A RANK-BASED PERMUTATION TEST FOR EQUIVALENCE AND NON-INFERIORITY

Rosa Arboretti

Department of Land Environment, Agriculture and Forestry, University of Padova, Italy

Eleonora Carrozzo[1]

Department of Management and Engineering, University of Padova, Italy.

Devin Caughey

Department of Political Science, Massachusetts Institute of Technology, USA.

Received: 28th November 2014/Accepted: 7th September 2015

Abstract. *Testing for the equivalence of two treatments has received attention in recent literature. Solutions typically considered are based on likelihood methods and the intersection-union (IU) principle. The IU principle focuses on the investigation of the equivalence between two treatments; that is a true equivalence has to be identified with a probability converging to one. The goal of this paper is to propose a rank-based permutation test through the Nonparametric Combination (NPC) of dependent tests, as an alternative to likelihood techniques.*

Keywords: *Intersection-union principle, Nonparametric combination, Permutation test, Testing for equivalence.*

1. INTRODUCTION

The need to test for the equivalence of two treatments occurs very frequently in many areas of applied research. In general, let us suppose we have a new treatment *A* with (constant) effect δ_A and a comparative treatment *B* with effect δ_B and we wish to test whether these two treatments may be considered equivalent or not. Following current literature (see, e.g., Berger, 1982; Berger and Hsu,1996; D'Agostino et al., 2003; Hung and Wang, 2009; Julious, 2010;

[1] Eleonora Carrozzo, email: carrozzo@gest.unipd.it

Laster and Johnson, 2003; Liu et al., 2002; Metha et al., 1984; Wellek, 2010; Zhong et al., 2012), this kind of problem is approached by the *intersection-union* (IU) principle. According to this principle the alternative hypothesis states that the difference between two treatments lies inside a given interval, namely the *equivalence interval*, of differences considered acceptable, whereas the null hypothesis states that the difference between the treatments lies outside that interval. Formally, the set of hypotheses can be written as:

$$\begin{cases} H_0: [(\delta_A \leq \delta_B - \varepsilon_I) \text{OR } (\delta_A \geq \delta_B + \varepsilon_S)] \\ H_1: (\delta_B - \varepsilon_I < \delta_A < \delta_B + \varepsilon_S) \end{cases} \tag{1}$$

where $\varepsilon_I > 0$ and $\varepsilon_S > 0$ are respectively the non-inferior and non-superior limits for the difference $\delta_A - \delta_B$. Let us consider the following partial sub-hypotheses:

$$\begin{cases} H_{0(I)}: (\delta_A \leq \delta_B - \varepsilon_I) \\ H_{1(I)}: (\delta_A > \delta_B - \varepsilon_I) \end{cases} \tag{2}$$

$$\begin{cases} H_{0(S)}: (\delta_A \geq \delta_B + \varepsilon_S) \\ H_{1(S)}: (\delta_A < \delta_B + \varepsilon_S) \end{cases} \tag{3}$$

Note that the null hypothesis of the IU test is $H_0 = H_{0(I)} \cup H_{0(S)}$; that is, it is true if only one between the sub-null hypotheses $H_{0(I)}$ and $H_{0(S)}$ is true. Similarly, the alternative hypothesis $H_1 = H_{1(I)} \cap H_{1(S)}$ is true if both sub-alternatives $H_{1(I)}$ and $H_{1(S)}$ are jointly true.

Following the IU principle the focus is on the equivalence between two treatments. When the alternative hypothesis is true, we want to identify it with a probability converging to one for divergent sample sizes. It is worth noting that when $\varepsilon_I = \varepsilon_S = 0$ the test does not admit any solution. As a matter of fact in this case there are no points in the set of the possible values for (δ_A, δ_B) which satisfy the alternative (it is empty $H_1: \emptyset$). It is possible to consider $H_0: [(\delta_A < \delta_B - \varepsilon_I) \text{OR } (\delta_A > \delta_B + \varepsilon_S)]$ instead of $H_0: [(\delta_A \leq \delta_B - \varepsilon_I) \text{OR } (\delta_A \geq \delta_B + \varepsilon_S)]$ against $H_1: \delta_A = \delta_B$ in order to reduce the effect of the drawback. But in this case when the involved variables are continuous, the probability of accepting H_1 when it is true is bounded at α for whatever sample size, and thus it is not possible to provide a valid solution because the test statistic is not consistent.

In this paper we propose a rank-based permutation test for testing for equivalence in the framework of the IU approach. We use the *Nonparametric Combination* (NPC) methodology (Bertoluzzo et al., 2013, Pesarin and Salmaso, 2010) to cope with *Multi-Aspect* Testing (MAT). In MAT a given inferential problem is analysed by at most a countable set of concurrent different partial tests, each of which highlights a different aspect useful for the analysis (Brombin et al., 2011; Marozzi, 2004; Marozzi et al., 2006; Salmaso et al., 2005). The aim of the paper is to go beyond the difficulties of the likelihood based methods.

Thus, in order to introduce our proposal, let us assume, without loss of generality, to have one endpoint variable *X* and a two-samples design. Let there be n_1 the IID observations from X_1 related to treatment *A* and n_2 the IID observations from X_2 related to treatment *B*. Let us also assume that the two variables X_1 and X_2 have the underling common variable *X* and that they differ only for a shift (which implies a constant treatment effect). Thus we define $X_1 = X + \delta_A$ and $X_2 = X + \delta_B$ and let $\boldsymbol{X}_1 = \left(X_{11}, \dots, X_{1n_1}\right)$ and $\boldsymbol{X}_2 = \left(X_{11}, \dots, X_{1n_2}\right)$ be respectively the sample from *A* and the sample from *B*.

2. IU PERMUTATION TEST

In this section we present our approach to deal with the hypotheses testing (1). The idea at the base of the proposal is to test separately albeit simultaneously the sub-hypotheses (2) and (3).

2.1 TESTING $\boldsymbol{H_{0(I)}}$ against $\boldsymbol{H_{1(I)}}$

For testing for $H_{0(I)}$ against $H_{1(I)}$ let us consider a test statistic based on the difference of means, denoted T_I, computed after some apposite transformations of the data. Let us consider the following transformations of original observations: $\boldsymbol{X}_{I1} = \boldsymbol{X}_1$ and $\boldsymbol{X}_{I2} = \boldsymbol{X}_2 + \varepsilon_I$. Then consider a rank transformation of the two samples $\boldsymbol{X}_{I1}$ and $\boldsymbol{X}_{I2}$. For the sake of simplicity let us continue to refer to the two transformed samples as $\boldsymbol{X}_{I1}$ and $\boldsymbol{X}_{I2}$. Note that we can see the hypotheses testing $H_{0(I)}: (\delta_A \leq \delta_B - \varepsilon_I)$ against $H_{1(I)}: (\delta_A > \delta_B - \varepsilon_I) \equiv F_{X_{I2}} > F_{X_{I1}}$ as a test for stochastic dominance. Thus a suitable test statistic is $T_I = \overline{\boldsymbol{X}}_{I2} - \overline{\boldsymbol{X}}_{I1}$, where $\overline{\boldsymbol{X}}_{Ij} = \frac{1}{n_j} \sum_{i \leq n_j} X_{Iji}\, , j = 1{,}2$. Large values of T_I are significant in favour of the alternative $H_{1(I)}$.

2.2 TESTING $\boldsymbol{H_{0(S)}}$ against $\boldsymbol{H_{1(S)}}$

For testing $H_{0(S)}$ against $H_{1(S)}$ let us consider a test statistic based on the difference of means, denoted T_S, computed after some appropriate transformations of the data. As above, let us consider the following transformations of the original observations: $\boldsymbol{X}_{S1} = \boldsymbol{X}_1$ and $\boldsymbol{X}_{S2} = \boldsymbol{X}_2 - \varepsilon_S$. Then consider a rank transformation of the two samples $\boldsymbol{X}_{S1}$ and $\boldsymbol{X}_{S2}$. For the sake of simplicity let us continue to refer to the two transformed samples as $\boldsymbol{X}_{S1}$ and $\boldsymbol{X}_{S2}$. Note that we can see the hypotheses testing $H_{0(S)}: (\delta_A \geq \delta_B + \varepsilon_S)$ against $H_{1(S)}: (\delta_A < \delta_B + \varepsilon_S) \equiv F_{X_{S2}} > F_{X_{S1}}$ as a test for stochastic dominance. Thus a suitable test statistic is $T_S = \overline{\boldsymbol{X}}_{S1} - \overline{\boldsymbol{X}}_{S2}$, where $\overline{\boldsymbol{X}}_{Sj} = 1/n_j \sum_{i \leq n_j} X_{Sji}$, $j = 1,2$. Large values of T_S are significant in favour of the alternative $H_{1(S)}$.

It is worth noting that the equivalence margins are expressed in the same unit of measurement of the observed variables. This fact represent an important point since in practice it is important to do not lose the precise meaning of the equivalence margins. Consider a clinical example in which we want to compare the effect of two treatments on the reduction of cholesterol. It is important for the sake of clarity that the interval be expressed in the same units as cholesterol levels are. By contrast, constructing an analogous nonparametric rank-based procedure following the optimal tests in the literature (such as that suggested by Lehman (1986)), requires that the equivalence interval be expressed in terms of ranks, which diminishes the substantive interpretability of the equivalence margins. Generally in the literature (see, e.g., Janssen and Wellek (2010)) the equivalence margins are indeed expressed in terms of standardized Wilcoxon-Mann-Whitney statistic, since transforming original data into ranks leads to a dependence between the tests which is difficult to model parametrically. With the Nonparametric Combination (NPC) methodology (see Pesarin and Salmaso (2010)) considered in this paper it is possible to manage the unknown dependence between the partial tests. Furthermore the rank transformation allows the execution of the test whatever be the underling distribution of the variables.

3. AN ALGORITHM FOR IU RANK-BASED PERMUTATION TEST

To better understand the proposed procedure, in this section we present more in detail the algorithm for the IU rank-based permutation test:

1. Let $\boldsymbol{X} = (\boldsymbol{X}_1, \boldsymbol{X}_2) = (X_i, i = 1, \dots, n); n_1, n_2$ be the set of data from the two groups and ε_I and ε_S the limits of the equivalence interval;
2. Define the two data vectors $\boldsymbol{X}_I = (\boldsymbol{X}_{I1}, \boldsymbol{X}_{I2}) = (X_{I1i} = X_{1i}, i = 1, \dots, n_1; X_{I2i} = X_{2i} - \varepsilon_I, i = 1, \dots, n_2)$ and $\boldsymbol{X}_S = (\boldsymbol{X}_{S1}, \boldsymbol{X}_{S2}) = (X_{S1i} = X_{1i}, i = 1, \dots, n_1; X_{S2i} = X_{2i} - \varepsilon_S, i = 1, \dots, n_2)$;
3. Compute the rank transformation of $\boldsymbol{X}_I$ and $\boldsymbol{X}_S$. For the sake of simplicity let us continue to refer to these transformed data with $\boldsymbol{X}_I$ and $\boldsymbol{X}_S$;
4. Compute and save the two observed values of the two statistics: $T_I^O = \bar{X}_{I1} - \bar{X}_{I2}$ and $T_S^O = \bar{X}_{S2} - \bar{X}_{S1}$;
5. Take a random permutation $\boldsymbol{u}^* = u_1^*, \dots, u_n^*$ of unit labels $\boldsymbol{u} = (1, \dots, n)$;
6. Define both permuted data sets $\boldsymbol{X}_I^* = \left(X_{Iu_i^*}, i = 1, \dots, n; n_1, n_2\right)$ and $\boldsymbol{X}_S^* = \left(X_{Su_i^*}, i = 1, \dots, n; n_1, n_2\right)$ on the same permutation $\boldsymbol{u}^*$;
7. Compute and save the related permuted values of two statistics: $T_I^* = \bar{X}_{I2}^* - \bar{X}_{I1}^*$ and $T_S^* = \bar{X}_{S1}^* - \bar{X}_{S2}^*$;
8. Independently repeat R times steps 5 to 7, obtaining results $[(T_{Ir}^*, T_{Sr}^*), r = 1, \dots, R]$ which simulate the bivariate permutation distribution of two partial tests (T_I, T_S);
9. Calculate two estimates of the two marginal p-value statistics $\lambda_I = \sum_{r=1}^{R} \boldsymbol{I}\left[T_{Ir}^* \geq T_I^O\right]/R$ and $\lambda_S = \sum_{r=1}^{R} \boldsymbol{I}\left[T_{Sr}^* \geq T_S^O\right]/R$ and the φ-combined observed value $\varphi^O = \varphi(\lambda_I, \lambda_S)$, small values of which are evidence in favour of null hypothesis H_0;
10. Transform the simulated bivariate distribution in step 8 into the bivariate empirical significance level function $\boldsymbol{L}^* = [(L_{Ir}^*, L_{Sr}^*), r = 1, \dots, R]$ where $L_{hr}^* = \left\{0.5 + \sum_{b=1}^{R} \boldsymbol{I}(T_{hb}^* \geq T_{hr}^*)\right\}/(R+1), h = I, S$;
11. Define the φ-combined distribution $[\varphi_r^* = \varphi(L_{Ir}^*, L_{Sr}^*), r = 1, \dots, R]$ that simulates the true bivariate permutation distribution of φ where thedependence between (T_I, T_S) is nonparametrically, albeit implicitly, taken into consideration;
12. The global NPC p-value statistic for testing equivalence is defined as $\lambda_\varphi = \sum_{r=1}^{R} \boldsymbol{I}[\varphi_r^* \geq \varphi^O]/R$;
13. If $\lambda_\varphi \leq \alpha$ then reject global H_0 in favour of H_1.

It is worth noting that $H_{0(I)}$ true implies $H_{0(S)}$ false and vice versa, hence the two partial p-values are negatively dependent. Thus the nonparametric combination function $\varphi: [0,1]^2 \rightarrow \mathbb{R}^+$is such that small values of which are significant. Combination functions useful for IU tests have to satisfy the following properties:

- φ is continuous and non-decreasing in each argument, i.e. $\lambda_q < \lambda'_q$ implies $\varphi(\dots, \lambda_q, \dots) < \varphi(\dots, \lambda'_q, \dots)$;
- φ must attain its infimum if all arguments attain zero;
- $\alpha > 0$ and larger than the minimum attainable value (Pesarin et al., 2010) implies the conditional critical value $\varphi_\alpha > 0$.

Some examples of combination functions for the IU rank-based permutation tests are: $\varphi_M = \max(\lambda_I, \lambda_S)$, $\varphi_A = \lambda_I + \lambda_S$, $\varphi_\pi = 1 - (1 - \lambda_I)(1 - \lambda_S)$.

4. A SIMULATION STUDY

In this section we present the results of a simulation study with the aim of demonstrating the general performance of the IU rank-based permutation test. In particular we consider two sample sizes (n=24, n=36), three different data generating distributions (Gaussian($N(0,1)$), Exponential ($\mathrm{Exp}(1)$), and Laplace ($\mathcal{L}(0,1)$)) and different types of equivalence intervals (symmetric and asymmetric). The rejection probability of the permutation test (based on 2000 permutations) at significance levels 0.05 and 0.10 is determined on the basis of 2000 Monte Carlo iterations, under the null hypothesis ($\delta_A = \varepsilon_I, \delta_A = \varepsilon_S$) and under the alternative ($\delta_A = 0$). The value of δ_B was set to zero. The boundaries used in our simulations are those suggested in Wellek (2010). We considered φ_M as combination function.

Tables 1-6 show the results of the simulation study in the case of balanced samples, but tests for unbalanced case were also performed and yielded similar results. From simulation results we can appreciate that nominal significance alpha level is substantially achieved for all distributions considered, both for symmetrical and asymmetrical equivalence intervals.

Table 1. Rejection rate of the IU rank-based permutation test for equivalence at both boundaries of the hypotheses $\delta_A = -\varepsilon_I, \delta_A = \varepsilon_S$and $\delta_A = 0$, with sample size $n_1 = n_2 = 24, 36$. Gaussian distribution and equivalence range $(-\varepsilon_I, \varepsilon_S) = (-0.7416, 0.7416)$.

δ_A	$n_1 = n_2$	$\alpha = 0.05$	$\alpha = 0.10$
ε_I	24	0.0505	0.1075
ε_I	36	0.0495	0.0880
ε_S	24	0.0495	0.1020
ε_S	36	0.0470	0.1010
0	24	0.6095	0.8005
0	36	0.8690	0.9370

Table 2.Rejection rate of the IU rank-based permutation test for equivalence at both boundaries of the hypotheses $\delta_A = -\varepsilon_I, \delta_A = \varepsilon_S$ and $\delta_A = 0$, with sample size $n_1 = n_2 = 24, 36$. Exponential distribution and equivalence range $(-\varepsilon_I, \varepsilon_S) = (-0.5108, 0.5108)$.

δ_A	$n_1 = n_2$	$\alpha = 0.05$	$\alpha = 0.10$
ε_I	24	0.0505	0.1100
ε_I	36	0.0510	0.1050
ε_S	24	0.0475	0.1000
ε_S	36	0.0485	0.0990
0	24	0.5710	0.7450
0	36	0.8150	0.8990

Table 3. Rejection rate of the IU rank-based permutation test for equivalence at both boundaries of the hypotheses $\delta_A = -\varepsilon_I, \delta_A = \varepsilon_S$ and $\delta_A = 0$, with sample size $n_1 = n_2 = 24, 36$. Laplace distribution and equivalence range $(-\varepsilon_I, \varepsilon_S) = (-0.8731, 0.8731)$.

δ_A	$n_1 = n_2$	$\alpha = 0.05$	$\alpha = 0.10$
ε_I	24	0.0395	0.0995
ε_I	36	0.0495	0.0970
ε_S	24	0.0450	0.0955
ε_S	36	0.0490	0.0975
0	24	0.5720	0.7600
0	36	0.8295	0.9215

Table 4. Rejection rate of the IU rank-based permutation test for equivalence at both boundaries of the hypotheses $\delta_A = -\varepsilon_I, \delta_A = \varepsilon_S$ and $\delta_A = 0$, with sample size $n_1 = n_2 = 24, 36$. Gaussian distribution and equivalence range $(-\varepsilon_I, \varepsilon_S) = (-0.5, 1)$.

δ_A	$n_1 = n_2$	$\alpha = 0.05$	$\alpha = 0.10$
ε_I	24	0.0535	0.1035
ε_I	36	0.0450	0.0985
ε_S	24	0.0505	0.0900
ε_S	36	0.0455	0.0860
0	24	0.4670	0.6235
0	36	0.6490	0.7855

Table 5. Rejection rate of the IU rank-based permutation test for equivalence at both boundaries of the hypotheses $\delta_A = -\varepsilon_I, \delta_A = \varepsilon_S$ and $\delta_A = 0$, with sample size $n_1 = n_2 = 24, 36$. Exponential distribution and equivalence range $(-\varepsilon_I, \varepsilon_S) = (-0.3235, 0.7348)$.

δ_A	$n_1 = n_2$	$\alpha = 0.05$	$\alpha = 0.10$
ε_I	24	0.0370	0.0995
ε_I	36	0.0530	0.1095
ε_S	24	0.0490	0.1020
ε_S	36	0.0505	0.0965
0	24	0.4485	0.6415
0	36	0.6755	0.8005

Table 6. Rejection rate of the IU rank-based permutation test for equivalence at both boundaries of the hypotheses $\delta_A = -\varepsilon_I, \delta_A = \varepsilon_S$ and $\delta_A = 0$, with sample size $n_1 = n_2 = 24, 36$. Laplace distribution and equivalence range $(-\varepsilon_I, \varepsilon_S) = (-0.5762, 1.17)$.

δ_A	$n_1 = n_2$	$\alpha = 0.05$	$\alpha = 0.10$
ε_I	24	0.0520	0.1070
ε_I	36	0.0455	0.0885
ε_S	24	0.0485	0.0940
ε_S	36	0.0540	0.1010
0	24	0.4375	0.6075
0	36	0.6375	0.7670

5. AN EXAMPLE APPLICATION

In this section we consider an example application of the IU rank-based permutation test. We want to further check the behavior and give some practical application guidelines of the test. The following example is not referred to a clinical application as is typical for the problem in matter. We want to show an example of equivalence testing problem in a social-political field, in order to take an idea of the different contexts in which the problem may occurs. The application is taken by the work of Caughey and Sekhon (2011), where it is of interest to test whether the congressional districts where the Democratic candidate barely won (and then the Republican barely lost) are equivalent to districts where the Democratic barely lost with respect to different pre-election covariates. We can easily see this example as an equivalence testing problem in which we want to assess if two districts can be considered equivalent on the base of a pseudo-treatment represented by a social variable of interest. We are interested in several variables and thus perform one test for each of them. The variables considered are: 1. the percent urban in the district (UrbanPct); 2. difference between the Democratic and Republican percentage of the presidential vote in the district, averaged over all elections in that decade (DifPVDec); 3. percent of the district population that worked for the government (GovWkPct); 4. percent black in the district (BlackPct); 5. percent foreign-born in the district (ForgnPct). The data consist of U.S. congressional elections decided by a margin less than 0.5%; see Caughey and Sekhon (2011) for details on the data. For each variable we tested equivalence in an interval of $\mp s/5$ where s is the standard deviation of the pooled data.

Table 7 shows for each covariate the equivalence interval and the corresponding p-value resulting from the IU rank-based permutation test. Fixed the significance level $\alpha = 0.05$, we can see from the results that there is no evidence of equivalence between the two congressional districts regards the five covariates (all p-values > 0.05).

Table 7. Results of the IU rank-based test applied to the example

Covariates	$(\varepsilon_I, \varepsilon_S)$	p-value
1. UrbanPct	**(-4.574, 4.574)**	**0.569**
2. DifPVDec	**(-0.031, 0.031)**	**0.307**
3. GovWkPct	**(-0.511, 0.511)**	**0.657**
4. BlackPct	**(-1.346, 1.346)**	**0.193**
5. ForgnPct	**(-0.949, 0.949)**	**0.245**

6. CONCLUSIONS

In this paper we presented a solution to testing equivalence and non-inferiority under IU approach. Because the proposed procedure is rank-based, the IU test can be performed independently of the underlying distribution of the variables involved in the problem. By contrast the parametric procedures proposed by Lehman can be applied only to distributions belonging to the regular exponential family. Following the results in Pesarin and Salmaso (2010) concerning the NPC methodology we are able to deal with this intriguing problems going beyond the likelihood-based methods and facing the generally too complex dependence structure of the several partial test statistics in which such analysis is often broken down. With the NPC methodology we are also able to face with dependent partial tests with a dependency structure more complicated respect to that linear, and we do not need to know it. Another important issue our proposed method overcomes is proving an easy interpretation of the results. This is done by the fact that equivalence margins may be expressed in the same unit of measurement of the observed variables.

Our proposed method is easy to be extended to one sample and C>2 sample designs, to ordered categorical end-point variables, to repeated measures, to multidimensional settings, to situations with missing or censored data using the corresponding solution described in Pesarin and Salmaso (2010, 2011) on MAT and NPC and properly modifying the combination function. We usually describe the problem of testing for equivalence speaking of two treatments to compare. Typically we refer to the treatments as drugs. With the applicative example in Section 5 we show that the term treatment may refer to various things, and hence the problem of testing equivalence may occur in different fields, such as social and political science.

Finally, we note that the problem of testing for equivalence of two treatments, can also be approached following a different principle, Roy's *union-intersection* (UI) principle. Under this perspective the alternative is that the effect of a new treatment lies outside the equivalence interval, $\widetilde{H}_1: [(\delta_A < \delta_B - \varepsilon_I) OR (\delta_A > \delta_B + \varepsilon_S)]$ (i.e. when treatments are substantially non-equivalent), and the null hypothesis is that the new effect lies inside the equivalence interval, $\widetilde{H}_0: (\delta_B - \varepsilon_I \leq \delta_A \leq \delta_B + \varepsilon_S)$ (i.e. the two treatments do not differ substantially). Following this alternative approach the logical drawback explained in Section 1 for $\varepsilon_I = \varepsilon_S = 0$ does not occur, since it becomes the traditional two-sided testing $\widetilde{H}_0: \delta_A = \delta_B$ against $\widetilde{H}_1: \delta_A \neq \delta_B$. A useful permutation solution following the UI principle is described in Pesarin et al. (2015).

REFERENCES

Berger, R.L. (1982). Multiparameter hypothesis testing and acceptance sampling. In *Technometrics* 24: 295-300.

Berger, R.L. and Hsu, J.C. (1996). Bioequivalence trials, intersection-union tests and equivalence confidence sets. In *Statistical Science* 11: 283-319.

Bertoluzzo, F., Pesarin, F. and Salmaso, L. (2013). On multi-sided permutation tests. In *Communications in Statistics – Simulation and Computation* 42(6): 1380-1390.

Brombin, C., Salmaso, L., Ferronato, G. and Galzignato, P.F. (2011). Multi-aspect procedures for paired data with application to biometric morphing. In *Communications in Statistics – Simulation and Computation* 40:1-12.

Caughey, D., and Sekhon, J.S. (2011). Elections and the regression discontinuity design: Lessons from close U.S. House races, 1942–2008. In *Political Analysis* 19(4): 385–408.

D'Agostino, R.B., Massaro, J.M. and Sullivan, L.M. (2003). Non-inferiority trials: Design concepts and issues – the encounters of academic consultants in statistics. In *Statistics in Medicine* 22: 169-186.

Hung, H.M.J. and Wang, S.U. (2009). Some controversial multiple testing problems in regulatory applications. In *Journal of Biopharmaceutical Statistics* 19: 1-11.

Janssen, A. and Wellek, S. (2010). Exact linear rank tests for two-sample equivalence problems with continuous data. In *Statistica Neerlandica* 64: 482–504.

Julius, S.A. (2010). *Sample Sizes for Clinical Trials.* Chapman & Hall/CRC, Boca Ratorn, USA.

Laster, L.L. and Johnson, M.F. (2003). Non-inferiority trials: the 'at least as good as' criterion. In'*Statistics in Medicine* 22: 187-200.

Lheman, E. L., (1986). *Testing Statistical Hypotheses*. 2nd' Edition. Wiley, New York.

Liu, J-p.,Hsueh, H-m., Hsieh, E., and Chen, J.J. (2002). Tests for equivalence or non-inferiority for paired binary data. In'*Statistics in Medicine* 21: 231-245.

Marozzi, M. (2004). A bi-aspect nonparametric test for the two-sample location problem. In *Computational Statistics & Data Analysis* 44(4): 639-648.

Marozzi, M., Salmaso L. (2006). Multivariate bi-aspect testing for two-sample location problem. In *Communication in Statistics – Theory and Methods* 35(3): 477-488.

Mehta, C.R., Patel, N.R and Tsiatis, A.A. (1984). Exact significance testing to establish treatment equivalence with ordered categorical data. In *Biometrics* 40: 819-825.

Pesarin, F., Salmaso, L., Carrozzo, E. and Arboretti, R., (2015). Union-Intersection permutation solution for two-sample equivalence testing. In *Statistics and Computing*. DOI. 10.1007/s11222-015-9552-y.

Pesarin, F. and Salmaso, L. (2011). A review and some new results on permutation testing for multivariate problem. In *Statistics and Computing* 22(2): 639-646.

Pesarin, F. and Salmaso, L. (2010). *Permutation Tests for Complex Data, Theory, Applications and Software.* Wiley: Chichester, UK.

Salmaso, L., Solari, A. (2005) Multiple aspect testing for case-control designs. In *Metrika* 62: 331-340.

Wellek, S. (2010). *Testing Statistical Hypotheses of Equivalence and Noninferiority.* Chapman & Hall/CRC: Boca Raton, USA.

Zhong, Z., Chen, W. and Jin, H. (2012). A new test for testing non-inferiority in matched-pairs design. In *Communications in Statistics – Simulation and Computation* 41: 1557-1565.