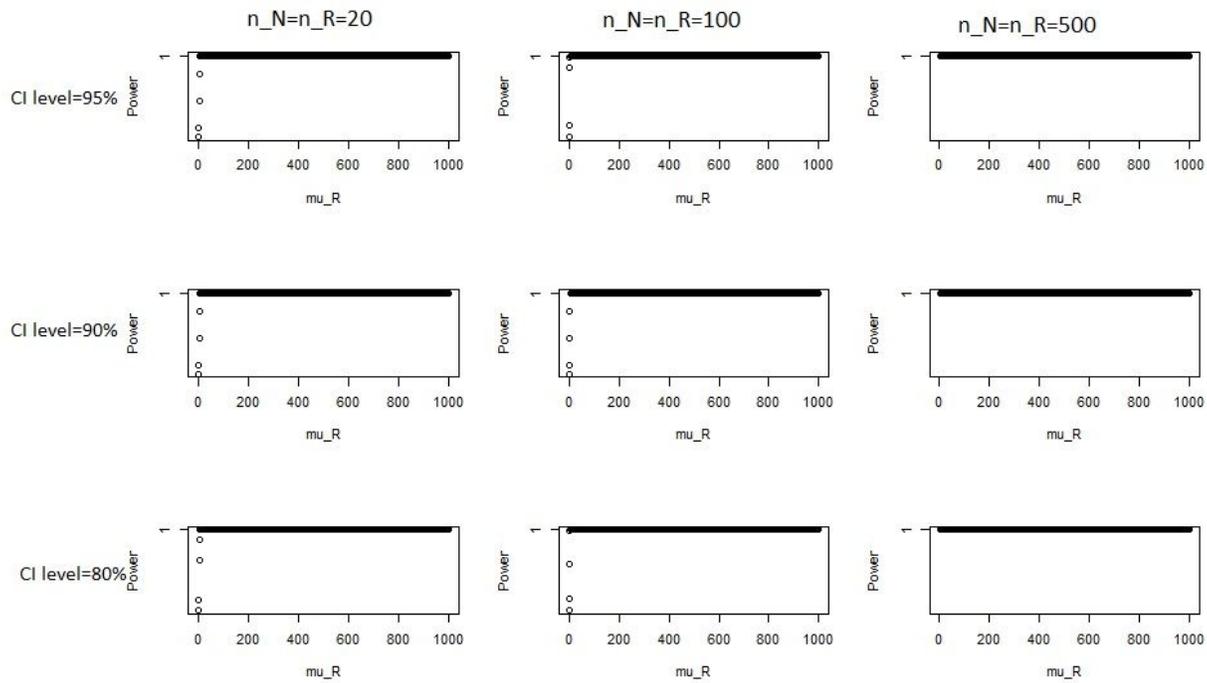


Figure 4

Power estimates according to sample sizes for the condence interval based test Power rate estimates as functions of reference tretment, for the the condence interval based test. From the rigth to the left sample sizes are $n_N = n_R = 20, 100$ and 500 respectively. From the top to bottom are condence interval level : 95% 90% and 80% respectively.