

On Some Nonparametric Tests For Partially Observed Correlated Data: Proposing New Tests

Hani M. Samawi

Department of Biostatistics
Jiann-Ping Hsu College Public Health
Georgia Southern University
P.O. Box 8015
Statesboro, GA 30460, USA
hsamawi@georgiasouthern.edu

Lili Yu

Department of Biostatistics
Jiann-Ping Hsu College Public Health
Georgia Southern University
P.O. Box 8015
Statesboro, GA 30460, USA

Robert Vogel

Department of Biostatistics
Jiann-Ping Hsu College Public Health
Georgia Southern University
P.O. Box 8015
Statesboro, GA 30460, USA

Received 27 July 2013

Accepted 5 January 2015

Abstract

Correlated or matched data is frequently collected under many study designs in applied sciences such as the social, behavioral, economic, biological, medical, epidemiologic, health, public health, and drug developmental sciences in order to have more efficient design and to control for potential confounding factors in the study. Challenges with respect to availability and cost commonly occur with matching observational or experimental study subjects, thus researchers frequently encounter situations where the observed sample consists of a combination of correlated and uncorrelated data due to missing responses (partially correlated data). Ignoring cases with missing responses, when analyzing the data, will introduce bias in the inference and lower the power of the testing procedure. This paper discusses and proposes some nonparametric testing procedures to handle data when partially observed correlated data is available without ignoring the cases with missing responses. Therefore, we will introduce more powerful testing

procedures which combine all cases in the study. A theoretical and numerical investigation will be provided. The proposed testing procedures will be applied to simulated and genetic data. This is an important research area and solving one of the challenging issues in missing data problem will enrich the research in this area and help the researcher to have more powerful test procedure.

Key Words: Power of the test, P-value of the test, efficiency, Matched pairs sign test, Sign test, Wilcoxon sign-rank test, Correlated and uncorrelated data

2010 Mathematics Subject Classification: 62G10

1. Introduction

The importance of statistical inferential methods and modeling approaches to the social, behavioral, economic, biological, medical, epidemiologic, health, public health, and drug developmental sciences has grown exponentially in the last few decades. Many study designs in the aforementioned applied sciences give rise to correlated and partially observed correlated data due to missing responses. For correlated data, subjects are matched to controls for confounding factors, and/or subjects are followed over time, are repeatedly treated under different experimental conditions, or are observed in logical units (e.g., clinics, families, litters). One assumption to consider is that observations are missing completely at random (MCAR), see for example [3] and [4]. However, Akritas et al [1] consider another missing value mechanism, missing at random (MAR). For quantitative responses, statistical methods, including linear and nonlinear models, are quite established for correlated data. However, for partially observed correlated data there are concerns which need addressing due to the complexity of the analysis, in particular for small sample sizes and when a normality assumption of the underlying populations is not valid.

As an example of partially correlated data for the MCAR design, consider the case where the researcher compares two different treatment regimens for eye redness or allergy and randomly assigns one treatment to each eye for each experimental subject. Some patients may drop out after the first treatment, while other patients may refuse to receive the first treatment, and then receive only the second treatment. In this situation, we may have two groups of patients: the first group of patients who received both treatments in each eye, and are considered as paired matched data; and the second group who received only one of the treatments in one of the eyes, and are considered as unmatched data. This study design is illustrated in Table 1.

Table 1.1 Partially matched study design.

Paired Subject	Treatment 1	Treatment 2
1	Yes	Yes
2	Yes	Yes
3	Yes	Yes
4	Yes	No
5	Yes	No
6	Yes	No
7	No	Yes
8	No	Yes
9	No	Yes
10	No	Yes

Moreover, additional examples for partially observed correlated data can be found in the literature (see for example [6], [16] and [20]). Several authors have presented various tests considering the problem of estimating the difference of means of a bivariate normal distribution when some observations corresponding to both variables are missing. Under the assumption of bivariate normality and MCAR, Ekbohm [8] summarized five procedures for testing the equality of two means. Using Monte Carlo results [8] indicated that the two tests based on a modified maximum likelihood estimator are preferred: one due to Lin and Stivers [13] when the number of complete pairs is large, and the other proposed in Ekbohm's paper otherwise, provided the variances of the two responses do not differ substantially. When the correlation coefficient between the two responses is small, two other tests may be used: a test proposed by Ekbohm when the homoscedasticity assumption is not strongly violated, and otherwise a Welch-type statistic suggested by [13] (for further discussion, see [8]).

Alternatively, researchers tend to ignore some of the data – either the correlated or the uncorrelated data depending on the size of each subset. However, Looney and Jones [15] argued that ignoring some of the correlated observations would bias the estimation of the variance of the difference in treatment means and would dramatically affect the performance of the statistical test in terms of controlling type I error rates and statistical power (see [19]). They propose a corrected z-test method to overcome the challenges created by ignoring some of the correlated observations. However, our preliminary investigation shows that the method of Looney and Jones [15] pertains to large samples and is not the most powerful test procedure. Furthermore,

Rempala and Looney [17] studied asymptotic properties of a two-sample randomized test for partially observed dependent data. They indicated that a linear combination of randomized t-tests is asymptotically valid and can be used for non-normal data. However, the large sample permutation tests are difficult to perform and only have some optimal asymptotic properties in the Gaussian family of distributions when the correlation between the paired observations is positive.

The aforementioned methods cannot be used for non-normal and moderate or small sample size data. There have been several attempts to provide nonparametric test procedures under MCAR and MAR designs (for example, see [3], [4], [1], [11] and [21]) some of those test are presented below. However, there is still a need for intensive investigation to develop more powerful nonparametric testing procedures for MCAR and MAR designs.

1.1. A Review of nonparametric methods for “combined” correlated and uncorrelated data

We have summarized five nonparametric found in the literature that compares medians in a paired case with unpaired data on both responses under MCAR design (see [14]). Those methods are found in [11] and [21].

The general setting of the two-sample problem is as follows: Suppose there are n paired subjects under methods 1 and 2. Let $\{(U_1, V_1), (U_2, V_2), \dots, (U_n, V_n)\}$ denote observed values of the paired subjects from a continuous bivariate distribution whose marginal distributions have the same shape. Let $\{X_1, X_2, \dots, X_{n_1}\}$ and $\{Y_1, Y_2, \dots, Y_{n_2}\}$ denote observed values for independent subjects exposed to method 1 and method 2, respectively. Assume that x - and u - observations originate from a common $F_X(x)$ parent population, and y - and v - observations originate from a (possibly different) common $F_Y(y)$ parent population but have the same shape. Suppose we are interested in testing the one-sided hypotheses throughout the paper $H_0 : \theta_1 = \theta_2$ and $H_1 : \theta_1 > \theta_2$ where θ_1 is a measure of location (median) of $F_X(x)$ and θ_2 is a measure of location (median) of $F_Y(y)$.

1) Test statistic S_1 : proposed by J. Wilson (unpublished work, see [21])

This test statistic is based on a sign test for the complete pairs and a Wilcoxon rank sum test for the unpaired cases. For complete pairs, two observed values within each pair are ranked, and the ranks are summed for one group. In this paper, ranks of group 1 are used throughout the analyses. Let $R_{1i} = \begin{cases} 1 & \text{if } u_i < v_i, i = 1, 2, \dots, n \\ 2 & \text{if } u_i > v_i, i = 1, 2, \dots, n. \end{cases}$

The test statistic for complete pairs is $\sum_{i=1}^n R_{1i}$. For unpaired cases, ranks are made among the total number of $n_1 + n_2$ unpaired observations and summed for the same group of X 's - group 1.

Let $R_{2j}, j = 1, 2, \dots, n_1$, be the rank of each observation in group 1 in the combined sample; then, the test statistic for unpaired data is $\sum_{j=1}^{n_1} R_{2j}$. Hence, the S_1 test statistic for the combined data is

defined as $S_1 = \sum_{i=1}^n R_{1i} + \sum_{j=1}^{n_1} R_{2j}$ (see [21]).

2) Test statistic S_2 : proposed by Brunner and Neumann [2]

This test statistic is based on a Wilcoxon rank sum test for both complete pairs and unpaired cases. For complete pairs, all the paired data from U and V are combined. Let $R_{3i}, i = 1, 2, \dots, n$, denote the rank of each observation in group 1 in the combined paired sample; then, the test statistic for the paired data is $\sum_{i=1}^n R_{3i}$. This rank sum test statistic for complete pairs does not

take pairing into consideration. Hence, the test statistic S_2 is defined as $S_2 = \sum_{i=1}^n R_{3i} + \sum_{j=1}^{n_1} R_{2j}$.

3) Test statistic S_3 : proposed by Sen, P.K. [18]

This test statistic is based on an aligned rank sum test (see also, [12]) for the complete pairs and a Wilcoxon rank sum test for the unpaired cases. For complete pairs, the mean of each pair is subtracted from the original observed values before ranking. All the modified paired data from U and V are then combined. Let $R_{4i}, i = 1, 2, \dots, n$, be the rank of each observation in group 1 in

the modified combined paired sample; then, the test statistic for complete pairs is $\sum_{i=1}^n R_{4i}$. Hence,

the test statistic S_3 is defined as $S_3 = \sum_{i=1}^n R_{4i} + \sum_{j=1}^{n_1} R_{2j}$.

4) Test statistic S_4 : proposed by Im KyungAh ([11], Master Thesis)

For this test statistic, if we assume that U and V are exchangeable random variables, then $U - V$ and $V - U$ both have symmetric distributions and the Wilcoxon procedure is clearly justified. Let $D_i = U_i - V_i, i = 1, 2, \dots, n$. Under H_0 , a similar procedure as in the Wilcoxon signed-rank test can be used. Let W_i be an indicator for when $|D|_{(i)}$ corresponds to a positive observation, where $|D|_{(1)} < \dots < |D|_{(n)}$ are the ordered absolute values. The [11] statistic is

defined as $\sum_{i=1}^n R_i s(D_i) - \sum_{i=1}^n R_i [1 - s(D_i)]$, which is a version of the Wilcoxon signed-rank statistic,

where R_i is the rank of $|D_i|$. Hence, the test statistic S_4 is defined as:

$$S_4 = \sum_{i=1}^n R_i s(D_i) - \sum_{i=1}^n R_i [1 - s(D_i)] + \left\{ \sum_{j=1}^{n_1} R_{2j} - \frac{n_1(n_1 + n_2 + 1)}{2} \right\}.$$

5) Test statistics S_5 : proposed by John Bryant (see [21], Master thesis).

In addition to the previous four test statistics, there is another option based on the Mann-Whitney U statistic. In particular, U is the number of times an x exceeds a y . It can be computed as follows:

$$U = \sum_{j=1}^{n_1} \sum_{k=1}^{n_2} I(X_j > Y_k), \text{ where } I(X_j > Y_k) = \begin{cases} 1 & \text{if } X_j > Y_k \\ 0 & \text{otherwise} \end{cases}.$$

Mann and Whitney have previously showed that the tests based on U are equivalent to tests based on W (the rank sum) in the case of no ties [9]. However, instead of using the

Mann-Whitney U test, they used the rank sum test. The data consisting of both complete pairs and unpaired observations are combined and a rank, R_{5i} , is defined for each observation in group 1 in the combined sample including both observed values of x 's and observed values of u 's. The test statistic

is then defined as $S_5 = \sum_{i=1}^{n+n_1} R_{5i}$.

This paper discusses and proposes some nonparametric testing procedures to handle data when partially observed correlated data is available without ignoring the cases with missing

responses. We will introduce more powerful testing procedure which combined all cases in the study. These procedures will be of special importance in meta-analysis where partially observed correlated data is a concern when combining results of various studies. Section 2 and 3 will propose new tests based on combining two-sample test with paired matched test. In section 4 we propose new test procedure. Numerical analysis and final remarks will be presented in section 5, while the illustration by using a real data will be given in section 6.

2. Combined Sign Tests for Correlated and Uncorrelated Data: Proposed Methods

2.1. Sign test for correlated data with no missing values

Correlated data consists of observations in a bivariate random sample $\{(U_i, V_i), i = 1, 2, \dots, n\}$, where there are n pairs of observations. A comparison is made within each pair (U, V) , and the pair is classified as “+” if $U > V$, or “-” if $U < V$. The underlying populations are assumed absolute continuous. Therefore, no ties are assumed. In this paper we assume the MCAR design and the marginal distributions have the same shape ($F_X(x)$ and $F_Y(y)$) to test the one-sided hypotheses $H_0 : \theta_1 = \theta_2$ and $H_1 : \theta_1 > \theta_2$ where θ_1 is a measure of location (median) of $F_X(x)$ and θ_2 is a measure of location (median) of $F_Y(y)$ (see also [5]). The matched pairs sign test statistic, denoted by T_1 , for testing the above hypotheses, equals the number of “+” pairs:

$$T_1 = \sum_{i=1}^n I(U_i > V_i) \tag{1}$$

where $I(U_i > V_i) = \begin{cases} 1 & \text{if } U_i > V_i \\ 0 & \text{otherwise} \end{cases}$.

Alternatively, define $D_i = U_i - V_i, i = 1, 2, \dots, n$. Then, the null hypothesis (H_0 : The median of the differences is zero) can be tested using the sign test. Therefore, the test

statistic can be written as: $T_1 = \sum_{i=1}^n I(D_i > 0)$. All tied pairs are discarded, and n represents the number of remaining pairs. Depending on whether the alternative hypothesis is one- or two-tailed, and if $n \leq 20$, then use the binomial distribution (i.e., $Bin(n, p = \frac{1}{2})$) for finding the critical

region of approximately size α . Under H_0 and for $n > 20$, $T_1 \sim N(\frac{n}{2}, \frac{n}{4})$. Therefore, the critical region can be defined based on the normal distribution:

$$Z_1 = \frac{T_1 - (\frac{n}{2})}{\sqrt{\frac{n}{4}}} \xrightarrow{L} N(0,1). \tag{2}$$

2.2. Mann-Whitney Wilcoxon test for uncorrelated data with no missing values data

For uncorrelated data, let $\{X_1, X_2, \dots, X_{n_1}\}$ and $\{Y_1, Y_2, \dots, Y_{n_2}\}$ denote two independent simple random samples of subjects exposed to method 1 and method 2, respectively. It can be shown that $H_0 : \theta_1 = \theta_2$ is valid for two-independent samples, where θ_1 is a measure of location (median) of $F_X(x)$ and θ_2 is a measure of location (median) of $F_Y(y)$. If the distributions of X and Y have the same shape, then the null hypothesis of interest is $H_0 : \theta_1 - \theta_2 = 0$. Define:

$$T_2 = \sum_{j=1}^{n_1} \sum_{k=1}^{n_2} I(X_j > Y_k), \tag{3}$$

where $I(X_j > Y_k) = \begin{cases} 1 & \text{if } X_j > Y_k \\ 0 & \text{otherwise.} \end{cases}$

Then, T_2 is the Mann-Whitney Wilcoxon two samples test. Therefore, $E(T_2) = \frac{n_1 n_2}{2}$ and $Var(T_2) = \frac{n_1 n_2 (n_1 + n_2 + 1)}{12}$, (for example, see [5]). For large samples and under $H_0 : \theta_1 - \theta_2 = 0$, the critical region can be defined based on the normal distribution (again, see [5]):

$$Z_2 = \frac{T_2 - (\frac{n_1 n_2}{2})}{\sqrt{\frac{n_1 n_2 (n_1 + n_2 + 1)}{12}}} \xrightarrow{L} N(0,1). \tag{4}$$

2.3. Combined sign test with Mann-Whitney Wilcoxon test

Case 1. Small sample sizes

For small sample sizes, we propose the following test procedure to combine the sign test for correlated data with the Mann-Whitney Wilcoxon test for uncorrelated data:

1) Let $T_c = T_1 + T_2$.

2) Let $0 < \gamma < 1$, then the two sign tests can be combined as follows:

define $T_\gamma = \gamma T_1 + (1 - \gamma) T_2$.

Using similar notation as that found in [10], we construct the following theorem:

Theorem 1. Given n_1 x 's, n_2 y 's, and n pairs of (u, v) and under H_0 , let:

$$P_{n_1, n_2}(l) = P_{H_0}(T_2 = l), l = 0, 1, 2, \dots, n_1 n_2; \quad \text{and} \quad P_n(i) = P_{H_0}(T_1 = i) = \binom{n}{i} \left(\frac{1}{2}\right)^n, i = 0, 1, 2, \dots, n.$$

Then,

$$(i) \quad P(T_c = t) = \sum_{l+i=t} \sum P_{n_1, n_2}(l) P_n(i), t = 0, 1, 2, \dots, n_1 n_2 + n; \quad \text{and} \quad (5)$$

$$(ii) \quad P(T_\gamma = t) = \sum_{\gamma l + (1-\gamma)i=t} \sum P_{n_1, n_2}(l) P_n(i), \quad (6)$$

where

$$P_{n_1, n_2}(l) = \frac{n_2}{n_1 + n_2} P_{n_1, n_2 - 1}(l - n_1) + \frac{n_1}{n_1 + n_2} P_{n_1 - 1, n_2}(l), \quad P_{n_1, n_2}(l) = \frac{\bar{P}_{n_1, n_2}(l)}{\binom{n_1 + n_2}{n_1}},$$

$$\bar{P}_{n_1, n_2}(l) = \bar{P}_{n_1, n_2 - 1}(l - n_1) + \bar{P}_{n_1 - 1, n_2}(l),$$

$$\bar{P}_{j, k}(l) = 0 \text{ if } l < 0, \bar{P}_{j, 0}(l), \text{ and } \bar{P}_{0, k}(l) \text{ is } 1 \text{ or } 0 \text{ as } l = 0 \text{ or } l \neq 0, \text{ respectively.}$$

The proof is a consequence of Theorems 3.2.2 and 3.2.3 in [9] and see also [10] Theorem 2.4.3.

Examples of these null distributions for selected sample sizes are provided as tables in the Appendix. Furthermore, R codes to calculate the exact discrete distribution of the proposed tests are provided on the following website: <http://personal.georgiasouthern.edu/~hsamawi/>.

Case 2. Large sample sizes

For large sample sizes and under H_0 , and let

$$\frac{n}{n + n_1 + n_2} \rightarrow \gamma \text{ as}$$

$\{n \rightarrow \infty \text{ and large } n_1, n_2 < \infty; \text{ or } n \rightarrow \infty \text{ and } n_1 \rightarrow \infty \text{ and large } n_2 < \infty; \text{ or } n, n_1, n_2 \rightarrow \infty\}$

we propose to use:

$$(i) Z_0 = \frac{T_1 + T_2 - (\frac{n}{2} + \frac{n_1 n_2}{2})}{\sqrt{\frac{n}{4} + \frac{n_1 n_2 (n_1 + n_2 + 1)}{12}}} \xrightarrow{L} N(0, 1); \text{ and} \tag{7}$$

$$(ii) T_Z = \sqrt{\gamma} Z_1 + \sqrt{1 - \gamma} Z_2 \xrightarrow{L} N(0, 1). \tag{8}$$

3. Combined Wilcoxon Signed Rank Test and Mann-Whitney Wilcoxon Test for Correlated and Uncorrelated Data

3.1 . Wilcoxon signed rank test for correlated data

If we assume that U and V are exchangeable random variables, then $U - V$ and $V - U$ both have symmetric distributions and the Wilcoxon test is clearly justified. Let $D_i = U_i - V_i, i = 1, 2, \dots, n$.

Under H_0 , we may use a Wilcoxon signed-rank test as follows. Let I_i be an indicator for when $|D|_{(i)}$ corresponds to a positive observation, where $|D|_{(1)} < \dots < |D|_{(n)}$ are the ordered absolute values. Then,

$$T_{WC} = \sum_{i=1}^n i I_i = \sum_{i=1}^n R_i s(D_i) \tag{9}$$

is the Wilcoxon signed-rank statistic, where R_i is the rank of $|D_i|$ and $I_i = s(D_i)$, where $s(D_i) = 1$ if $D_i > 0$ and 0 otherwise. It has been shown that under the null hypothesis,

$$E(T_{WC}) = \frac{n(n+1)}{4} \text{ and } Var(T_{WC}) = \frac{n(n+1)(2n+1)}{24}.$$

Since D_i has a symmetric distribution under the H_0 assumption, T_{WC} is a linear combination of i.i.d. *Bernoulli*($1/2$) random variables.

However, for large samples

$$Z_{WC} = \frac{T_{WC} - \frac{n(n+1)}{4}}{\sqrt{\frac{n(n+1)(2n+1)}{24}}} \xrightarrow{L} N(0, 1). \tag{10}$$

Conover [5] provides an example for such large sample sizes.

3.2 . Combined Wilcoxon rank test with Mann-Whitney Wilcoxon test

Case 1. Small sample sizes

For small sample sizes, we propose the following test procedure to combine the two tests:

(i) Let $T_{cw} = T_{WC} + T_2$. (see [7])

(ii) Let $0 < \gamma < 1$, then the two sign tests can be combined as follows:

define $T_{\gamma w} = \gamma T_{WC} + (1 - \gamma)T_2$.

Again, using similar notation as that found in [9] and [10], we construct the following theorem:

Theorem 2 Given n_1 x's , n_2 y's, and n pairs of (u, v) and under H_0 , let

$$P_{n_1, n_2}(l) = P_{H_0}(T_2 = l), l = 0, 1, 2, \dots, n_1 n_2 \text{ and} \tag{11}$$

$$P_{wn}(b) = P_{H_0}(T_{WC} = b) = \frac{\bar{P}_{wn}(b)}{2^n}, b = 0, 1, 2, \dots, \frac{n(n+1)}{2}. \tag{12}$$

Then:

(i) $P(T_{cw} = t) = \sum_{l+b=t} \sum P_{n_1, n_2}(l)P_{wn}(b), t = 0, 1, 2, \dots, n_1 n_2 + \frac{n(n+1)}{2}$; and

(ii) $P(T_{\gamma w} = t) = \sum_{\gamma l + (1-\gamma)b=t} \sum P_{n_1, n_2}(l)P_{wn}(b),$

where $P_{n_1, n_2}(l)$ is the same as that in Theorem 2.1 and $\bar{P}_n(b) = \bar{P}_{n-1}(b) + \bar{P}_{n-1}(b-n),$

$\bar{P}_0(0) = 1, \bar{P}_0(b) = 0$ and $\bar{P}_n(b) = 0$ for $b < 0$.

The proof is a result of Theorems 3.2.2 and 3.2.3 as well as Exercise 3.7.3 in [9]. Similarly, examples of these null distributions for selected sample sizes are provided in the Appendix. Additionally, R codes are provided on the following website to calculate the exact discrete distribution of the proposed tests: <http://personal.georgiasouthern.edu/~hsamawi/>.

Case 2. Large sample sizes

For large samples and under H_0 , and let

$$\frac{n}{n + n_1 + n_2} \rightarrow \gamma \text{ as}$$

$\{n \rightarrow \infty \text{ and large } n_1, n_2 < \infty; \text{ or } n \rightarrow \infty \text{ and } n_1 \rightarrow \infty \text{ and large } n_2 < \infty; \text{ or } n, n_1, n_2 \rightarrow \infty\}$

, we can combine the two tests as follows:

$$(i) Z_{0D} = \frac{T_{WC} + T_2 - \left(\frac{n(n+1)}{4} + \frac{n_1 n_2}{2}\right)}{\sqrt{\frac{n(n+1)(2n+1)}{24} + \frac{n_1 n_2 (n_1 + n_2 + 1)}{12}}} \xrightarrow{L} N(0,1); \text{ and}$$

$$(ii) T_Z = \sqrt{\gamma} Z_{WC} + \sqrt{1-\gamma} Z_2 \xrightarrow{L} N(0,1).$$

4. New Test Procedure

Under the assumption of MCAR and in order to test the null hypothesis $H_0 : \theta_1 = \theta_2$, the following notation are used: $D_i = U_i - V_i, i = 1, 2, \dots, n$, and $D_{jk} = X_j - Y_k, j = 1, 2, \dots, n_1, k = 1, 2, \dots, n_2$. Let $N = n + n_1 n_2$; $DD_1 = D_1, \dots, D_n = DD_n; DD_{n+1} = D_{11}, \dots, D_{n_1 n_2} = DD_N$; and I_m be an indicator when $|DD|_{(m)}$ corresponds to a positive DD observation, where $m = 1, 2, \dots, N$ and $|DD|_{(1)} < \dots < |DD|_{(N)}$ are the ordered absolute values. Then,

$$T_{New} = \sum_{m=1}^N m I_m = \sum_{m=1}^N R_m s(DD_m) \tag{13}$$

is the new signed-rank statistic, where R_m is the rank of $|DD_m|$ and $I_m = s(DD_m)$ is defined above. As in Hettmansperger [9] Theorem 2.2.1 and Hettmansperger and McKean [10] Lemma 1.7.1, under the assumption that $\{s(DD_1), \dots, s(DD_N)\}$ and $\{R_1, \dots, R_N\}$ are independent we will provide the following results. Using simple probability algebra, we can show that

$$P(s(DD_m) = 1) = P(s(DD_m) = 1, DD_m \in A_1) + P(s(DD_m) = 1, DD_m \in A_2)$$

$$= \frac{1/2}{\binom{N}{n}} + \frac{1/2}{\binom{N}{n_1 n_2}} = \frac{1}{\binom{N}{n}},$$

where $A_1 = \{DD_1, \dots, DD_n\}$ and $A_2 = \{DD_{n+1}, \dots, DD_N\}$. The number of permutations of the sequence of N positive and negative signs of DD 's is 2^N . Using the convolution of the distribution of Wilcoxon's signed rank statistic in the Wilcoxon-Mann-Whitney statistic, the null

distribution of T_{new} can be written as

$$P(T_{new} = t) = \frac{1}{\binom{N}{n}} \sum_{m=1}^{2^N} P_m(T_{WC} = t_{WC}) P_m(T_2 = t - t_{WC}), \quad t = 0, 1, 2, \dots, \frac{N(N+1)}{2}, \quad (14)$$

where $P_m(T_{WC} = t_{WC})$ and $P_m(T_2 = t - t_{WC})$ are expressed in Theorem 2. Eq.(12) for the m^{th}

permutation as
$$P_m(b) = P_m(T_{WC} = b) = \frac{\bar{P}_m(b)}{2^m}, \quad b = 0, 1, 2, \dots, \frac{m(m+1)}{2}$$
 and

$$\bar{P}_m(b) = \bar{P}_{m-1}(b) + \bar{P}_{m-1}(b-m), \quad \bar{P}_0(0) = 1, \bar{P}_0(b) = 0 \text{ and } \bar{P}_m(b) = 0 \text{ for } b < 0.$$

R codes to find (14) by calculating the exact discrete distribution of the proposed test are provided on the following website: <http://personal.georgiasouthern.edu/~hsamawi/>. Using Eq. (13), it is easy to show that the mean and the variance of our proposed test statistic are as follows:

$$E(T_{New}) = \frac{N(N+1)}{2 \cdot \binom{N}{2}}, \quad (15)$$

$$V(T_{New}) = \frac{4N(N+1)(2N+1) \binom{N}{2} - 6N^2(N+1)^2}{24 \binom{N}{2}^2}. \quad (16)$$

Note that both the mean and the variance in Eq. (15) and Eq. (16) are finite and decreasing as N increases, provided that

$$\frac{n}{n+n_1+n_2} \rightarrow \gamma \text{ as}$$

$\{n \rightarrow \infty \text{ and large } n_1, n_2 < \infty; \text{ or } n \rightarrow \infty \text{ and } n_1 \rightarrow \infty \text{ and large } n_2 < \infty; \text{ or } n, n_1, n_2 \rightarrow \infty\}$

Under the MCAR design and the null hypothesis, the asymptotic distribution of T_{New} is as

follows:
$$Z_{New} = \frac{T_{New} - E(T_{New})}{\sqrt{V(T_{New})}} \xrightarrow{L} N(0,1). \quad (17)$$

5. Numerical Analysis and Final Remarks

Simulations were conducted to compare the power of the proposed tests. In order to compare the proposed tests with other tests found in the literature, (those methods are further described in [11] and [21] as $S_1, S_2, S_3, S_4,$ and S_5) sample sizes $n=3, n_1=2$ and $n_2=3$ and different shift

parameters $d=0.0, 0.5, 1.0, 1.5, 2.0, 2.5$ and 3 were considered. Underlying bivariate normal distributions with different values of the correlation coefficient (r), ranging from -0.9 to 0.7 , were considered. The power estimation was based on $10,000$ iterations.

Table 1 shows that all proposed tests for small sample sizes have a simulated probability of Type I errors very close to the nominal value of 0.05 except for test statistics S_2 and S_5 . Also, note that S_1 is equivalent to T_c (i.e., the combined sign test with Mann-Whitney Wilcoxon test). Therefore, S_1 is excluded from any further comparisons. Moreover, depending on the strength and direction of the correlation between the paired variables, test statistics S_2 and S_5 appear not to be invariant. Table 2 shows that our proposed tests are competing with the other test in most of the cases. However, due to the variation in the estimated nominal values in table 1, ROC curve of the power of the presented tests is used to compare between the test's curves average areas. The ROC curves show that all our proposed tests have areas above 60% . The test statistics S_2, S_4 and S_5 have areas below 60% . Test S_3 is competing with our tests with area of ROC curve above 60% . Figures 1-6 show that, test S_1 has the largest average curve area over the given correlation coefficient values. Therefore, our proposed tests and S_3 test are recommended.

Table 1. Estimated probability of a Type I error using randomized tests for small sample sizes with a 0.05 level of significance. (Design 3-2-3)

r	T_c	T_γ	$T_{\gamma w}$	T_{cw}	S_1	S_2	S_3	S_4	S_5	T_{New}
-0.900	0.052	0.065	0.061	0.050	0.052	0.126	0.049	0.070	0.050	0.058
-0.500	0.049	0.062	0.063	0.051	0.041	0.110	0.050	0.073	0.040	0.063
-0.100	0.046	0.058	0.057	0.045	0.046	0.088	0.046	0.065	0.026	0.058
0.300	0.048	0.059	0.058	0.048	0.048	0.072	0.049	0.068	0.017	0.068
0.700	0.050	0.062	0.064	0.052	0.050	0.050	0.052	0.073	0.011	0.057

Table 2. Power comparison for testing the difference in medians (H_0) for all methods using a correlation coefficient of r (Design 3-2-3)

D	r	T_c	T_γ	$T_{\gamma w}$	T_{cw}	S_1	S_2	S_3	S_4	S_5	T_{New}
0.500	-0.900	0.145	0.178	0.151	0.181	0.145	0.296	0.149	0.196	0.164	0.178
1.000	-0.900	0.291	0.347	0.312	0.366	0.291	0.514	0.309	0.382	0.345	0.376
1.500	-0.900	0.486	0.563	0.521	0.589	0.486	0.727	0.515	0.600	0.575	0.613
2.000	-0.900	0.680	0.761	0.726	0.781	0.681	0.875	0.720	0.789	0.787	0.816
2.500	-0.900	0.823	0.894	0.867	0.906	0.823	0.951	0.860	0.906	0.912	0.933
3.000	-0.900	0.906	0.957	.943	0.962	0.906	0.981	0.939	0.962	0.971	0.980
0.500	-0.500	0.144	0.175	0.152	0.185	0.143	0.281	0.152	0.200	0.136	0.190
1.000	-0.500	0.310	0.368	0.336	0.388	0.310	0.511	0.335	0.410	0.323	0.400
1.500	-0.500	0.520	0.594	0.570	0.627	0.520	0.751	0.568	0.648	0.579	0.655
2.000	-0.500	0.708	0.786	0.769	0.809	0.708	0.903	0.766	0.827	0.796	0.849
2.500	-0.500	0.850	0.912	0.901	0.926	0.850	0.967	0.899	0.933	0.930	0.951
3.000	-0.500	0.926	0.968	0.961	0.973	0.926	0.991	0.960	0.975	0.978	0.986
0.500	0.300	0.163	0.198	0.183	0.215	0.163	0.243	0.184	0.237	0.097	.238
1.000	0.300	0.372	0.433	0.431	0.468	0.372	0.544	0.435	0.515	0.309	0.504
1.500	0.300	0.600	0.670	0.682	0.699	0.600	0.809	0.689	0.760	0.599	0.757
2.000	0.300	0.797	0.854	0.867	0.867	0.796	0.948	0.872	0.910	0.839	0.906
2.500	0.300	0.907	0.945	0.953	0.949	0.907	0.988	0.956	0.971	0.951	0.974
3.000	0.300	0.961	0.980	0.984	0.981	0.961	0.999	0.985	0.991	0.990	0.994
0.500	0.700	0.199	0.239	0.236	0.262	0.199	0.248	0.240	0.303	0.077	0.285
1.000	0.700	0.440	0.493	0.528	0.528	0.437	0.595	0.541	0.617	0.297	0.570
1.500	0.700	0.668	0.725	0.763	0.738	0.668	0.861	0.776	0.832	0.608	0.799
2.000	0.700	0.829	0.873	0.895	0.875	0.829	0.966	0.902	0.932	0.851	0.924
2.500	0.700	0.919	0.946	0.957	0.946	0.919	0.992	0.960	0.975	0.958	0.976
3.000	0.700	0.965	0.980	0.985	0.980	0.965	0.999	0.986	0.992	0.993	0.995

Figure 1. ROC curve for the empirical power of T_c test versus the empirical nominal value of the test

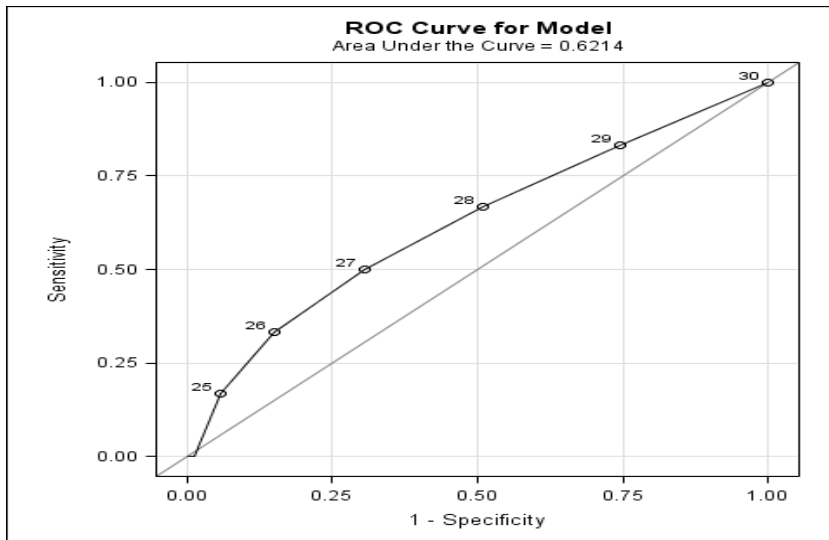


Figure 2. ROC curve for the empirical power of T_{cv} test versus the empirical nominal value of the test

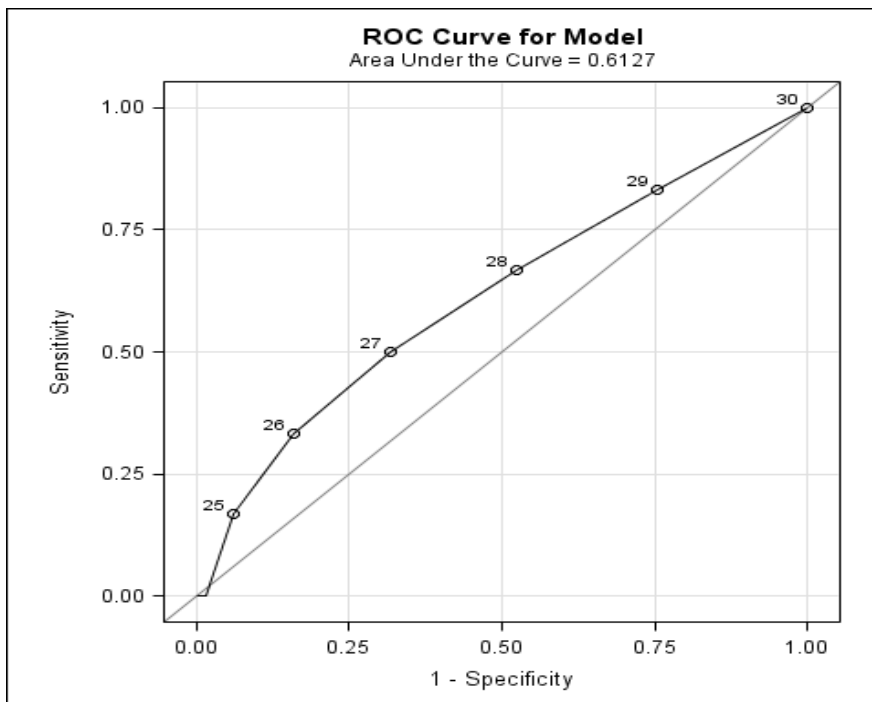


Figure 3. ROC curve for the empirical power of T_γ test versus the empirical nominal value of the test

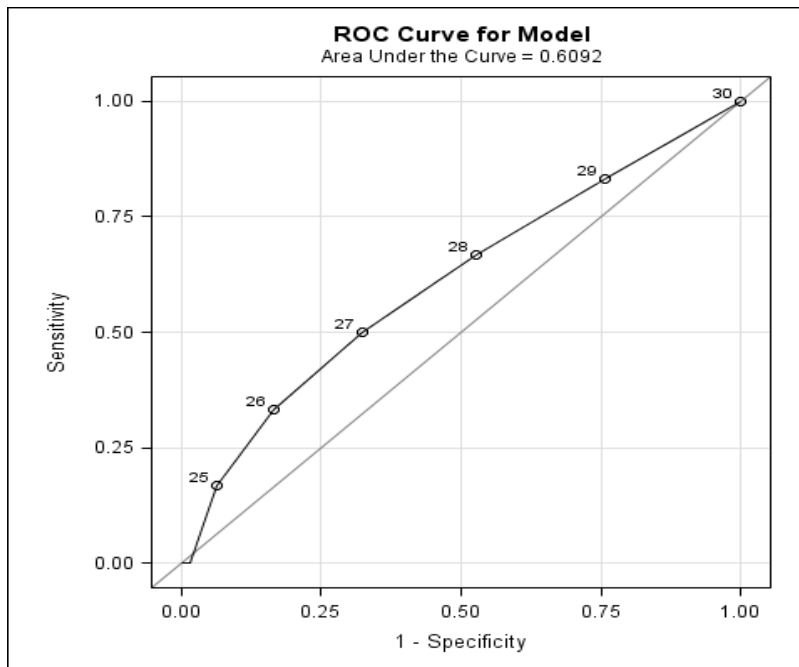


Figure 4. ROC curve for the empirical power of T_{New} test versus the empirical nominal value of the test

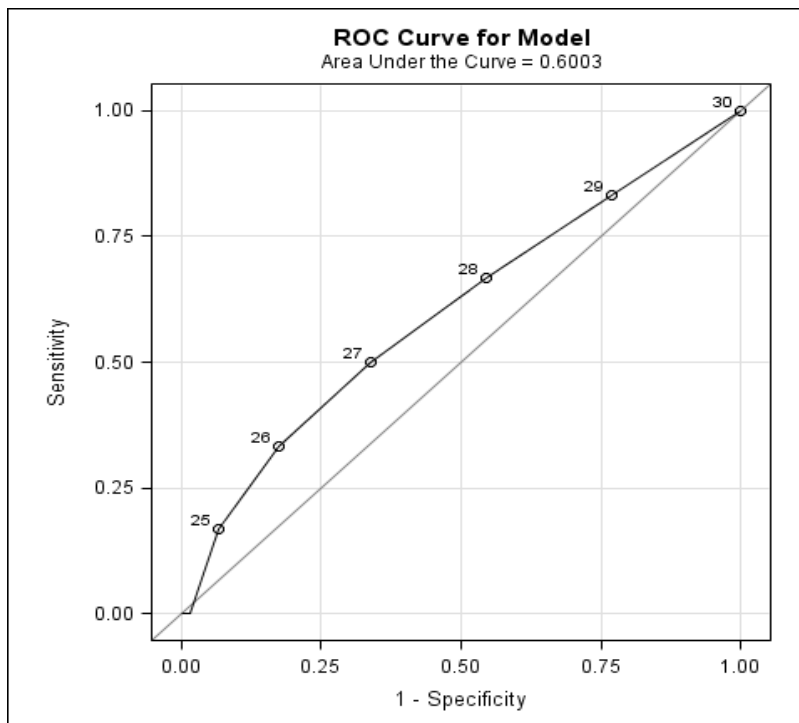


Figure 5. ROC curve for the empirical power of S_2 test versus the empirical nominal value of the test

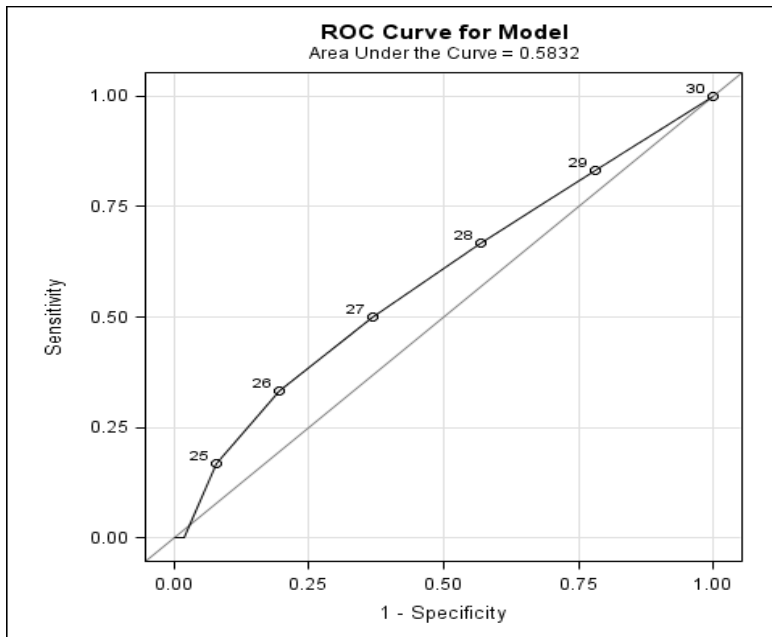
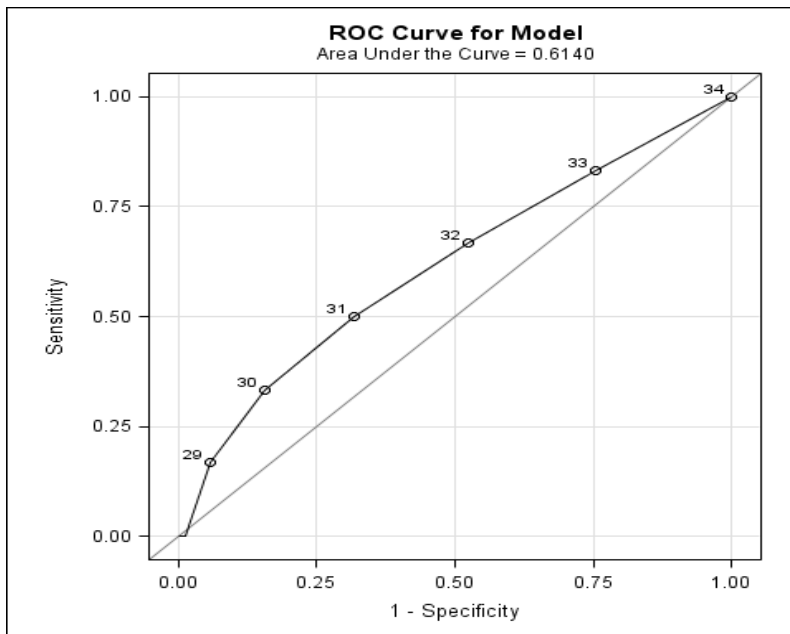


Figure 6. ROC curve for the empirical power of S_3 test versus the empirical nominal value of the test



6. Illustration Using Genetic Data

Table 3 contains data on eight patients taken from Weidmann et al. [22]. The purpose of this study was to compare the proportions of certain T cell receptor gene families (the $V\beta$ gene families) on tumor infiltrating lymphocytes (TILs) and peripheral blood lymphocytes (PBLs) in patients with hepatocellular carcinoma (i.e., matched pairs or block design). Weidmann et al [22] expected that more changes would be detected in surface receptors of T lymphocytes in the presence of a tumor. The outcome represented the percentage of T cells of each type showing the $V\beta 22$ receptor. The null hypothesis of equal medians was tested by estimating and comparing the relative proportions of $V\beta$ gene family usage for several patients' TILs and PBLs. However, data are missing for some patients due to factors unrelated to the measurements themselves.

Table 3. T cell receptor data ([22]).

<i>VB%</i>		
Patient	TIL	PBL
1	6.7	2.8
2	3.7	3.5
3	4.4	4.1
4	2.3	.
5	4.5	.
6	.	4.0
7	.	14.7
8	.	3.2

Table 4. Results of all suggested tests.

Test Procedure	Value of the test	P-value
S_1	11.000	0.5000
S_2	19.000	0.3000
S_3	20.000	0.2103
S_4	5.000	0.1625
S_5	33.000	0.3500
Sign test for matched pairs T_1 (ignoring unmatched data)	3.000	0.1250
Wilcoxon test for matched pairs T_{wc} (ignoring unmatched data)	6.000	0.1250
Mann-Whitney test for two independent samples T_2 (ignoring matched data)	2.000	0.8000
Combined Sign with Mann-Whitney tests T_c	5.000	0.5000
Weighted Sign with Mann-Whitney tests T_γ , $\gamma = \frac{n}{n + n_1 + n_2}$	2.375	0.4000
Combined Wilcoxon with Mann-Whitney tests T_{cw}	8.000	0.2878
Weighted Wilcoxon with Mann-Whitney tests $T_{\gamma w}$	3.500	0.3875
Optimal Weighted Sign with Mann-Whitney tests $\gamma = \frac{w_i}{\sum_{i=1}^2 w_i}, w_i = \frac{1}{\sigma_i}$, where σ_i is the standard deviation of the i th test.	2.800	0.1500
Optimal Weighted Wilcoxon with Mann-Whitney tests	3.840	0.2875
New proposed test (T_{New})	18.000	0.6350

None of the tests provided in Table 4 indicate any statistical significance. Therefore, regardless of the test used, H_0 is not rejected.

Acknowledgment

The Authors would like to thank the reviewers and the associate editor for their valuable comment which improved the manuscript.

References

- [1] Akritas, M.G., Kuha, J., and Osgood, D.W. , A nonparametric approach to matched pairs with missing data. *Sociological Methods & Research*, **30** (2002), 425-457.
- [2] Brunner, E. and Neumann, N., Two-sample rank tests in general models. *Biometrical Journal*, **28**, (1986) 394-402.
- [3] Brunner, E. and Puri, M.LN., *Nonparametric methods in design and analysis of experiments*. (In S. Ghosh and C.R. Rao (Eds.), *Handbook of Statistics 13* (pp. 631-703). Amsterdam: North-Holland/Elsevier, 1996) .
- [4] Brunner, E., Domhof, S., and Langer, F. ,*Nonparametric analysis of longitudinal data In factorial designs*. (New York: John Wiley & Sons, Inc.,2002).
- [5] Conover, W. J. , *Practical nonparametric statistics* (Third ed.) (New York: John Wiley & Sons, Inc. , 1999).
- [6] Dimery, I.W., Nishioka, K., Grossie, B., Ota, D.M., Schantz, S.P., Robbins, K.T., and Hong, W.K. , Polyamine metabolism in carcinoma of oral cavity compared with adjacent and normal oral mucosa. *American Journal of Surgery*, **154**, (1987) 429-433.
- [7] Dubnicka, S.R, Blair , R.C. and. Hettmansperger, T.P , Rank-based procedures for mixed paired and two-sample designs. *Journal of Modern Applied Statistical Methods*, 1 (1), (2002) 32-41.
- [8] Ekbohm, G. Comparing means in the paired case with missing data on one response. *Biometrika*, **63** (1), (1976) 169-172.
- [9] Hettmansperger, T.P. , *Statistical inference based on ranks*.(New York: John Wiley & Sons, Inc., 1984).
- [10] Hettmansperger, T. P. and McKean, J. W. ,*Robust nonparametric statistical method*. (2nd ed.) (New York: CRC Press, Taylor & Francis Group, 2011).
- [11] Im KyungAh., *A modified signed rank test to account for missing in small samples*

- With paired data.* (M.S. Thesis, Department of Biostatistics, Graduate School of Public Health, University of Pittsburgh, Pittsburgh, PA. <http://www.worldcat.org/title/modified-signed-rank-test-to-account-for-missing-data-in-small-samples-with-paired-data/oclc/52418573>., 2002).
- [12] Lehmann, E.L., *Nonparametrics: Statistical methods based on ranks.* (San Francisco: Holden-Day., 1975).
- [13] Lin, P. and Stivers, L.E., On difference of means with incomplete data. *Biometrika*, **61**(2) (1974) 325-334.
- [14] Little, R.J. , A test of missing completely at random for multivariate data with Missing values. *Journal of the American Statistical Association*, **83**, (1988)1198-1202.
- [15] Looney, S.W. and Jones, P.W. , A method for comparing two normal means using combined samples of correlated and uncorrelated data. *Statistics In Medicine*, **22**, (2003)1601-1610.
- [16] Nurnberger, J., Jimerson, D., Allen, J.R., Simmons, S., and Gershon, E., Red cell ouabain-sensitive Na⁺-K⁺-adenosine triphosphatase: a state marker in affective disorder inversely related to plasma cortisol. *Biological Psychiatry*, **17** (9), (1982) 981-992.
- [17] Rempala, G. and Looney, S., Asymptotic properties of a two-sample randomized test For partially dependent data. *Journal of Statistical Planning and Inference*, **136**, (2006) 68-89.
- [18] Sen, P.K., Asymptotic efficiency of a class of aligned rank order tests for Multiresponse experiments in some incomplete block designs. *Annals of Mathematical Statistics*, **42**, (1971)1104–1112.
- [19] Snedecor, G.W. and Cochran, W.G., *Statistical Methods* (7th ed.). Ames, IA: Iowa State University Press, 1980).
- [20] Steere, A.C., Green, J., Schoen, R.T., Taylor, E., Hutchinson, G.J., Rahn, D.W., and Malawista, S.E., Successful parenteral penicillin therapy of established Lyme arthritis. *New England Journal of Medicine*, **312** (14), (1985) 8699-874.
- [21] Tang, X., *New test statistic for comparing medians with incomplete paired data.* (M.S. Thesis, Department of Biostatistics, Graduate School of Public Health, University of Pittsburgh, Pittsburgh, PA, http://www.google.com/search?hl=en&rlz=1T4ADRA_enUS357US357&q=Tang+X.+%

[282007%29New+Test+Statistic+for+Comparing+Medians+with+Incomplete+Paired+Data&btnG=Search&aq=f&aqi=&aql=&oq=2007\)..](#)

[22] Weidmann, E., Whiteside, T.L., Giorda, R., Herberman, R.B., and Trucco, M., The T-cell receptor V beta gene usage in tumor-infiltrating lymphocytes and blood of patients with hypatocellular carcinoma. *Cancer Research*, **52** (21), (1992) 5913-5920.

Appendix

Table 7. Null distribution to T_γ , T_c , $T_{\gamma w}$, and T_{cw} for $n = 2, n_1 = 2, n_2 = 2$ and $\gamma = 0.33$.

t_c	$P(T_c = t_c)$	t_γ	$P(T_\gamma = t_\gamma)$	t_{cw}	$P(T_{cw} = t_{cw})$	$t_{\gamma w}$	$P(T_{\gamma w} = t_{\gamma w})$
0	0.041667	0.00	0.041667	0	0.041667	0.00	0.0416665
1	0.125000	0.33	0.083333	1	0.083333	0.33	0.0416665
2	0.208332	0.66	0.041667	2	0.166666	0.66	0.0416665
3	0.249998	0.67	0.041667	3	0.208332	0.67	0.0416665
4	0.208332	1.00	0.083333	4	0.208332	0.99	0.0416665
5	0.125000	1.33	0.041667	5	0.166665	1.00	0.0416665
6	0.041667	1.34	0.083333	6	0.083333	1.33	0.0416665
		1.67	0.166665	7	0.041667	1.34	0.0833325
		2.00	0.083333			1.66	0.0416665
		2.01	0.041667			1.67	0.0833325
		2.34	0.083333			2.00	0.0833325
		2.67	0.041667			2.01	0.0416665
		2.68	0.041667			2.33	0.0833325
		3.01	0.083333			2.34	0.0416665
		3.34	0.041667			2.67	0.0416665
						2.68	0.0416665
						3.00	0.0416665
						3.01	0.0416665
						3.34	0.0416665
						3.67	0.0416665

Table 8. Null distribution to T_γ , T_c , $T_{\gamma w}$, and T_{cw} for $n = 3, n_1 = 3, n_2 = 2$ and $\gamma = 0.375$.

t_c	$P(T_c = t_c)$	t_γ	$P(T_\gamma = t_\gamma)$	$t_{\gamma w}$	$P(T_{\gamma w} = t_{\gamma w})$	t_{cw}	$P(T_{cw} = t_{cw})$
0	0.0125	0.0000	0.01250	0	0.0125	0.0000	0.0125
1	0.0500	0.3750	0.03750	1	0.0250	0.3750	0.0125
2	0.1000	0.6250	0.01250	2	0.0500	0.6250	0.0125
3	0.1500	0.7500	0.03750	3	0.0875	0.7500	0.0125
4	0.1875	1.0000	0.03750	4	0.1125	1.0000	0.0125
5	0.1875	1.1250	0.01250	5	0.1375	1.1250	0.0250
6	0.1500	1.2500	0.0250	6	0.1500	1.2500	0.0250
7	0.1000	1.3750	0.03750	7	0.1375	1.3750	0.0125
8	0.0500	1.6250	0.0750	8	0.1125	1.5000	0.0125
9	0.0125	1.7500	0.01250	9	0.0875	1.6250	0.0250
		1.8750	0.0250	10	0.0500	1.7500	0.0250
		2.0000	0.0750	11	0.0250	1.8750	0.0375
		2.2500	0.0750	12	0.0125	2.0000	0.0250
		2.3750	0.0250			2.1250	0.0125
		2.5000	0.0250			2.2500	0.0375
		2.6250	0.0750			2.3750	0.0500
		2.8750	0.0750			2.5000	0.0375
		3.0000	0.0250			2.6250	0.0250
		3.1250	0.0125			2.7500	0.0250
		3.2500	0.0750			2.8750	0.0375
		3.5000	0.0375			3.0000	0.0500
		3.6250	0.0250			3.1250	0.0375
		3.7500	0.0125			3.2500	0.0250
		3.8750	0.0375			3.3750	0.0250
		4.1250	0.0375			3.5000	0.0375
		4.2500	0.0125			3.6250	0.0500
		4.5000	0.0375			3.7500	0.0375
		4.8750	0.0125			3.8750	0.0125
						4.0000	0.0250

						4.1250	0.0375
						4.2500	0.0250
						4.3750	0.0250
						4.5000	0.0125
						4.6250	0.0125
						4.7500	0.0250
						4.8750	0.0250
						5.0000	0.0125
						5.2500	0.0125
						5.3750	0.0125
						5.6250	0.0125
						6.0000	0.0125