

Original Research Paper

Functional Non-Inferiority Hypothesis Testing for Longitudinal Data

¹Arsène Brunelle Sandie, ²Anthony Wanjoya and ^{3,4}Jules Brice Tchatchueng-Mbouguia

¹Department of Mathematics-Statistics, Pan African University Institute of basic Science, Technology and Innovation (PAUSTI)/Jomo Kenyatta University of Agriculture and Technology (JKUAT), Nairobi, Kenya

²Department of Statistics and Actuarial Sciences,

Jomo Kenyatta University of Agriculture and Technology (JKUAT), Nairobi, Kenya

³Service d'épidémiologie et de Santé Publique du Centre Pasteur du Cameroun,

Membre du Réseau International des Instituts Pasteur, Yaoundé, Cameroun

⁴Université de Yaoundé I - CETIC, UPMC Université Paris 06, IRD,

Unité de Modélisation Mathématique et Informatique des Systèmes Complexes (UMMISCO), F-93143, Bondy, France

Article history

Received: 14-05-2019

Revised: 01-08-2019

Accepted: 28-08-2019

Corresponding Authors:

Dr. Jules Brice Tchatchueng-Mbouguia

Service d'épidémiologie et de Santé Publique du Centre Pasteur du Cameroun, Membre du Réseau International des Instituts Pasteur, Yaoundé, Cameroun

Email: tchatchueng@pasteuryaounde.org

Abstract: The study pattern of non-inferiority trials is increasingly used to show the non-inferiority of new health intervention. Although in such studies the data are longitudinally collected (data held over a period of time), the conclusion of these non-inferiority trials is based on data observed at a specific time during the study period (usually at the end of the study period). In this study, we present a method that takes into account all the data observed during the study period to perform non-inferiority test. Thus, we approximate the observed data on a statistical unit by a function of time. This allows to transform the observed data on a time grid into functional data on a continuum domain. Although it could have some relevant applications, the functional data analysis for non-inferiority test has not been addressed. In this study, the functional non-inferiority hypothesis testing has been introduced. The optimal point-wise test and simultaneous confidence bands have been adapted and adopted for the purpose. The assessment of the methods has been done through simulations example. Both methods have good performances for large sample sizes. For small sample sizes, the optimal point-wise test would be too conservative while the simultaneous confidence bands based test would be a bit liberal.

Keywords: Confidence Bands, False Discovery Rate, Functional Data, Non-Inferiority Test, Pointwise Test

Introduction

The study design of non-inferiority randomized cohort trials is increasingly applied to show the non-inferiority of new health interventions (Ng, 2015). One of the advantages of such a study scheme is the longitudinal aspect of the collected data. Indeed, this study scheme makes it possible to collect repeated measurements on the people included in the study during the follow-up period. Depending on the follow-up duration and the delay between measurements, the number of repeated measurements per person could be important. Thus, longitudinal trials allow to get an array of data on the variation of the main endpoint on a

predefined time grid. The general goal of the non-inferiority trials is to show the non-inferiority of a new health intervention compared to a reference intervention. The evaluation may be related to the variation of the main endpoint in the whole study period. We situate our work, in this framework, motivated by the non-inferiority randomized trial conducted by Laurent *et al.* (2011). However, although the interest is on the variation of the endpoint on the follow-up period, the assessment of non-inferiority is carried out at a precise moment during the study (generally at the end of the study), thus reducing the problem of the non-inferiority hypothesis testing in the finite dimension. Indeed, in the finite dimension, the

non-inferiority test is a well-studied problem. However, the infinite dimensional functional case processes have additional difficulties. In this work, we will adopt functional data analysis approach to perform the non-inferiority hypothesis testing.

Although in practice the data are recorded discretely, with large measurement errors, in functional data analysis the data are treated as though they are in the form of a curve. Observed data are subject to a pre-processing steps, usually based on local polynomial or spline methods (Eubank, 1999; Wand and Jones, 1995; Green and Silverman, 1994; Ruppert *et al.*, 2003), to transform them to the smooth curve from which the methods of functional data analysis are applied. In many instances, the pre-processing step is not of great importance. However, some studies have shown that it has the potential to significantly reduce power Hall and Keilegom (2007).

There has been a prosperous period in the development of functional hypothesis testing procedures to deal with simple hypothesis (Darlin, 1957; Johnson and Kotz, 1990). However, the more common situation involving composite hypothesis is more challenging. Little development includes functional composite hypothesis testing and still less for functional non-inferiority hypothesis test. The point-wise test, L2-norm-based tests (Faraway, 1997; Zhang *et al.*, 2010a; Zhang and Chen, 2007), F-type tests (Shen and Faraway, 2004; Zhang and Liang, 2013), bootstrap tests (Faraway, 1997; Cuevas *et al.*, 2004; Zhang *et al.*, 2010a; Zhang and Sun, 2010) are well described for two-sided composite hypothesis test. However, for the one-sided composite hypothesis test, all those tests could not be applied.

In this work, we used a Bayesian approach to construct an optimal point-wise test and simultaneous confidence bands to construct a global test procedure to perform the non-inferiority hypothesis testing on the whole period of the follow-up.

The section 2 present the formulation of the problem and introduce the notion of functional non-inferiority hypothesis test. The section 3 present the optimal test for functional data which is adopted for functional non-inferiority test on a continuum domain. In this section, it is also presented a global test procedure based on simultaneous confidence bands. The assessment of the adopted methods is done in the section 4 through a simulation example.

Formulation of Functional Non-Inferiority Hypothesis Test

Let's consider an active controlled non-inferiority trial with repeated measurements over time, where the goal is non-inferiority testing of a new treatment or health intervention (N) versus a reference health intervention (R). We assume that the endpoint is a

continuous variable X , X_N and X_R for group N and R respectively that should be observed on finite grid point $\{t_1, \dots, t_m\}$. Generally, the non-inferiority hypothesis testing is performed at the end of the follow-up period t_m . This makes the data collected before the end of follow-up period useless for non-inferiority test. However, the repeated outcomes can be viewed as variable of dimension m , $(X_N(t_1), \dots, X_N(t_m))$ and $(X_R(t_1), \dots, X_R(t_m))$ for the groups N and R respectively. The idea is to get a decision on non-inferiority testing on the whole continuum domain $\tau = [t_1, t_m]$ not only at the end of follow-up period t_m . To overcome that, some tools such as longitudinal and functional data analysis can be used. Longitudinal data analysis has been much used in the context of cohort studies. However, the functional data analysis is more flexible and consists of modelling or converting the initial data set on the discrete grid $\{t_1, \dots, t_m\}$ into curves or functional dataset on the continuum interval $\tau = [t_1, t_m]$. Ramsay and Silverman (2005) have provided a broad overview. Basically, it's consist of assuming there exist functions f_i on t which modelled individual trajectory $(X_{i,c}(t_1), \dots, X_{i,c}(t_m))$, $i = 1 \dots n$, $c \in \{R, N\}$. That is by:

$$X_{i,c}(t_j) = f_{i,c}(t_j) + \varepsilon_j^{i,c}, j = 1, \dots, m$$

where, $\varepsilon_j^{i,c}$ is the error measurement process for i th individual in group $c \in \{R, N\}$ at t_j . In the practise, the true expression of individual function $f_{i,c}$ is unknown, but can be approximated by using observed data and approximations techniques such as local polynomial kernel smoothing, P-spline, regression and smoothing splines. The general principle is to get the approximation of functions $f_{i,c}$ such that the error measurements are minimised. It follow that the underline individual functions:

$$f_{i,c}(t_j) = \mu_c(t_j) + v_{i,c}(t_j)$$

are independent and identically distributed (i.i.d). copies of the underlying stochastic process $f_c(t) \sim GP(\mu_c(t), \gamma_c(t))$, $t \in \tau$, $c \in \{R, N\}$, where μ_c , are mean function and γ_c the covariance matrix defined on τ in each group. By using functional data analysis modelling, the non-inferiority testing would be the comparison of mean functions $\mu_N(t)$ and $\mu_R(t)$ of the new and reference treatments respectively in term of non-inferiority on the continuum domain $\tau = [t_1, t_m]$. That requires the formulation of non-inferiority with functional endpoints. A simple and intuitive way is to formulate similarly to the case of scalar endpoint, at every point of the domain τ .

$$\begin{cases} H_0 : \mu_N(t) \leq \mu_R(t) - \Delta_L(t), \forall t \in \tau & \text{There is no non-inferiority on the continuum } \tau \\ H_1 : \exists t \in \tau \text{ such that } \mu_N(t) > \mu_R(t) - \Delta_L(t) & \text{There is non-inferiority on the continuum } \tau \end{cases} \quad (1)$$

where, $\Delta_L(\cdot)$ is the predefined non-inferiority margin function, such that $\Delta_L(t) > 0$, for all $t \in \tau$. The hypothesis test in Equation 1 will refer to **functional non-inferiority hypothesis test** on the continuum domain τ .

Most of the statistical tests for overall testing addressed are for two-sample problem for functional mean differences. Zhang *et al.* (2010b; Zhang, 2014) is proposed an overall test for mean difference based on L^2 -Norm, Staicu *et al.* (2014) and Shen and Faraway (2004) have proposed pseudo likelihood ratio test. All such methods could not be applied for the one-sided hypothesis test problem for the functional means difference in Equation 1. In fact, the construction of the test statistic uses the L^2 Norm or other global test statistic based on *Sup* norm (Taylor *et al.*, 2007), which cannot allow knowing to the direction of the inequality when the null hypothesis is rejected.

One could then alternatively use the pointwise approach which had been used in Forgy and Small (2014) for equivalence testing for functional data. In that case, it is required to control the compound error for the test on the whole continuum domain.

Methodology

Notations and Assumptions

1. We assume that the functional random variables $X_{i,c}$, $c \in \{N, R\}$ are i.i.d. That means, for all t in τ , the real individual trajectories $X_{i,c}$ are independent and identically distributed
2. Gaussian Process (GP) with mean function m and covariance matrix g will be denoted $GP(\mu, \gamma)$ and $f_c(t) \sim GP(\mu_c(t), \gamma_c(t))$, $t \in \tau$, $c \in \{R, N\}$
3. Lets by $\hat{\mu}_c$, $c \in \{N, R\}$ the estimator of the mean function μ_c :

$$\hat{\mu}_c(t) = n_c^{-1} \sum_{i=1}^{n_c} X_{i,c}(t)$$

4. Lets by $\hat{\gamma}_c(t, s)$, $c \in \{N, R\}$ the classical unbiased covariance estimate of γ_c :

$$\hat{\gamma}_c(t, s) = (n_c - 1)^{-1} \sum_{i=1}^{n_c} (X_{i,c}(t) - \hat{\mu}_c(t))(X_{i,c}(s) - \hat{\mu}_c(s))$$

Pointwise Approach based Test

To perform the functional non-inferiority hypothesis testing in Equation 1, one can perform the pointwise non-inferiority hypothesis testing using scalar case at

each point. Then, for a given $t \in \tau$, let define by $r(t) = 0$ and $r(t) = 1$ when the null and alternative hypothesis are respectively true, $w(t) = 0$ and $w(t) = 1$ when the null and alternative hypothesis are declared respectively true from the observed data. The total of set of true and false null hypothesis is defined by: $T_k = \{t \in \tau, r(t) = k\}$, $k = 0, 1$, the total of set of declaration and non-declaration of null hypothesis from data is defined by: $D_k = \{t \in \tau, w(t) = k\}$, $k = 0, 1$ respectively. The Table 1 summarizes the outcomes of pointwise functional non-inferiority hypothesis testing : Let assume that at every point $t \in \tau$ there is type I error α_t . For the decision about the non-inferiority test on the whole domain τ , it is required a compound error measure like in multiple hypothesis testing. The False Discovery Rate (FDR) formally introduced in Benjamini and Hochberg (1995) and Family Wise Error Rate (FWER) introduced in Hochberg and Tamhane (1987) are respectively the main indicators used for evaluating the compound type error in the setting of multiple hypothesis testing. The FWER for functional data by Cox and Lee (2007) is well appropriated on the condition of permutation pivotality which may not be held in the functional non-inferiority hypothesis testing. In fact, for functional non-inferiority hypothesis testing, the inequality in the null hypothesis tests, prevent this condition to be satisfied. In the setting of this work, the false discovery rate is used as the compound type I error measure. Then, there are some methods devoted to the control of the false discovery rate in the context of mean differences, such as Benjamin-Hochberg method and Benjamin-Yekutieli methods in Benjamini and Hochberg (1995). Xu *et al.* (2018) proposed another point-wise method controlling the false discovery, which had been found better than Benjamin-Hochberg method and Benjamin-Yekutieli methods. Therefore, in this work, the optimal pointwise test introduced in Xu *et al.* (2018) is adopted for the functional non-inferiority hypothesis testing.

This marginal false discovery rate it is defined by:

$$mfd_r = \frac{E(\mathcal{L}(V_1))}{E(\mathcal{L}(D_1))} \quad (2)$$

where, \mathcal{L} is the Lebesgue measure of interval subsets. The overall power for the functional non-inferiority test is define in a similar way as in Leventhal and Huynh (1996) and Sun *et al.* (2015) for multiple two-sided testing by:

$$\pi = 1 - \frac{E(\mathcal{L}(V_2))}{E(\mathcal{L}(T_1))} \quad (3)$$

Table 1: Outcomes of multiple non-inferiority hypothesis testing

	No non-inferiority declared (w(t) = 0)	Non-inferiority declared (w(t) = 1)	Total
No non-inferiority (r(t)=0)	/	Type I error (V ₁)	T ₀
Non-inferiority (r(t)=1)	Type II error(V ₂)	/	T ₁
Total	D ₀	D ₁	τ

For a given compound nominal type I error α_c , we want to determine the test w with a false discovery rate $\hat{m}fdr$ such that $\hat{m}fdr \leq \alpha_c$. As shown in Xu *et al.* (2018), the optimal test w should be searched among the test family w_λ , $\lambda > 0$ defined by:

$$w_\lambda(t_j) = \begin{cases} 1 & \text{If } \frac{Pr(r(t_j)=1)}{Pr(r(t_j)=0)} \geq \lambda \\ 0 & \text{If } \frac{Pr(r(t_j)=1)}{Pr(r(t_j)=0)} < \lambda \end{cases} \quad (4)$$

The false discovery rate for a test w_λ being estimated by:

$$\hat{m}fdr(w_\lambda) = \frac{\sum_{i=1}^N Pr(r(t_i)=0) 1_{w_\lambda(t_i)=1}}{\sum_{i=1}^N 1_{w_\lambda(t_i)=1}} \quad (5)$$

where, t_1, \dots, t_N are the center points of the sequence of subintervals $[s_{i-1}, s_i]$, $i = 1, \dots, N$, which is a partition of τ . The probability $S_0(t_j) = Pr(r(t_j) = 0)$ is the probability of true null hypothesis of no non-inferiority at the point t_j , it is unknown and can be estimated only in a Bayesian setting. Therefore, one would assume a prior distribution on μ_N and μ_R and get the posterior distribution from the observed data X_N and X_R . Then, from the posterior distribution, one uses Monte Carlo simulations techniques to get a couple of M functional process samples μ_N^i and μ_R^i , $i = 1, \dots, M$. Thus, estimate $S_0(t_j)$ by:

$$\hat{S}_0(t_j) = \sum_{i=1}^M 1_{\mu_N^i(t_j) \leq \mu_R^i(t_j) - \Delta(t_j)} \quad (6)$$

Therefore, the test decision at any point t is given by:
 $w_\lambda(t) = \sum_{i=1}^N 1_{t \in [s_{i-1}, s_i]} w_\lambda(t_i)$. The optimal test w controlling the false discovery rate at a given nominal compound type I error rate α_c is then $w = w_{\lambda^*}$, where

$$\lambda^* = \min \{ \lambda > 0, \hat{m}fdr(w_\lambda) \leq \alpha_c \}.$$

Algorithm for Determining the Optimal Test

The expression of $\hat{m}fdr$ is not explicit, therefore, the determination of λ^* requires numerical computation techniques. As shown in Sun *et al.*

(2015), $\hat{m}fdr$ is monotone decreasing, then, the λ which provides the maximum false discovery rate should be found in a positive neighbourhood of 0. Also, when λ tends to infinity, the function $\hat{m}fdr$ will tend to null. Therefore, the computational algorithm will depend on α_c (smaller or greater). For larger α_c (for example 10% or 5%), the forward algorithm presented in the Algorithm 1 can be suitable with an initial value λ_0 closer to 0. In the case of smaller α_c (for example 2.5% or 1%) the backward algorithm presented in the Algorithm 2 can be preferable with larger initial value λ_0 . Whatever the case, as any computational problem with the initial input parameter, the results will depend on.

Algorithm 1: Algorithm for the computation of the optimal λ^* which controls the false discovery rate using forward direction:

Data: $\lambda_0, tol, Max_{iter}, \alpha$
Result: $\lambda^*, mfrd(\lambda^*)$
 $\lambda = \lambda_0;$
 $N_{iter} = 1;$
 $m = \hat{m}fdr(\lambda);$
while $m > \alpha$ **and** $N_{iter} \leq Max_{iter}$ **do**
 $\lambda = \lambda + tol;$
 $m = \hat{m}fdr(\lambda);$
 $N_{iter} = N_{iter} + 1;$
end
 Return λ^*, m .

Algorithm 2: Algorithm for the computation of the optimal λ^* which controls the false discovery rate using backward direction:

Data: $\lambda_0, tol, Max_{iter}, \alpha_c$
Result: $\lambda^*, mfrd(\lambda^*)$
 $\lambda = \lambda_0;$
 $N_{iter} = 1;$
 $m = \hat{m}fdr(\lambda);$
while $m > \alpha_c$ **and** $N_{iter} \leq Max_{iter}$ **do**
 $\lambda = \lambda - tol;$
 $m = \hat{m}fdr(\lambda);$
 $N_{iter} = N_{iter} + 1;$
end
 Return λ^*, m .

Confidence Bands based Test

Likewise in the non-inferiority hypothesis testing for scalar data Ng (2008); Elie *et al.* (2008); Food and Drug Administration (2016), one can adopt confidence bands for the formulated functional non-inferiority testing in Equation 1. The idea is to reject the null hypothesis when the margin function $-\Delta_L$ is under the lower confidence band of $\mu_N - \mu_R$ on a subset $\tilde{\tau}$ of τ . The Fig. 1 gives an illustration of functional non-inferiority testing based on confidence bands. Denoting by l and u the lower and upper confidence bands of $\mu_N - \mu_R$ respectively, η the confidence bands level, following are the steps of the test procedure:

1. Construct a confidence bands $[l, u]$ of level η of $\mu_N - \mu_R$
2. Compare the two functions l and $-\Delta_L$ on τ
3. Reject the null hypothesis H_0 of no non-inferiority on τ if and only if there exist a subset $\tilde{\tau}$ of τ such that l is greater than $-\Delta_L$ on $\tilde{\tau}$

The construction of confidence bands can be done using a pointwise approach in parametric or non-parametric settings. That will not be valid for overall or simultaneous inference, since the coverage level of confidence bands will be less than η Degras (2017). In this work, the simultaneous confidence bands (SCB) by Degras (2017) is adopted. The lower and upper confidence bands l and u for $\mu_N - \mu_R$ are defined for all $t \in \tau$ respectively by:

$$\begin{aligned}
 l(t) &= \hat{\mu}_N(t) - \hat{\mu}_R(t) - z_{\eta, \hat{\rho}_{NR}} \left(\frac{\hat{\gamma}_R(t,t)}{n_R} + \frac{\hat{\gamma}_N(t,t)}{n_N} \right)^{\frac{1}{2}} \\
 u(t) &= \hat{\mu}_N(t) - \hat{\mu}_R(t) + z_{\eta, \hat{\rho}_{NR}} \left(\frac{\hat{\gamma}_R(t,t)}{n_R} + \frac{\hat{\gamma}_N(t,t)}{n_N} \right)^{\frac{1}{2}}
 \end{aligned} \tag{7}$$

With $z_{\eta, \hat{\rho}_{NR}}$ the η -quantile of $\sup_{t \in \tau} |Z(t)|$, $Z \sim GP(0, \hat{\rho}_{NR})$ and $Z \sim GP(0, \hat{\rho}_{NR})$, $\hat{\rho}_{NR} = \hat{\rho}_R + \hat{\rho}_N$, $\hat{\rho}_R(s,t) = \frac{\hat{\gamma}_R(t,s)}{\hat{\gamma}_R(t,t)\hat{\gamma}_R(s,s)}$, $\hat{\rho}_R$ is then define similarly.

The application of SCB with level η for overall testing of the equality of two mean functions in a two-sided problem in Degras (2017), provided a type I error which is asymptotically equal to $1 - \eta$ and the statistical power tending to 1. The level of the proposed test procedure for functional non-inferiority test and the statistical power will be evaluated through a simulation example by using the Monte-Carlo method.

Simulations Example

Simulations Scenario and Settings

The purpose of the simulation example is the evaluation of the proposed functional non-inferiority

hypothesis testing based on the optimal pointwise test and confidence bands based test. The nominal type I error rate is evaluated for the SCB based test, the *mfrd* is evaluated for the optimal pointwise test, the power is estimated for both proposed tests. Two scenario were considered: A scenario simulating functional data with the null hypothesis satisfies and a scenario with the alternative satisfies. Each scenario purposing respectively the evaluation of actual type I error rate and *mfd*r for the optimal pointwise test and SCB based test and statistical power for both tests.

The inputs parameters using for generating functional data sets are functional margin Δ_L , functional means μ_N and μ_R , co-variance matrix γ_N and γ_R . In all simulations, we consider the discrete grid point (0,6,12,24) and continuum domain $\tau = [0,24]$. In all simulations, it has been assumed the equality co-variance matrix, $\gamma(t, s) = \gamma_N(t, s) = \gamma_R(t, s)$. It was considered the case of correlated data, which is the most encountered in the practice. Then, it was assumed that the correlation depends on the distance between time points, the data at two closer points are assumed more correlated than the data at two distant points. Therefore, one could take $\gamma(t, s) = 1902 \exp(-(t-s)^2)$. The equal sample sizes are considered ($n_R = n_N = n$), $n = 30, 100, 1000$. The Gaussian process was used to simulate process on the discrete grid point (0,6,12,24), then splines was used for smoothing on the continuum domain $\tau = [0,24]$. The nominal compound type I error has been set to $\alpha_c = 10\%$. In all simulations, the estimation of the *mfrd* was based on the forward algorithm 1, therefore, the initial parameter λ_0 was chosen smaller and in the closer neighborhood of 0.

The estimation of false discovery rate for the optimal pointwise test and the actual type I error rate for the SCB based methods, the data are generated on the null hypothesis, for example, that is when $\mu_N(t) = \mu_R(t) - \Delta_L(t)$ for all $t \in \tau$. In that case we chose $\mu_R(t) = 30t$, $\Delta_L(t) = (35/3)t + 50$. The power for both methods is estimated by drawing the functional data on the alternative. Similarly to the case of scalar data (Zhang, 2006; Flight and Julious, 2016), it is considered the particular case when $\mu_N = \mu_R$ on $\tau = [0,24]$. The actual type I error rate and statistical for the procedure based on SCB are also evaluated according to the level of SCB. Therefore, it is considered a confidence bands level $\eta = 95\%, 90\%$ and 80% .

The **R** software programming language (R Core Team, 2016) has been used to conduct all the simulations and codes are accessible in a separate file. However, the packages FDA by Ramsay *et al.* (2018) and mvtnorm by Genz *et al.* (2018) have been specially useful for the simulations and the manipulations of functional process data.

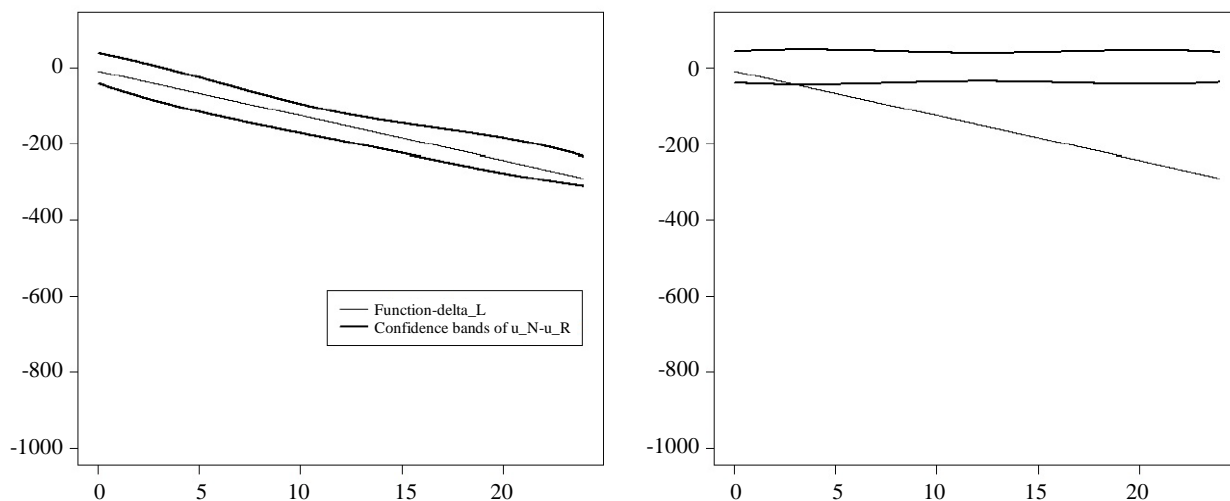


Fig. 1: Illustration of functional non-inferiority hypothesis testing based on confidence bands. The null hypothesis is not rejected (left), the alternative, non-inferiority on $[0,24]$ demonstrated (right)

The estimation of the probability of getting the null and alternative in the optimal test require a prior distribution on the process μ_N and μ_R . Most time, one choose Gaussian processes with zero means and co-variances functions K_N and K_R (It is assumed equal co-variances functions: $k = K_N = K_R$) which belongs to a known family of functions. There are constant, polynomial, Matérn, rational quadratic, exponential and so on. In this work, we will use the exponential family defines by:

$$k(x, y, w, \theta) = \theta \exp\left(-w(x-y)^2 / 2\right) \quad (8)$$

where, w and θ are called hyper-parameters, x and y are the points where the co-variance function is estimated. In this work, θ will be chosen fixed and equal to 190^2 and $w = 100$.

Simulation Procedures

The estimation of mfd_r and π , are done by the following steps:

1. Simulate two couples of Gaussian process X_N and X_R on the discrete grid $\{0,6,12,18,24\}$ with functional mean μ_N and μ_R as described above such that the null hypothesis is satisfied and with sample size n
2. Convert Gaussian process X_N and X_R into functional data on the continuum $[0,24]$, then get two functional data set $X_N(t)$ and $X_R(t)$
3. Compute the mean functions estimated $\hat{\mu}_N(t)$ and $\hat{\mu}_R(t)$. Then, get the posterior expression of the mean and co-variance functions for μ_N and μ_R
4. Simulate M couples of Gaussian process sample $\hat{\mu}_N^i$ and $\hat{\mu}_R^i$ form the posterior distribution

obtained in the step 3. Then estimate

$$\hat{S}_0(t) = \frac{1}{m} \sum_{i=1}^M 1_{\hat{\mu}_N^i(t) \leq \hat{\mu}_R^i(t) - \Delta(t)}$$

5. Define the function: $m\hat{f}dr(\lambda) = \frac{\sum_{i=1}^N \hat{S}_0(t_i) 1_{w(t_i)=1}}{\sum_{i=1}^N 1_{w(t_i)=1}}$,

where the function $1_{w(t)=1} = 1$ if $\hat{S}_0(t) > (1 + \lambda)^{-1}$

6. Compute $\lambda^* = \min\{\lambda, m\hat{f}dr(\lambda) \leq \alpha\}$. Replace λ^* in the step 5 to get the optimal test which controls the $m\hat{f}dr$ at level α_c as well an estimate of $m\hat{f}dr$
7. Repeat the steps 1-6 B times, then get a distribution $m\hat{f}dr_1, \dots, m\hat{f}dr_B$ of $m\hat{f}dr$

The estimation of the power is done in a similar way, but at the step one, the data are generated on the alternative hypothesis. then, estimate the power by

$$\hat{\pi} = 1 - \frac{\sum_{i=1}^N \hat{S}_1(t_i) 1_{w(t_i)=1}}{24N}$$

Repeat B times then get the a distribution $\hat{\pi}_1, \dots, \hat{\pi}_B$.

For the test based on confidence bands, it is considered a level of $\eta = 95\%$, 90% and 80% for confidence bands and all the input parameters as in the optimal pointwise test are considered. The type I error and statistical power is estimated as follows:

1. Simulate two couples of Gaussian process X_N and X_R on the discrete grid $\{0, 6, 12, 18, 24\}$ with functional mean μ_N and μ_R as described above such that the null hypothesis is satisfied and with sample size n
2. Convert Gaussian process X_N and X_R into functional data on the continuum $[0,24]$, then get two functional data set $X_N(t)$ and $X_R(t)$

3. Compute the functional means estimated $\hat{\mu}_N(t)$ and $\hat{\mu}_R(t)$ as well as the functional variances estimated $\hat{\gamma}_N(t,t)$ and $\hat{\gamma}_R(t,t)$
4. Then, determine the lower confidence bands l of μ_N - μ_R of level η with the formula in Equation 7.
5. Set the function $f(t) = l(t) + \Delta_L(t)$ defined on τ and solve the equation, denoting by Ω the subset of τ such that $f(t) > 0$, the null hypothesis of no non-inferiority on τ is rejected when $\Omega \neq \emptyset$
6. Repeat the steps 1-5 B times, then get B subset Ω denoted $\Omega_1, \dots, \Omega_B$. The type 1 error is then estimated by: $\hat{\alpha} = \sum_{i=1}^B 1_{\Omega_i \neq \emptyset}$

The estimation of statistical power is done in the similar way with the data sample on the alternative hypothesis at step 1.

Stability Analysis and Simulation Results for Optimal Pointwise Test

The optimal pointwise test is based on the numerical algorithm with the input parameters. The optimal pointwise test is studied according to the initial parameter λ_0 by fixing the tolerance tol and the maximal number of iteration Max_{iter} . A smaller value in a closer neighborhood of 0 has been chose by $\lambda_0 = 1e-10$, then larger values of $\lambda_0 = 1$ and $\lambda_0 = 100$. The idea is to provide a better guess of the initial parameter for more accurate

results of the pointwise test. The results are presented in the Fig. 2. With the same value of λ_0 , the test with large sample sizes gets closer to the nominal compound type I error rate. While for the smaller sample size the test would be more conservative. Whatever the sample sizes, the $m\hat{f}dr$ estimate gets smaller as the initial parameter λ_0 gets larger. The larger value of λ_0 would lead to a more conservative test. Therefore, the results suggest that the smaller value of the initial guess λ_0 at a closer neighborhood of 0 would be preferred. For more accuracy, the stability analysis is done to study the pollution of results for a random choice of λ_0 at a closer neighborhood of 0. Due to computation time, which is high, the stability analysis was limited to three cases of a random choice of λ_0 at a closer neighborhood of 0. Let's estimate the $m\hat{f}dr$ thrice with three different random values of λ_0 at a closer neighborhood of 0 (For instance, λ_0 chooseen with uniform distribution on $[1e-10, 1e-1]$). The results are represented in the Fig. 3, Whatever the sample size, the results show that the boxplot distribution of $m\hat{f}dr$ is not much affected by a random choice of λ_0 at a closer neighborhood of 0. Fixing the tolerance and the number of iterations, the optimal pointwise test would be stable for λ_0 guess at a closer neighborhood of 0 and for small sample size $n = 30$, the $m\hat{f}dr$ estimate would be around 5%, for the medium sample size $n = 100$, the $m\hat{f}dr$ estimate would be around 7% and for large sample size $n = 1000$, it would be around 9%. Concerning the statistical power, as shown in Table 2, the power tends to 1.

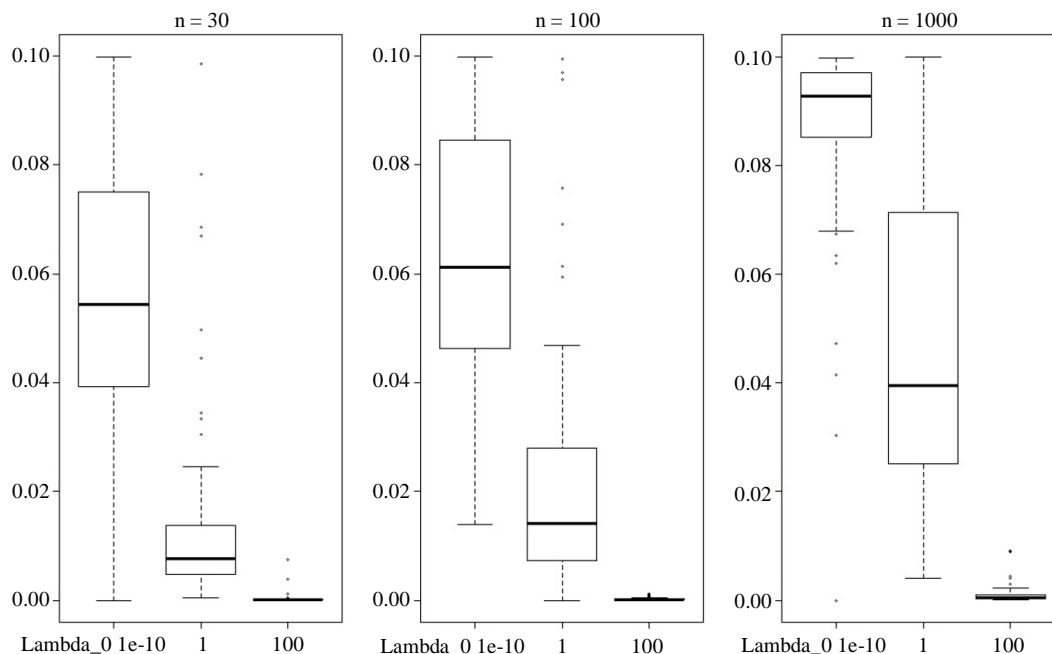


Fig. 2: The boxplots of $m\hat{f}dr$ controlling the test at the nominal level $\alpha_c = 0.1$ according to the initial parameters ($\lambda_0 = 1e-10$, $\lambda_0 = 1$, $\lambda_0 = 100$) and sample sizes ($n = 30, 100, 1000$)

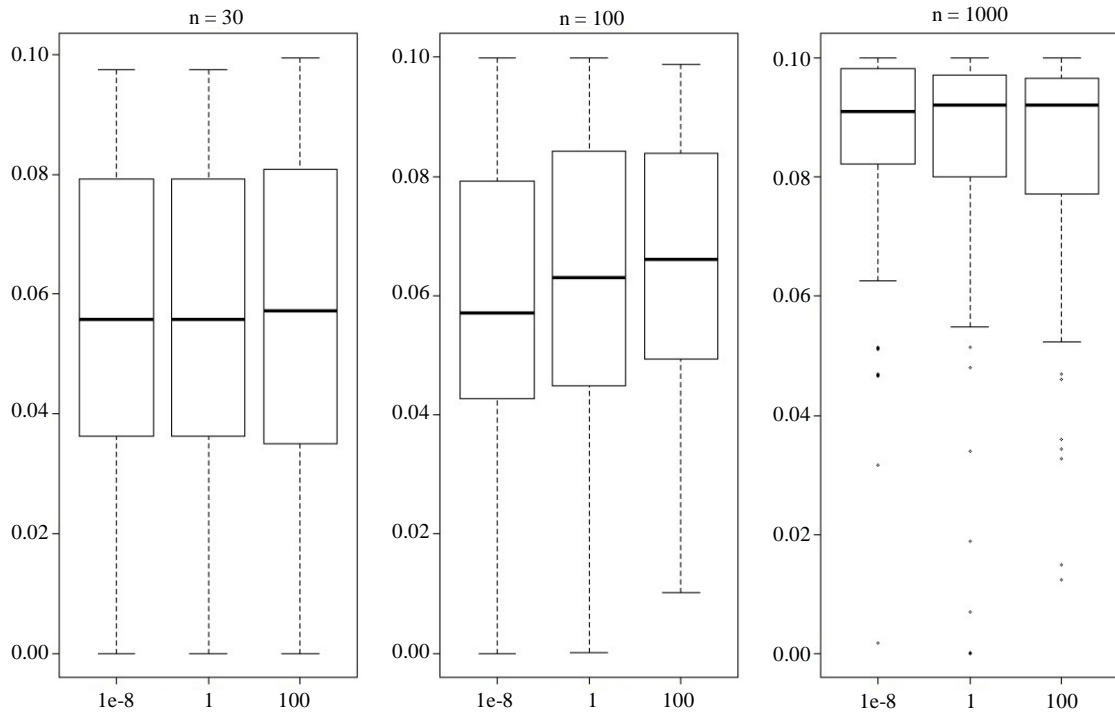


Fig. 3: The boxplots of \hat{mfrd} controlling the test at the nominal level $\alpha_c = 0:1$ according to random initial parameters λ_0 at a closer neighborhood of 0 and sample sizes ($n = 30;100;1000$)

Table 2: Estimation of the power according to sample sizes for the optimal pointwise test

	Power
$n_N = n_R = 30$	0.93
$n_N = n_R = 100$	0.99
$n_N = n_R = 1000$	1.00

Table 3: Estimation of the type I error rate according to the level of confidence bands (95, 90 and 80%), sample sizes ($n = 30,100,1000$) and the type of data (non-correlated and correlated data) using the SCB based test

	95%	90%	80%
$n_N = n_R = 30$	0.041	0.076	0.14
$n_N = n_R = 100$	0.038	0.074	0.12
$n_N = n_R = 1000$	0.03	0.058	0.11

Results for the SCB Based Test

The results for SCB based test are summarized in Table 3. The type I error rate and power are estimated according to the sample sizes and confidence bands level. As the sample sizes increase, the simulated type I error rate decrease and seem converging to specific values. The results can allow concluding that the method based on confidence bands with level of 95%, 90% and 80% lead approximately to a type I error rate of 2.5%, 5% and 10% for large sample size respectively. Therefore, for a given confidence bands level η , the methods would produce a test with a type I error rate approximately to $(1-\eta)/2$ asymptotically. Whatever the sample sizes and the confidence bands level and the statistical power estimation was equal to 1.

Discussion

This work has introduced functional non-inferiority hypothesis testing for the continuous variable in longitudinal trial. After formulating the hypothesis test, the optimal pointwise test in Xu *et al.* (2018) and simultaneous confidence bands in Degras (2017) were adopted. The pointwise test has the advantage to show the significant area. It was not possible to adopt the classical global test for functional two-sample mean problem in Zhang *et al.* (2010b); Zhang (2014). Since they are based on a norm which could not allow evidence for the direction of the inequality when the null hypothesis is rejected. However, the proposed SCB based test procedure can be regarded as a global test. The both proposed test

procedure for functional non-inferiority testing got good performances for large sample sizes.

The added value of this work is that the non-inferiority hypothesis testing is determined on a whole continuum domain not only at the end of follow-up period. That can allow more flexibility in the interpretation of the results of the trial. Also, that could be relevant for the determination of non-inferiority delay which could be helpful to determine the follow-up duration of future trials with the similar treatment effect. The proposed methods based on the pointwise multiple testing procedure involved numerical techniques methods which approximations depend on parameters such as the initial value and tolerance from which the results would depend on. This study provided a stability analysis for a proper guess of the initial value, which had not been studied in Xu *et al.* (2018).

Like any scientific study, this study presents some limitations, for example, an improper guess of the initial parameter would lead to a too conservative test. Also, the usage of optimal test required non-linear recursive programming which is costly in terms of execution time. All these, as well as a global test for functional non-inferiority based on a test statistic, could be another interesting future avenue of research. The study has introduced the non-inferiority test with functional endpoint which has not been previously studied in the literature. But, that involve many methodological aspects for non-inferiority trial such as assay sensitivity, constancy assumption and non-inferiority margin which should be studied for functional endpoint. This may constitute an interesting issue for future research work in the non-inferiority trials.

Conclusion

This article introduced the non-inferiority hypothesis testing with a functional endpoint. The pointwise based test and simultaneous confidence bands based test were proposed. Both proposed test procedures got good performances for large sample sizes. For small sample sizes, the pointwise based test would be too conservative while the simultaneous confidence bands based test would be a bit liberal. The functional endpoint is less used in clinical trial studies. We hope that this study could attract the attention of practitioners in the area of clinical trials of the relevance of functional endpoint in clinical trials in general and non-inferiority trials in particular.

Acknowledgment

The first author is thankful to the African Union (AU), he was a recipient of a doctorate scholarship. The comments of the reviewers and Editor-in-Chief have significantly improved the content of this study, the authors are grateful.

Authors Contribution

Arsène Brunelle Sandie: Contributed to the writing of the manuscript, simulations study and reviewed the manuscript.

Anthony Wanjoya: Contributes to the writing and reviewing the manuscript.

Jules Brice Tchatchueng-Mbougua: Contributed to the writing of the manuscript, provided guidance and reviewed the manuscript.

Ethics

This article is original and contains unpublished material. The corresponding author confirms that all of the other authors have read and approved the manuscript and no ethical issues involved.

References

- Benjamini, Y. and Y. Hochberg, 1995. Controlling the false discovery rate: A partical and powerfull approach to multiple testing. *J. Royal Stat. Society B*, 57: 289-300. DOI: 10.1111/j.2517-6161.1995.tb02031.x
- Cox, D.D. and J.S. Lee, 2007. Pointwise testing with functional data using the westfall-young randomization method. Technical Report, College of Humanities and Social Sciences at Research Showcase.
- Cuevas, A., M. Febrero-Bande and R. Fraiman, 2004. An anova test for functional data. *Comput. Stat. Data Anal.*, 47: 111-122. DOI: 10.1016/j.csda.2003.10.021
- Darlin, D.A., 1957. The kolmogorov-smirnov, cramer-von mises tests. *Annals Math. Stat.*, 28: 823-838. DOI: 10.1214/aoms/1177706788
- Degras, D., 2017. Simultaneous confidence bands for the mean of functional data. *WIREs Comput. Stat.*, 9: e1397-e1397. DOI: 10.1002/wics.1397
- Elie, C., Y.D. Rycke, J.P. Jais, R. Marion-Gallois and P. Landais, 2008. Methodological and statistical aspects of equivalence and non inferiority trials. *Revue d'Épidémiologie et de Santé Publique*, 56: 267-277. DOI: 10.1016/j.respe.2008.05.027
- Eubank, R.L., 1999. *Nonparametric Regression and Spline Smoothing*. 2nd Edn., CRC Press, ISBN-10: 0824793374, pp: 360.
- Faraway, J.J., 1997. Regression analysis for a functional response. *Technometrics*, 39: 254-261. DOI: 10.1080/00401706.1997.10485118
- Flight, L. and S.A. Julious, 2016. Practical guide to sample size calculations: non-inferiority and equivalence trials. *Pharm. Stat.*, 9: 80-89. DOI: 10.1002/pst.1716
- Food and Drug Administration, 2016. Non-inferiority clinical trials to establish effectiveness-guidance for industry. Technical report, U.S. Department of Health and Human Services.

- Forgaty, C.B. and D.S. Small, 2014. Equivalence testing for functional data with application to comparing pulmonary function devices. *Annals Applied Stat.*, 8: 2002-2026. DOI: 10.1214/14-AOAS763
- Genz, A., F. Bretz, T. Miwa, X. Mi and F. Leisch *et al.*, 2018. MVTNORM: Multivariate Normal and t Distributions. R Package Version 1.0-8.
- Green, P.J. and B.W. Silverman, 1994. Nonparametric Regression and Generalized Linear Models: A Roughness Penalty Approach. 1st Edn., CRC Press, ISBN-10: 1482229757, pp: 184.
- Hall, P. and I.V. Keilegom, 2007. Two-sample tests in functional data analysis starting from discrete data. *Stat. Sinica*, 17: 1511-1531.
- Hochberg, Y. and A.C. Tamhane, 1987. Multiple Comparison Procedures. 1st Edn., Wiley, New York, ISBN-10: 0471822221, pp: 450.
- Johnson, N.L. and S. Kotz, 1990. Use of moments in deriving distributions and some characterizations. *Math. Sci.*
- Laurent, C., C. Kouanfack, G. Laborde-Balen, A.F. Aghokeng, Aghokeng and J.B. Mbougua *et al.*, 2011. 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.*, 11: 825-833. DOI: 10.1016/S1473-3099(11)70168-2
- Leventhal, L. and C. Huynh, 1996. Directional decisions for two-tailed tests: Power, error rates and sample size. *Psychol. Meth.*, 1: 278-292. DOI: 10.1037/1082-989X.1.3.278
- Ng, T., 2008. Noninferiority hypotheses and choice of noninferiority margin. *Stat. Med.*, 27: 5392-5406. DOI: 10.1002/sim.3367
- Ng, T.H., 2015. Noninferiority Testing in Clinical Trials: Issues and Challenges. 1st Edn., CRC Press, ISBN-10: 1466561491, pp: 208.
- R Core Team, 2016. A language and environment for statistical computing. R foundation for statistical computing. Vienna, Austria.
- Ramsay, J.O. and B.W. Silverman, 2005. Functional Data Analysis. 2nd Edn., Springer, New York, ISBN-10: 038740080X, pp: 426.
- Ramsay, J.O., H. Wickham, S. Graves and G. Hooker, 2018. FDA: Functional Data Analysis. R Package Version 2.4.8.
- Ruppert, D., M.P. Wang and R.J. Carroll, 2003. Semiparametric regression. Cambridge Series in Statistical and Probabilistic Mathematics.
- Shen, Q. and J. Faraway, 2004. An f test for linear models with functional responses. *Stat. Sinica*, 14: 1239-1257.
- Staicu, A.M., Y. Li, C.M. Crainiceanu and D. Ruppert, 2014. Likelihood ratio tests for dependent data with applications to longitudinal and functional data analysis. *Scandinavian J. Stat.*, 41: 932-949. DOI: 10.1111/sjos.12075
- Sun, W., B.J. Reich, T.T. Cai, M. Guindani and A. Schwartzman, 2015. False discovery control in large-scale spatial multiple testing. *J. R Stat. Soc. Ser. B Stat. Methodol.*, 77: 59-83. DOI: 10.1111/rssb.12064
- Taylor, J.E., K.J. Worsley and F. Gosselin, 2007. Maxima of discretely sampled random fields with an application to 'bubbles'. *Biometrika*, 94: 1-18. DOI: 10.1093/biomet/asm004
- Wand, M.P. and M.C. Jones, 1995. Kernel Smoothing. 1st Edn., CRC Press, ISBN-10: 0412552701, pp: 224.
- Xu, P., Y. Lee, J.Q. Shi and J. Eyre, 2018. Automatic detection of significant areas for functional data with directional error control. *Stat. Med.*, 38: 376-397. DOI: 10.1002/sim.7968
- Zhang, C., H. Peng and J.T. Zhang, 2010a. Two samples tests for functional data. *Commun. Stat. Theory Meth.*, 39: 559-578. DOI: 10.1080/03610920902755839
- Zhang, J.T., 2014. Analysis of variance for functional data. Taylor and Francis Group, 6000 Broken Sound Parkway NW, Suite 300. Monographs on Statistics and Applied Probability.
- Zhang, J.T. and J. Chen, 2007. Statistical inferences for functional data. *Annals Stat.*, 35: 1052-1079. DOI: 10.1214/009053606000001505
- Zhang, J.T. and X. Liang, 2013. One-way ANOVA for functional data via globalizing the pointwise f -test. *Scandinavian J. Stat.*, 41: 51-71. DOI: 10.1111/sjos.12025
- Zhang, J.T., Y. Liang and S. Xiao, 2010b. On the two-sample Behrens-fisher problem for functional data. *J. Stat. Theory Pract.*, 4: 571-587. DOI: 10.1080/15598608.2010.10412005
- Zhang, J.T. and Y. Sun, 2010. Two-sample test for equal covariance function for functional data. *Oriental J. Math.*
- Zhang, Z., 2006. Non-inferiority testing with a variable margin. *Biometr. J.*, 48: 948-965. DOI: 10.1002/bimj.200610271