Current Research in Biostatistics

Original Research Paper

Global Test for Superiority and Non-Inferiority Trials with Functional Endpoints Data

[1]Arsene Brunelle Sandie, [2,3]Jules Brice Tchatchueng-Mbougua and [4]Anthony Wanjoya

*[1]Department of Mathematics-Statistics, Pan African University Institute of basic Science,
Technology and Innovation (PAUSTI)/ Jomo,
Kenyatta University of Agriculture and Technology (JKUAT), Nairobi, Kenya
[2]Service d´´epid´emiologie et de sant´e publique du Centre Pasteur du Cameroun,
Membre du R´eseau International des Instituts Pasteur, Yaound´e, Cameroun
[3]Universit´e de Yaound´e I - CETIC, UPMC Universit´e Paris 06, IRD,
Unit´e de Mod´elisation Math´ematique et Informatique des Syst`emes Complexes (UMMISCO),
F-93143, Bondy, France
[4]Department of Statistics and Actuarial Sciences,
Jomo Kenyatta University of Agriculture and Technology (JKUAT), Nairobi, Kenya*

Article history
Received: 29-04-2019
Revised: 13-06-2019
Accepted: 05-07-2019

Corresponding Authors:
Arsene Brunelle Sandie
Departement of Statistics-Mathematics, Pan African University, Nairobi, Kenya

Email: sandiearsene@gmail.com

Abstract: The global two-sided test for functional means difference got much attention in the literary works. While the global one-sided test for functional means difference has not been studied. However, that could have some applications and implications for longitudinal trials or trials with functional data outcomes. This work introduced superiority and noninferiority hypothesis tests for functional data outcomes, which were viewed as a one-sided problem for functional mean differences. Two global tests statistic have been proposed: a test based on integral and a test based on supremum. The performances of the tests have been evaluated via a simulation example. Both tests got an estimated actual type I error closer to the nominal type I error. The test based on supremum got good estimated power in all considered cases of the alternative hypothesis, while the test based on integral got very poor power performances for some cases of considered alternative hypothesis.

Keywords: Functional Data Analysis, Global Test, Non-Inferiority Test, Superiority Test

Introduction

Functional data analysis is becoming increasingly used in different applied areas such as economics, clinical trials, meteorology and so on. It owes its applicability to the development of technologies, which allows nowadays to store and process data with larger dimensions. In fact, contrary to the scalar or multivariate data analysis, functional data analysis itself requires infinite dimension. Therefore, the functional data analysis uses its own inference tools. Ramsay and Silverman (2005) provided a large overview of different inferential methods for functional data analysis. Such as in the case of scalar data, in practice, it sometimes needs to compare statistically two functional means on a continuum domain. There are basically two approaches to do so: An approach based on pointwise test and an approach based on the global test. The pointwise test consists of performing the test at each point of the domain. But, for the decision on the whole continuum domain, it is required to control the compound type I error rate. The False Discovery Rate (FDR) and Family-Wise Error Rate (FWER) introduced by Benjamini and Hochberg (1995); Hochberg and Tamhane (1987) are respectively the main measures used to evaluate the compound type error. The global test consists of defining a scalar test statistic for the functional hypothesis testing on the whole continuum domain.

In the context of functional means difference, most of the works in the literature using a global test are for the two-sided problem. Zhang *et al*. (2010); Zhang (2014) proposed a global test for two-sided means difference based on L^2-Norm. Taylor *et al*. (2007) proposed as well a global test based on *Sup*-Norm. All those global tests cannot be used for the functional one-sided problem. Since, when the null hypothesis is rejected, the direction of the inequalities could not be determined. This work proposes global tests for the functional one-sided problem. The superiority and non-inferiority hypothesis tests with functional endpoints which are particular cases of one-sided problems are introduced. This works can lead to some considerations for the design and the interpretation of some clinical trials. For example, in the non-inferiority or superiority longitudinal trials the hypothesis testing is generally performed after a fixed period of time which is generally the end of the follow-up period. However, these longitudinal data can be converted into functional data on the continuum follow-up period and then get a decision about superiority or non-inferiority not only at the end of the follow-up period but on the whole domain. That will allow more flexibility in the interpretation of the results of the hypothesis testing.

The Monte-Carlo simulations methods are used for the assessment of the performances of the proposed tests for a one-sided problem in functional mean differences through a simulation example. The actual type I error rate and statistical power are therefore evaluated.

Methods

Formulation of Superiority and Non-Inferiority Hypothesis Test for Functional Data Outcomes

Non-inferiority and superiority test with binary, continuous and survival outcomes have been commonly used in clinical trials. They got much attention and many works of literature and technical reports were dedicated to it Food and Dug Administration (2016); Committee for Proprietary Medicinal Products (2000).

While, the same cannot be said for the functional outcomes, which are rarely considered in clinical trials. However, this latter could have some relevant applications. For example, let's consider a longitudinal trial with multiple outcomes, where the purpose is the comparison of an experimental treatment to placebo or reference. Generally, the multiple outcomes data are collected on a discrete grid time points up to a fixed period of time. Then, the test (superiority or non-inferiority) is performed with the data at the end of the follow-up period. The data collected before the end of the period are then useless for the test, however, the data can be modeled in the setting of functional data analysis by converting the data observed on the discrete grid into curves or functions on the whole continuum domain. Methods such as local polynomial kernel smoothing, P-spline, regression and smoothing splines, etc described in Ramsay and Silverman (2005) can be used for the purpose.

Let's assume we are interested in longitudinal trial with a continuous endpoint, let's X_1 and X_2 the endpoints for control and experimental treatment respectively which are each observed on a discrete grid point $\{t_1,...,t_m\}$. Denoting μ_1 and μ_2 the true and unknown means for X_1 and X_2 at the point tm, the superiority or non-inferiority is generally performed at t_m, by testing the hypothesis:

$$\begin{cases} H_0 : \mu_1 \le \mu_2 \\ H_1 : \mu_1 > \mu_2 \end{cases} \qquad (1)$$

and:

$$\begin{cases} H_0 : \mu_1 \le \mu_2 - \Delta_c \\ H_1 : \mu_1 > \mu_2 - \Delta_c \end{cases} \qquad (2)$$

where, Δ_c is the non-inferiority margin for the continuous endpoints, Δ_c>0. The idea of functional data analysis is to model scalar random variables X_1 and X_2 defined on the discrete grid $\{t_1,...,t_m\}$ by functional random variables F_1 and F_2 respectively defined on the continuum $D = [t_1, t_m]$, with true and unknown mean functions f_1 and f_2 respectively, co-variance matrix γ_1 and γ_2 respectively. Then, instead of superiority hypothesis test in 1(respectively non-inferiority hypothesis test in 2) at the point t_m, one may be interested on the superiority (respectively the non-inferiority) on the whole continuum domain D. The functional superiority hypothesis test on D(respectively functional non-inferiority hypothesis test on D) are then formulated by:

$$\begin{cases} H_0 : f_1 \le f_2 \ \ on \ \ D \\ H_1 : f_1 > f_2 \ \ on\, a\, subset\, D^*\, of\ D \end{cases} \qquad (3)$$

and:

$$\begin{cases} H_0 : f_1 \le f_2 - \Delta_f \ \ on \ \ D \\ H_1 : f_1 > f_2 - \Delta_f \ \ on\, a\, subset \ \ D^*\, of\ D \end{cases} \qquad (4)$$

where, Δ_f is the non-inferiority margin for functional outcomes, it is defined such that Δ_f(t)>0, for all $t\in D$. The inequality of two functions on D is the usual definition of inequality of functions, that is $f_1 \le f_2$ on D when for all x in D, $f1(x) \le f_2(x)$. The functional superiority and noninferiority hypothesis tests in equations 3 and 4 is regarded as one-sided problem for functional means difference. For functional hypothesis, there are basically two approaches, an approach based on pointwise test such as in Xu *et al.* (2018), Sun *et al.* (2015) and Cox and Lee (2007) and a global approach which provide a single summarized for the testing on the whole continuum domain. In the framework of two-sided problem there are global statistic test which have been proposed generally based on a norm such as in Zhang *et al.* (2010); Zhang (2014) based on L^2-norm or Taylor *et al.* (2007) based on *Sup*-norm. Such methods cannot be applied for one-sided hypothesis test, since that could not allow to know the direction of the inequality when the null hypothesis is rejected.

Proposed Global Test Statistics

Denoting by $\hat{f}_1$ and $\hat{f}_2$ the functional mean estimates of f_1 and f_2 respectively and assuming that there are positive and integrable on D, let's define the test following statistic:

$$G = \int_D \left(\hat{f}_1(t) - \hat{f}_2(t) \right) dt \tag{5}$$

On the null hypothesis, the statistic G would be smaller, while on the alternative, it would be larger. The non-parametric bootstrap method which is free-assumption can be used to estimate the distribution of the test statistic G on the null hypothesis H_0' . Then, denoting by α the nominal type 1 error, the null hypothesis is rejected when $G > G_{1-\alpha}$, where $G_{1-\alpha}$ is the $(1-\alpha)$-quantile estimate of the distribution of the statistic G. Similarly to Taylor *et al.* (2007), one may define a global test based on supremum, then let's consider the following statistic:

$$S = Sup_D\left(\hat{f}_1 - \hat{f}_2 \right) \tag{6}$$

Similarly to test statistic G, the statistic G would be smaller, while on the alternative, it would be larger. In this case, the null hypothesis is rejected when $S > S_{1-\alpha}$, where $S_{1-\alpha}$ is the $(1-\alpha)$-quantile estimate of the distribution of the statistic S, which also can be determined using non-parametric bootstrap estimate method. Let's assume there is given two functional data set F_1 and F_2 with sample sizes n_1 and n_2, following description steps are used to determine $G_{1-\alpha}$ and $S_{1-\alpha}$:

1. From original pair of data set F_1 and F_2, make B random samples with replacement and with sizes n_1 and n_2 for each respectively, then get B pairs of sampled data
2. From the B pairs of sampled data, estimate statistics G_i or S_i , i = 1...B
3. Determine $G_{1-\alpha}$ and $S_{1-\alpha}$ by $G^*_{B,1-\alpha}$ and $S^*_{B,1-\alpha}$ respectively such that $\frac{1}{B} card\{j \in \{1,...,B\}, G_j \leq G^*_{B,1-a}\} \approx 1-a$ and $\frac{1}{B} card\ \{j \in \{1,...,B\}, S_j \leq S^*_{B,1-\alpha}\} \approx 1-\alpha$

Simulations

Simulations Settings

The performances of the proposed methods are done trough Monte-Carlo simulations method. Actual type I error rate and statistical power are evaluated according to the sample sizes. In all simulations, with consider equal sample sizes $n_1 = n_2 = n$, where n_1 and n_2 are the sample sizes of groups 1 and 2 respectively, then three cases were considered: n = 30(small), n = 100(medium) and n = 1000(large). The continuum domain considered was D = [0,24]. In the evaluation of type I error rate, data were simulated on the null hypothesis, in this case, it was considered the common function mean for example $f_1 = f_2 = 30t^2+17$, represented in the Fig. 1. The statistical power was evaluated with the data drawn on the alternative hypothesis, three cases represented in the Fig. 2 were considered:

- Case 1: $D^* = D$, the function f_1 was greater than f_2 on the whole domain D. Therefore, it was considered $f_1 = 30t^2+1000$ and $f_2 = 30t^2+1$
- Case 2: $D^* \subset D$, the function f_1 was greater than f_2 on the proper subset D^* = [12, 24] of D, $f_1 = 30t^2$ and $f_2 = 350t + 120$
- Case 3: $D^* \subset D$, the function f_1 was greater than f_2 on the proper subset D^* = [22,24] of D, $f_1 = 30t^2$ and $f_2 = 650t + 220$

In all simulations, the nominal type I error was a = 5%, the co-variance matrix function were assumed equal ($\gamma_1 = \gamma_2 = \gamma$), it was considered correlated data, such that data at two closer points were more correlated than data at two distant points, $\gamma(t, s) = 80^2 exp(-0.5(t-s)^2)$. Following are the Monte-Carlo simulations steps for the estimation of type I error and power:

1. Simulate a pair F_1 and F_2 of functional data with equal sample size n on the domain D, with equal covariance function g and functional means f_1 and f_2 respectively such that null hypothesis is satisfied for the type I error estimation (such that the alternative is satisfied for the power estimation)
2. From the pair of sample F_1 and F_2, estimated the statistic G and S respectively in the Equation 5 and 6. The bootstrap procedure described in 2.2 is used to get the quantiles G_{1-a} and S_{1-a} of the distribution of the statistics G and S respectively. Then, reject the null hypothesis hypothesis if $G > G_{1-a}$ or $S > S_{1-a}$
3. Repeat **1** and **2** N times and get N test decisions $G^i > G^i_{1-\alpha}\ or\ S^i > S^i_{1-\alpha}, i = 1,...,N$
4. The type I error is then estimated by: $\hat{a} = \sum_{i=1}^N \mathbf{1}_{G^i > G^i_{1-\alpha}}\ or\ \hat{a} = \sum_{i=1}^N \mathbf{1}_{S^i > S^i_{1-\alpha}}$. The power is estimated with the same formulas, but with data generated on the alternative hypothesis at step 1

The number of bootstrap replication was B = 1000 and the number of replicated samples was N = 10000. The *R* software programming language 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 especially useful for the simulations and the manipulations of functional process data. The programming codes can be provided on the demand of the user.

Simulations Results

The type I error rate and statistical power estimates are summarized in the Table 1 and 2 respectively. The two global tests with small sample sizes would be somehow liberal as the actual type I error in this case is about 6% and 6:5% for the G and S test statistics respectively. However, as the sample sizes get larger, the actual type I error estimated gets closer to the fixed nominal type I error 5%.

The power was estimated considering some cases of the alternative hypothesis. Based on the considered example of functional means on the alternative hypothesis, the S statistic test got good performances with a power equal to 100% whatever the sample sizes. While the G test statistics got good performances in the cases where D* = D and D* = [12, 24] and very poor performances when the length of D* is smaller.

Table 1: Estimation of the actual type I error rate according to the sample sizes for both proposed test statistics *G* and *S* respectively

	Test statistic *G*	Test statistic *S*
$n_N = n_R = 30$	0.062	0.065
$n_N = n_R = 100$	0.055	0.06
$n_N = n_R = 1000$	0.051	0.54

Table 2: Estimation of the statistical power according to the sample sizes and three different cases of the alternative hypothesis, for the both proposed test statistics *G* and *S* respectively

	Test statistic *G*	Test statistic *S*
	Case 1	
$n_N = n_R = 30$	1	1
$n_N = n_R = 100$	1	1
$n_N = n_R = 1000$	1	1
	Case 2	
$n_N = n_R = 30$	1	1
$n_N = n_R = 100$	1	1
$n_N = n_R = 1000$	1	1
	Case 3	
$n_N = n_R = 30$	0	1
$n_N = n_R = 100$	0	1
$n_N = n_R = 1000$	0	1

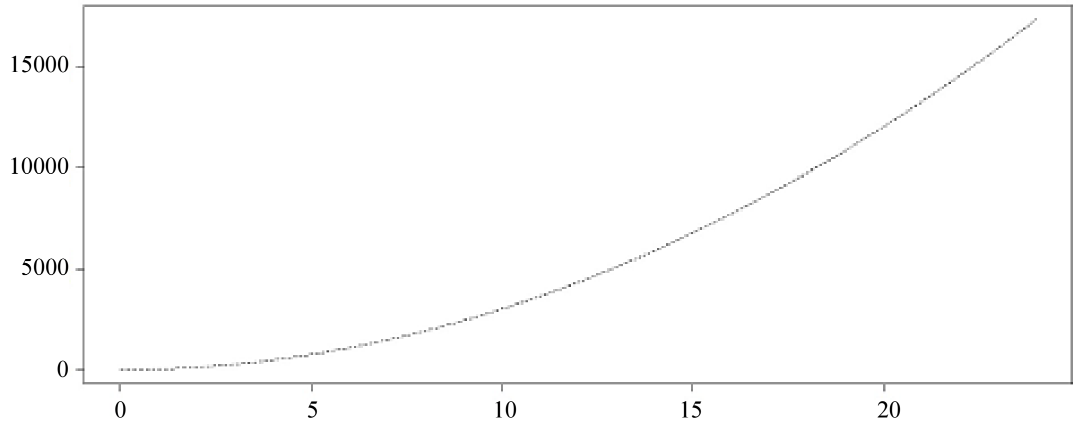

Fig. 1: Functional mean used for samples replication for actual type I error estimation. The functional means f_1 and f_2 are such that the null hypothesis is satisfied and the boundary of the null hypothesis was considered $f_1 = f_2$

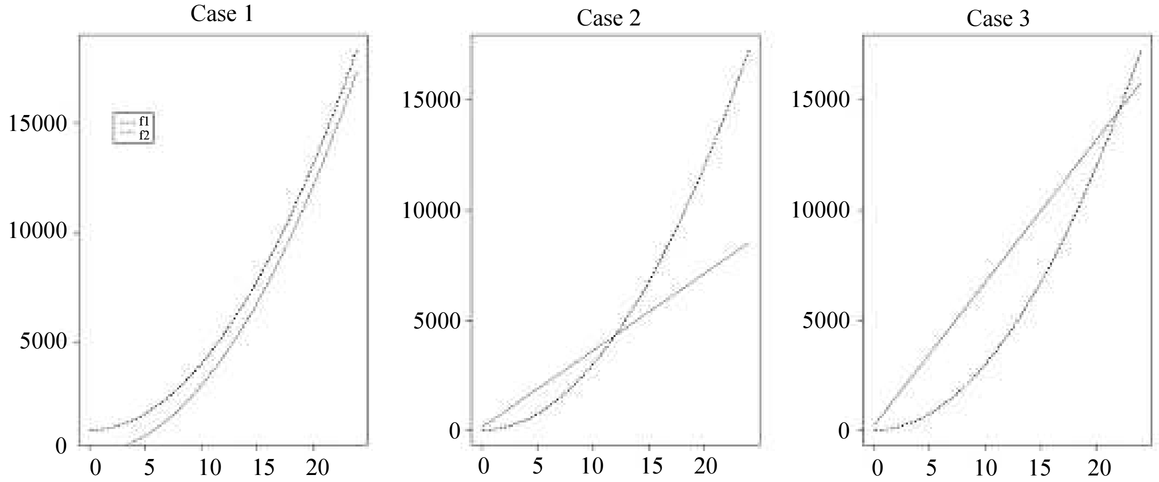

Fig. 2: Functional means used for samples replication for power estimation. The functional means f_1 and f_2 are such that the alternative hypothesis is satisfied and three cases are considered

Conclusion

This work has introduced the superiority and non-inferiority hypothesis test with functional data endpoints. Those hypothesis tests have been regarded as a one-sided problem in functional means difference, where the global test has not been addressed in the literature. Under the assumption of positive functional means in both groups comparison, two global tests statistics were proposed: a test based on the integral and a test based on the supremum. The performances of proposed tests were evaluated through a simulation example, by computing the actual type I error rate and statistical power according to the sample sizes. The two tests got acceptable actual type I error, especially when sample sizes tend to be larger.

However, the test based on supremum got good power performances whatever the sample sizes and whatever the case of alternative, while the test based on integral is very poor for some cases of the true alternative. Therefore, in practice, one would recommend or prefer the test based on supremum.

The assumption of positive mean functions in both treatment groups is satisfied in many practical situations. That is the case where higher values of the endpoints are preferred, when the increase of the endpoint corresponds to more efficiency, for example, red blood cells increase, CD4 count cells, etc... However, it is suitable to get aglobal test in a more generalized context, that can be addressed in future researches works. This pioneering work introducing non-inferiority test with functional data endpoint leads an interesting avenue of future research works especially by considering methodological aspects such as assay sensitivity, constancy assumption and non-inferiority margin which have been broadly studied for binary and continuous endpoints (Zhang, 2006; Tsong *et al*., 2003; Ng, 2008).

Author's Contributions

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

Jules Brice Tchatchueng-Mbougua: Contributed to the writing and reviewed the manuscript.

Anthony Wanjoya: Contributed to the writing of the m 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 pratical and powerfull approach to multiple testing. J. Royal. Stat. Soc. Series B Methodol., 57: 289-300.

Committee for Proprietary Medicinal Products, 2000. Point to consider on switching between superiority and non-inferiority. Eur. Med. Agency.

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.

Food and Dug Administration, 2016. Non-inferiority clinical trials to establish effectiveness-guidance for industry. Technical report, U.S. Department of Health And Human Services.

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.

Hochberg, Y. and A.C. Tamhane, 1987. Multiple Comparison Procedures. 1st Edn., Wiley, New York.

Ng, T., 2008. Noninferiority hypotheses and choice of noninferiority margin. Stat. Med., 27: 5392-5406.

Ramsay, J.O. and B.W. Silverman, 2005. Functional Data Analysis. 1st Edn., Springer.

Ramsay, J.O., H. Wickham, S. Graves and G. Hooker, 2018. FDA: Functional Data Analysis. R Package Version 2.4.8.

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.

Taylor, J.E., K.J. Worsley and F. Gosselin, 2007. Maxima of discretely sampled random fields with an application to 'bubbles'. Biometrika.

Team, R.C., 2016. A language and environment for statistical computing. R Foundation for statistical computing. Vienna, Austria. http://www.R-project.org/

Tsong, Y., S.J. Wang, H.M. Hung and L. Cui, 2003. Statistical issues on objectives, designs and analysis of non-inferiority test active controlled clinical trials. J. Biopharmaceutical Stat., 13: 29-41.

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.

Zhang, J.T., 2014. Analysis of variance for functional data. Taylor Francis Group, 6000 Broken Sound Parkway NW, Suite 300.

Zhang, J.T., Y. Liang and S. Xiao, 2010. On the two-sample behrens-fisher problem for functional data. J. Stat. Theory Practice, 4: 571-587.

Zhang, Z., 2006. Non-inferiority testing with a variable margin. Bio. J., 48: 948-965.