

**NEW TEST STATISTIC FOR COMPARING MEDIANS WITH INCOMPLETE
PAIRED DATA**

by

Xinyu Tang

B.A., East China University of Science and Technology, 2003

Submitted to the Graduate Faculty of
Graduate School of Public Health in partial fulfillment
of the requirements for the degree of
Master of Science

University of Pittsburgh

2007

UNIVERSITY OF PITTSBURGH
GRADUATE SCHOOL OF PUBLIC HEALTH

This thesis was presented

by

Xinyu Tang

It was defended on

April 9, 2007

and approved by

Jong-Hyeon Jeong, PhD., Assistant Professor, Department of Biostatistics,
Graduate School of Public Health, University of Pittsburgh

Maria M. Brooks, PhD., Associate Professor, Department of Epidemiology,
Graduate School of Public Health, University of Pittsburgh

Thesis Director: John W. Wilson, PhD., Associate Professor, Department of Biostatistics,
Graduate School of Public Health, University of Pittsburgh

**NEW TEST STATISTIC FOR COMPARING MEDIANS WITH INCOMPLETE
PAIRED DATA**

Xinyu Tang, M.S.

University of Pittsburgh, 2007

This paper is concerned with nonparametric methods for comparing medians of paired data with unpaired values on both responses. A new nonparametric test statistic is proposed in this paper based on a Mann-Whitney U test making comparisons across complete and incomplete pairs. A method of finding the null hypothesis distribution for this statistic is presented using a permutation approach. A Monte Carlo simulation study is described to make power comparisons among four already-existing nonparametric test statistics and this new test statistic. It is concluded that this new test statistic is fairly powerful in handling this kind of data compared to the other four test statistics. Finally, all five test statistics are applied to a real dataset for comparing the proportions of certain T cell receptor gene families in a cancer study. The introduction of this new nonparametric test statistic is of public health importance because it is a powerful statistical method for dealing with a pattern of missing data that may be encountered in clinical and public health research.

TABLE OF CONTENTS

1.0	INTRODUCTION.....	1
1.1	INCOMPLETE PAIRED DATA CONFIGURATION AND NULL HYPOTHESIS	2
1.2	PREVIOUS NONPARAMETRIC TEST STATISTIC ($T_1 - T_4$)	5
1.2.1	Test statistic: T_1 proposed by J. Wilson.....	5
1.2.2	Test statistic: T_2 proposed by Brunner & Neumann (1984)	6
1.2.3	Test statistic: T_3 proposed by P.K. Sen.....	6
1.2.4	Test statistic: T_4 proposed by KyungAh Im.....	7
1.3	NEW TEST STATISTIC: T_5 PROPOSED BY JOHN BRYANT	8
1.4	NULL HYPOTHESIS DISTRIBUTIONS FOR $T_1 - T_5$	9
2.0	APPLICATION.....	13
3.0	MONTE CARLO SIMULATION STUDY	25
4.0	CONCLUSIONS AND DISCUSSION	33
	APPENDIX A NULL HYPOTHESIS DISTRIBUTION FOR T1.....	36
	APPENDIX B NULL HYPOTHESIS DISTRIBUTION FOR T3.....	37
	APPENDIX C NULL HYPOTHESIS DISTRIBUTION FOR T4	38
	BIBLIOGRAPHY	39

LIST OF TABLES

Table 1. Data and application of test statistic T_1	14
Table 2. Null hypothesis distribution of test statistic T_1	15
Table 3. Data and application of test statistic T_2	16
Table 4. Null hypothesis distribution of test statistic T_2	17
Table 5. Data and application of test statistic T_3	18
Table 6. Null hypothesis distribution of test statistic T_3	19
Table 7. Data and application of test statistic T_4	20
Table 8. Null hypothesis distribution of test statistic T_4	21
Table 9. Data and application of test statistic T_5	22
Table 10. Null hypothesis distribution of test statistic T_5	23
Table 11. All combinations of parameters for datasets generated from a bivariate normal distribution.	28
Table 12. Power comparisons for five test statistics under various conditions with $\sigma^2=1$ and $\alpha=0.05$	29
Table 13. Power comparisons for five test statistics under various conditions with $\sigma^2=1$ and $\alpha=0.025$	30

Table 14. Power comparisons for five test statistics under various conditions with $\sigma^2=1$ and $\alpha=0.0125$ 31

Table 15. 95% confidence intervals for a sample size of 5000 simulations with regard to varied Type I errors and test powers..... 32

LIST OF FIGURES

Figure 1. Incomplete paired data with configuration J-K-L.	4
---	---

1.0 INTRODUCTION

The way to compare population location parameters using complete paired data, as in before-after, repeated measures, or matched pair studies, has been widely known for a long period of time. Differences may be calculated within each pair and the single sample of differences is examined. If the sample is normally distributed, normal theory applies and the difference in sample means will be within approximately two standard errors of the difference in population means 95% of the time.

However, in real studies, researchers occasionally must compare location parameters in the paired case with unpaired data on one or both responses. This problem arises in many different applications, e.g. in vivisectional experiments when some animals die before all observations are obtained, in public health controls when some people only take part at one time and in other cases when observations are lost or do not become available. An important assumption made in treating this problem is that missing observations are missing completely at random (MCAR).

This paper covers the topic of comparing location parameters in the paired case with unpaired data on both responses. So far, several authors have presented various tests considering the problem of estimating the difference of means of a bivariate normal distribution where some observations corresponding to both variables are missing. Ekbohm (1976) summarized five procedures for testing the equality of two correlated means with incomplete data on both

responses and compared them using Monte Carlo studies. The results are that the two tests based on a modified maximum likelihood estimator are to be preferred, one due to Lin & Stivers (1974) when the number of complete pairs is large and the one proposed in Ekbohm's paper otherwise, provided the variances of the two responses do not differ by much. When the correlation between the two responses is small, two other tests may be used; a test proposed in Ekbohm's paper when the homoscedasticity assumption is not strongly violated, and a Welch type statistic suggested by Lin & Stivers (1974) otherwise (Ekbohm 1976).

All five procedures mentioned by Ekbohm are presented under the assumption of bivariate normality and MCAR. However, we may need to deal with paired data that are not bivariate normally distributed; some data have a small sample size where normality can't be assessed. In those situations, we turn to use nonparametric tests to compare medians for incomplete paired data. In this paper, five nonparametric tests that compare medians in paired case with unpaired data on both responses are examined. Among them, four tests were proposed previously, and one is proposed herein. After introducing the five nonparametric test statistics and their respective null hypothesis distributions, a Monte Carlo study of the powers and level of significance is conducted. Comparisons among these tests are made for different combinations of the correlation coefficient ρ and differences of means. For the purpose of this paper, we will assume that the variances in the compared populations are equal.

1.1 INCOMPLETE PAIRED DATA CONFIGURATION AND NULL HYPOTHESIS

Let Y_1 and Y_2 be jointly distributed according to a continuous bivariate distribution whose marginal distributions have the same shape. We consider the situation in which J paired

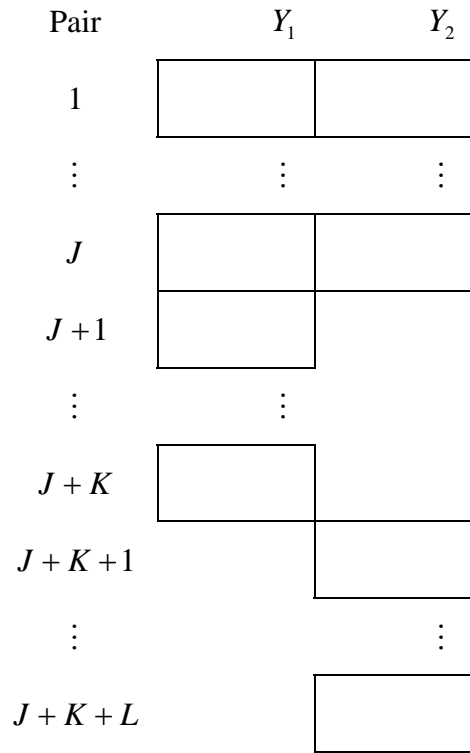
observations $(y_{1\alpha}, y_{2\alpha})'$, $\alpha = 1, \dots, j$ are made on $(Y_1, Y_2)'$, and in addition K unpaired observations are available on Y_1 alone and L unpaired observations are available on Y_2 alone (Figure 1 on Page 5). We denote this kind of incomplete paired data configuration as $J - K - L$, where there are J complete pairs, K incomplete pairs with missing observations in Y_2 only, and L incomplete pairs with missing observations in Y_1 only.

The following one-sided hypotheses are being tested throughout the analyses assuming that the probability of missing an observation is independent of the observed responses (i.e., MCAR: Little 1988), populations studied are continuous and have the same shape under the null hypothesis.

$$H_0 : M_1 = M_2$$

$$H_a : M_1 > M_2$$

Where M_1 and M_2 are the medians from each population.



$$Y_1 : y_{1,1}, y_{1,2}, \dots, y_{1,j}, y_{1,j+1} \dots y_{1,j+k}$$

$$Y_2 : y_{2,1}, y_{2,2}, \dots, y_{2,j}, \quad y_{2,j+k+1}, y_{2,j+k+2}, \dots, y_{2,j+k+l}$$

$y_{m,n}$ = observation from mth group and nth pair

Figure 1. Incomplete paired data with configuration J-K-L.

1.2 PREVIOUS NONPARAMETRIC TEST STATISTIC ($T_1 - T_4$)

Most notations follow those in KyungAh Im's paper (2002).

1.2.1 Test statistic: T_1 proposed by J. Wilson

Test statistic T_1 is based on a sign test for the complete pairs and a Wilcoxon rank sum test for the unpaired cases.

For the complete pairs, rank two observed values within each pair and sum the ranks of one group. In this paper, ranks of Y_1 group are used throughout the analyses. Denote

$$R_{n1} = \begin{cases} 1 & \text{if } y_{1,n} < y_{2,n}, \quad n = 1, \dots, j \\ 2 & \text{if } y_{1,n} > y_{2,n}, \quad n = 1, \dots, j \end{cases}$$

Then the test statistic for the complete pairs is $T_{c1} = \sum_{n=1}^j R_{n1}$.

For the unpaired cases, rank among total number of $k + l$ unpaired observations and sum the ranks for the same group, Y_1 group here.

Combined sample with unpaired data in both Y_1 and Y_2 :

$$y_{1,j+1}, \dots, y_{1,j+k}, y_{2,j+k+1}, \dots, y_{2,j+k+l}$$

Denote r_n be the rank of each observation in Y_1 group in the combined sample, then the

test statistic for the unpaired data is $T_i = \sum_{n=j+1}^{j+k} r_n$.

Hence the test statistic T_1 is defined as $T_1 = T_{c1} + T_i$.

1.2.2 Test statistic: T_2 proposed by Brunner & Neumann (1984)

Test statistic T_2 is based on a Wilcoxon rank sum test for complete pairs and unpaired cases respectively.

For the complete pairs, combine all the paired data from Y_1 and Y_2 :

$$y_{1,1}, \dots, y_{1,j}, y_{2,1}, \dots, y_{2,j}$$

Denote R_{n2} be the rank of each observation in Y_1 group in the combined paired sample,

then the test statistic for the paired data is $T_{c2} = \sum_{n=1}^j R_{n2}$. This rank sum test statistic for the

complete pairs doesn't take pairing into consideration.

Hence the test statistic T_2 is defined as $T_2 = T_{c2} + T_i$, where T_i is defined as for the test statistic T_1 .

1.2.3 Test statistic: T_3 proposed by P.K. Sen

Test statistic T_3 is based on an aligned rank sum test (Lehmann 1975) for the complete pairs and a Wilcoxon rank sum test for the unpaired cases.

For the complete pairs, subtract the mean of each pair from the original observed values before ranking. Denote $\hat{y}_{m,n} = y_{m,n} - \mu_n$, where μ_n is the mean of the two observed values in pair n.

Combine all the modified paired data from Y_1 and Y_2 as following:

$$\hat{y}_{1,1}, \dots, \hat{y}_{1,j}, \hat{y}_{2,1}, \dots, \hat{y}_{2,j}$$

Denote R_{n3} be the rank of each observation in Y_1 group in the modified combined paired

sample, then the test statistic for the complete pairs is $T_{c3} = \sum_{n=1}^j R_{n3}$.

Hence the test statistic T_3 is defined as $T_3 = T_{c3} + T_i$, where T_i is defined as for the test statistic T_1 .

1.2.4 Test statistic: T_4 proposed by KyungAh Im

For this test statistic, in addition to the assumptions mentioned for the previous three tests, the two populations being compared are assumed to be symmetrically distributed about their respective medians (Im 2002).

For complete pairs, calculate the difference d_n within each pair, i.e. $d_n = y_{1,n} - y_{2,n}$

$$\text{Let } s(d_n) = \begin{cases} 1 & \text{if } d_n > 0, \quad n = 1, \dots, j \\ 0 & \text{if } d_n < 0, \quad n = 1, \dots, j \end{cases}$$

Rank the absolute value of the difference, denote $r(|d_n|)$

Then the test statistic for the complete pairs is

$$T_{c4} = \left\{ \sum_{n=1}^j r(|d_n|) * s(d_n) - \sum_{n=1}^j r(|d_n|) * [1 - s(d_n)] \right\}$$

Where $s(d_n)$ are independent Bernoulli random variables with expected value of 0.5 under the null hypothesis. Then T_{c4} is a version of the Wilcoxon Signed Rank Test (Gibbons

1971). The partial test statistic T_{c4} has the property of natural symmetry around the value zero in its distribution under the null hypothesis (Im 2002).

For the unpaired cases, define $T_i^* = T_i - \mu$,

where T_i is defined as for the test statistic T_1 , μ is the expected sum of ranks for Y_1 group, which can be shown to be $K(K + L + 1)/2$ (Lehmann 1975).

Hence the test statistic T_4 is defined as $T_4 = T_{c4} + T_i^*$.

1.3 NEW TEST STATISTIC: T_5 PROPOSED BY JOHN BRYANT

In addition to the previous 4 tests, there is another option based on the Mann-Whitney U Statistic; we call it Test statistic T_5 .

For testing the hypothesis $H_0 : \Delta = 0$, Mann and Whitney (1947) proposed the statistic

$$U = \sum_{i=1}^m \sum_{j=1}^n \phi(X_i, Y_j)$$

Where

$$\phi(X_i, Y_j) = \begin{cases} 1, & \text{if } X_i < Y_j, \\ 0, & \text{otherwise.} \end{cases}$$

Namely, U is the number of times a y precedes an x. It can be computed as follows. For each pair of values X_i and Y_j , observe which is smaller. If the X_i value is smaller, score one for that pair; if the Y_j value is smaller, score 0 for that pair. Add up the 0s and 1s and call the sum U . The advantage of this new test statistic over the previous four test statistics is that each value

in one group is compared with every value of the other group despite it is paired or unpaired observation, hence comparisons are made across the complete pairs and unpaired 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 (Hollander and Wolfe 1999). Therefore, instead of using Mann-Whitney U Test, I will just use the rank sum test for simplicity of computation. Combine the data consisting both complete pairs and unpaired observations as following:

$$y_{1,1}, \dots, y_{1,j}, y_{1,j+1}, \dots, y_{1,j+k}, y_{2,1}, \dots, y_{2,j}, y_{2,j+k+1}, \dots, y_{2,j+k+l}$$

Let R_{n5} be the rank of each observation in Y_1 group in the combined sample including both observed values from Y_1 and observed values from Y_2 , then the test statistic is defined as

$$T_5 = \sum_{n=1}^{j+k} R_{n5} .$$

1.4 NULL HYPOTHESIS DISTRIBUTIONS FOR $T_1 - T_5$

Permutation algorithms are used to obtain the null hypothesis distribution for $T_1 - T_5$. Permutation tests are well understood and thoroughly documented in the statistical literature. Though not always as powerful as their parametric counterparts, they sometimes have equal or even greater power. Often they can be used when asymptotic theory falls short (e.g. small samples), and when fully enumerated, they provide exact results (as opposed to approximations based on asymptotic theory). Most pertinent to the current discussion is their reliance on few

distributional assumptions, giving permutation tests a much broader range of application (Opdyke 2002).

For the complete pairs, permute within each pair. Thus, there are 2^j possible arrangements, with each of these distinct arrangements equally likely to occur under the null hypothesis. Take the data configuration of 3-2-3 (3 complete pairs, 2 unpaired values for Y_1 and 3 unpaired values for Y_2) as an example.

For test statistic T_1 , the arrangements of ranks for the three ranks of paired value from Y_1 group will always occur as one of the following:

(1,1,1), (1,1,2), (1,2,1), (1,2,2), (2,1,1), (2,1,2), (2,2,1), (2,2,2).

For test statistic T_3 , whatever the original value and order of ranks would be for the three paired observations in Y_1 group, the permutations will always be performed between 1 and 6, 2 and 5, 3 and 4. So the possible value of rank sum for paired cases will be one of the following:

6, 7, 9, 10, 11, 12, 14, 15.

For test statistic T_4 , despite of the original ranks for the three absolute difference values, the rank sum for the complete pairs will always be the sum of ranks with positive differences minus the sum of ranks with negative differences; hence the possible value would be one of the following:

-6, -4, -2, 0, 2, 4, 6.

The possible arrangements for test statistics T_2 and T_5 depend on the original ranks given to each observation. If for T_2 the original ranks for complete pairs are as follows:

Pair	Y_1	Y_2
1	4	3
2	1	5
3	6	2

Then the possible arrangements for three ranks of Y_1 are as followings:

(4,1,6), (4,1,2), (4,5,6), (4,5,2), (3,1,6), (3,1,2), (3,5,6), (3,5,2).

This is not always the case for T_5 , since the ranks for T_5 could include values up to 11, say the original ranks for complete pairs are as follows:

Pair	Y_1	Y_2
1	11	3
2	1	5
3	6	8

Then the possible arrangements for three ranks of Y_1 are as followings:

(11,1,6), (11,1,8), (11,5,6), (11,5,8), (3,1,6), (3,1,8), (3,5,6), (3,5,8).

For unpaired cases, we choose K ranks from a total of $K + L$ ranks, that is

$\binom{K+L}{K}$ total number of rank sums based on these combinations in the sample space. Still take

the data configuration of 3-2-3 as an example, any two values of ranks can be chosen for Y_1 out of five possible values of ranks from combined sample. All the combinations of two ranks are as following:

(1,2), (1,3), (1,4), (1,5), (2,3), (2,4), (2,5), (3,4), (3,5), (4,5).

The combinations for test statistic T_5 are slightly different from the above. Instead of only choosing from 1 to 5, the combinations for ranks of unpaired cases in T_5 depend on the original ranks given to each unpaired observation. If the original ranks for unpaired cases are as follows:

Pair	Y_1	Y_2
4	5	
5	10	
6		1

7		4
8		2

Then the possible combinations for three ranks of Y_1 are as follows:

(5,10), (5,1), (5,4), (5,2), (1,10), (4,10), (2,10), (1,4), (1,2), (4,2).

Calculate the test statistic for each permutation, and then the probability distribution of

$T_1 - T_5$ under the null hypothesis is

$$P(T_q = t_q) = \frac{N(t_q)}{2^{J*} \binom{K+L}{K}}$$

where $q = 1,2,3,4,5$ and $N(t_q)$ is the number of occurrence of the test statistic t_q based on the data.

2.0 APPLICATION

All the previous four test statistics and the new test statistic proposed in this paper are applied to data comparing 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 (Weidmann 1992). We expected that more changes would be detected in surface receptors of T lymphocytes in the presence of tumor. The outcome variable was the percentage of T cells of each type showing the $V\beta 22$ receptor. For the purpose of this analysis, this percentage was treated as a continuous value rather than as a binomial proportion. 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.

The following one-sided test is used:

$$H_0 : M_{TIL} = M_{PBL}$$

$$H_a : M_{TIL} > M_{PBL}$$

Where M_{TIL} = median proportion of $V\beta$ gene in TIL

M_{PBL} = median proportion of $V\beta$ gene in PBL

Table 1. Data and application of test statistic T_1 .

Data from 8 patients showing the data configuration of 3-2-3 and respective ranks based on test statistic T_1 :

Patient	Data (3-2-3)		Ranks	
	$V\beta\%$		T_1	
	TIL	PBL	TIL	PBL
1	6.7	2.8	2	1
2	3.7	3.5	2	1
3	4.4	4.1	2	1
4	2.3	.	1	.
5	4.5	.	4	.
6	.	4.0	.	3
7	.	14.7	.	5
8	.	3.2	.	2

According to Table 1, $T_1 = (2+2+2) + (1+4) = 11$.

Table 2. Null hypothesis distribution of test statistic T_1 .

Possible values of T_1	Probability under H_0
6	1/80
7	4/80
8	8/80
9	12/80
10	15/80
11	15/80
12	12/80
13	8/80
14	4/80
15	1/80

$$P(T_1 \geq 11) = P(T_1 = 11) + P(T_1 = 12) + P(T_1 = 13) + P(T_1 = 14) + P(T_1 = 15) = 40/80 = 0.50$$

Table 3. Data and application of test statistic T_2 .

Data from 8 patients showing the data configuration of 3-2-3 and respective ranks based on test statistic T_2 :

Patient	Data (3-2-3)		Ranks	
	$V\beta\%$		T_2	
	TIL	PBL	TIL	PBL
1	6.7	2.8	6	1
2	3.7	3.5	3	2
3	4.4	4.1	5	4
4	2.3	.	1	.
5	4.5	.	4	.
6	.	4.0	.	3
7	.	14.7	.	5
8	.	3.2	.	2

According to Table 3, $T_2 = (6+3+5) + (1+4) = 19$.

Table 4. Null hypothesis distribution of test statistic T_2 .

Possible values of T_2	Probability under H_0
10	1/80
11	3/80
12	5/80
13	7/80
14	8/80
15	8/80
16	8/80
17	8/80
18	8/80
19	8/80
20	7/80
21	5/80
22	3/80
23	1/80

$$P(T_2 \geq 19) = P(T_2 = 19) + P(T_2 = 20) + P(T_2 = 21) + P(T_2 = 22) + P(T_2 = 23) = 24/80 = 0.30$$

Table 5. Data and application of test statistic T_3 .

Data from 8 patients showing the data configuration of 3-2-3 and respective ranks based on test statistic T_3 :

Patient	Data (3-2-3)		Ranks	
	$V\beta\%$		T_3	
	TIL	PBL	TIL	PBL
1	6.7	2.8	6	1
2	3.7	3.5	4	3
3	4.4	4.1	5	2
4	2.3	.	1	.
5	4.5	.	4	.
6	.	4.0	.	3
7	.	14.7	.	5
8	.	3.2	.	2

According to Table 5, $T_3 = (6+4+5) + (1+4) = 20$.

Table 6. Null hypothesis distribution of test statistic T_3 .

Possible values of T_3	Probability under H_0
9	1/80
10	2/80
11	3/80
12	5/80
13	6/80
14	7/80
15	8/80
16	8/80
17	8/80
18	8/80
19	7/80
20	6/80
21	5/80
22	3/80
23	2/80
24	1/80

$$P(T_3 \geq 20) = P(T_3 = 20) + P(T_3 = 21) + P(T_3 = 22) + P(T_3 = 23) + P(T_3 = 24) = 17/80 = 0.2125$$

Table 7. Data and application of test statistic T_4 .

Data from 8 patients showing the data configuration of 3-2-3 and respective ranks based on test statistic T_4 :

Patient	Data (3-2-3)		Ranks	
	$V\beta\%$		T_4	
	TIL	PBL	TIL	PBL
1	6.7	2.8	3 (+)	
2	3.7	3.5	1 (+)	
3	4.4	4.1	2 (+)	
4	2.3	.	1	.
5	4.5	.	4	.
6	.	4.0	.	3
7	.	14.7	.	5
8	.	3.2	.	2

Note: (+) means the sign of the difference within a pair is positive; (-) means the sign of the difference within a pair is negative.

According to Table 7, $T_4 = [(3+1+2)-0] + [(1+4)-2*(2+3+1)/2] = 5$.

Table 8. Null hypothesis distribution of test statistic T_4 .

Possible values of T_4	Probability under H_0
-9	1/80
-8	1/80
-7	3/80
-6	3/80
-5	5/80
-4	4/80
-3	7/80
-2	5/80
-1	8/80
0	6/80
1	8/80
2	5/80
3	7/80
4	4/80
5	5/80
6	3/80
7	3/80
8	1/80
9	1/80

$$P(T_4 \geq 5) = P(T_4 = 5) + P(T_4 = 6) + P(T_4 = 7) + P(T_4 = 8) + P(T_4 = 9) = 13/80 = 0.1625$$

Table 9. Data and application of test statistic T_5 .

Data from 8 patients showing the data configuration of 3-2-3 and respective ranks based on test statistic T_5 :

Patient	Data (3-2-3)		Ranks	
	$V\beta\%$		T_5	
	TIL	PBL	TIL	PBL
1	6.7	2.8	10	2
2	3.7	3.5	5	4
3	4.4	4.1	8	7
4	2.3	.	1	.
5	4.5	.	9	.
6	.	4.0	.	6
7	.	14.7	.	11
8	.	3.2	.	3

According to Table 9, $T_5=10+5+8+1+9=33$.

Table 10. Null hypothesis distribution of test statistic T_5 .

Possible values of T_5	Probability under H_0
17	1/80
18	2/80
19	1/80
20	1/80
21	2/80
22	2/80
23	3/80
24	3/80
25	4/80
26	6/80
27	4/80
28	4/80
29	5/80
30	4/80
31	5/80
32	4/80
33	4/80
34	6/80
35	4/80
36	3/80
37	3/80
38	2/80
39	2/80
40	1/80
41	1/80
42	2/80
43	1/80

$$\begin{aligned} P(T_5 \geq 33) &= P(T_5 = 33) + P(T_5 = 34) + P(T_5 = 35) + P(T_5 = 36) + P(T_5 = 37) \\ &+ P(T_5 = 38) + P(T_5 = 39) + P(T_5 = 40) + P(T_5 = 41) + P(T_5 = 42) + P(T_5 = 43) \\ &= 29/80 = 0.3625 \end{aligned}$$

Therefore, based on the above five nonparametric test statistics, T_1, T_2, T_3, T_4 and T_5 , there is no evidence that the proportions of the $V\beta$ gene family differ in tumor infiltrating lymphocytes (TILs) and peripheral blood lymphocytes (PBLs).

3.0 MONTE CARLO SIMULATION STUDY

A Monte Carlo simulation study was carried out to compare the powers of the five test statistics mentioned in this paper including the one newly proposed. Bivariate normally distributed datasets were generated with means μ_1 and μ_2 , correlation coefficient ρ , and common variance σ^2 . Each dataset generated was of the same missing data configuration 3-2-3 (J=3, K=2 and L=3 in Figure 1). 5000 simulations were performed for each combination of the parameters for a bivariate normal distribution (shown in Table 11); the variances of two groups will be kept equal to 1 without loss of generality, while changing the mean difference or the correlation between two groups. The situation of mean difference ranging from 0 to 3, as well as the correlation varying from 0 (no correlation), 0.2 (a small correlation) to 0.5 (a moderate correlation) is analyzed to assess the effect of changing mean difference and correlation on the results of the power comparisons for the five test statistics. Programs are written in SAS (Statistical Analysis Systems, v9.1, NC) to calculate the five test statistics and their respective null hypothesis distributions. One-sided alpha values of 0.05, 0.025, and 0.0125 were used to obtain the critical values for rejecting the null hypothesis for each test statistic. Then the program compared each test statistic to its unique critical value to get the total number of rejections from 5000 iterations, so that the empirical power can be calculated by dividing the total number of rejections by 5000 for each test. The power for each test under various conditions is presented in Table 12-14.

With Type I error set to be 0.05, T_5 was the most powerful test statistic among the five test statistics, followed by T_2 and T_3 . T_3 had a slightly higher power than T_2 in most cases, except for ones with a moderate correlation of 0.5 and relatively larger mean differences of 2.0 or 3.0.

With Type I error set to be 0.025, T_1 and T_3 shared the same power, although both were less powerful than the others. T_4 ranked the most powerful test statistics most of the time, except being exceeded by T_5 in some cases with relatively bigger mean differences and smaller correlations.

With Type I error set to be 0.0125, all five test statistics have the same power under all parameter combinations. The main reason for this happening is because analyses were done on the same dataset generated for each test statistic. Therefore, we will get same number of rejections for each test statistic according to a quite large critical value for an alpha value of 0.0125.

After doing all the simulations, it is easy to find out that the null hypothesis distributions for T_1 , T_3 and T_4 are invariant if datasets with the same missing data configuration are being analyzed without ties, which means that the null hypothesis distribution remain the same no matter what the original order of ranks would be for each dataset generated with the same missing data configuration. This fact contributes to the simplicity of SAS program coding and improves the efficiency of running the simulation. However, the null hypothesis distributions for T_2 and T_5 do change with the dataset generated although all the dataset are of the same missing data configuration.

Although the Type I error was set before performing each test, the actual significance level should still be examined after all the tests were done. From Table 12, true significance levels under various conditions are all below 0.05, with T_5 having a relatively higher Type I error. From Table 13, it is evident that T_4 has the highest significance level among the five tests. Although most Type I errors are still under 0.025, there is one exception for the significance level of T_4 under the correlation of 0.2 with a slightly higher significance level of 0.0264. The actual significance levels under correlation of 0.0 and 0.2 are also slightly higher than 0.0125 according to Table 14. These figures greater than alpha must be due to “simulation” variation. So generally speaking, all the true significance levels are close or even below the Type I errors being set. So the power comparisons among the five test statistics can be relied on.

Checking was also done to see if 5000 simulations was a reasonable sample size to detect the power difference among the five test statistics. Table 15 shows confidence intervals for each type I error and various power values of 0.30, 0.50, 0.80, and 0.90. All the confidence intervals seem acceptable for a sample size of 5000, lending credibility to the sample size chosen for this Monte Carlo study.

Table 11. All combinations of parameters for datasets generated from a bivariate normal distribution.

μ_1	μ_2	σ^2	ρ
1	1	1	0.0
2	1	1	0.0
3	1	1	0.0
4	1	1	0.0
1	1	1	0.2
2	1	1	0.2
3	1	1	0.2
4	1	1	0.2
1	1	1	0.5
2	1	1	0.5
3	1	1	0.5
4	1	1	0.5

Table 12. Power comparisons for five test statistics under various conditions with $\sigma^2 = 1$ and $\alpha = 0.05$.

μ_d : mean differences

ρ : correlation coefficient

T_1 : test statistic proposed by J. Wilson

T_2 : test statistic proposed by Brunner & Neumann (1984)

T_3 : test statistic proposed by P.K. Sen

T_4 : test statistic proposed by KyungAh Im

T_5 : new test statistic first proposed by John Bryant

		T_1	T_2	T_3	T_4	T_5
ρ	μ_d	$\Pr\{rejectH_0\}$	$\Pr\{rejectH_0\}$	$\Pr\{rejectH_0\}$	$\Pr\{rejectH_0\}$	$\Pr\{rejectH_0\}$
0.0	0.0	0.0128	0.0338	0.0362	0.025	0.0434
	1.0	0.1614	0.3122	0.332	0.2458	0.3664
	2.0	0.551	0.7758	0.795	0.6768	0.8344
	3.0	0.8746	0.9672	0.9736	0.9342	0.9864
0.2	0.0	0.0136	0.0364	0.037	0.0264	0.0438
	1.0	0.1794	0.3412	0.3568	0.2736	0.3858
	2.0	0.5958	0.8152	0.8298	0.7336	0.8626
	3.0	0.8966	0.9802	0.9802	0.9588	0.9908
0.5	0.0	0.0122	0.0352	0.0376	0.0234	0.0428
	1.0	0.221	0.3992	0.414	0.3366	0.44
	2.0	0.6668	0.8598	0.856	0.8138	0.8968
	3.0	0.9188	0.9836	0.98	0.9768	0.994

Table 13. Power comparisons for five test statistics under various conditions with $\sigma^2 = 1$ and $\alpha = 0.025$.

Notations follow those in Table 12.

		T_1	T_2	T_3	T_4	T_5
ρ	μ_d	$\Pr\{rejectH_0\}$	$\Pr\{rejectH_0\}$	$\Pr\{rejectH_0\}$	$\Pr\{rejectH_0\}$	$\Pr\{rejectH_0\}$
0.0	0.0	0.0128	0.0182	0.0128	0.025	0.0202
	1.0	0.1614	0.1998	0.1614	0.2458	0.232
	2.0	0.551	0.6192	0.551	0.6768	0.6862
	3.0	0.8746	0.9132	0.8746	0.9342	0.9478
0.2	0.0	0.0136	0.0168	0.0136	0.0264	0.0198
	1.0	0.1794	0.2164	0.1794	0.2736	0.248
	2.0	0.5958	0.6784	0.5958	0.7336	0.7286
	3.0	0.8966	0.9388	0.8966	0.9588	0.9618
0.5	0.0	0.0122	0.0144	0.0122	0.0234	0.019
	1.0	0.221	0.2634	0.221	0.3366	0.2936
	2.0	0.6668	0.762	0.6668	0.8138	0.7892
	3.0	0.9188	0.966	0.9188	0.9768	0.9742

Table 14. Power comparisons for five test statistics under various conditions with $\sigma^2 = 1$ and $\alpha = 0.0125$.

Notations follow those in Table 12.

		T_1	T_2	T_3	T_4	T_5
ρ	μ_d	$\Pr\{rejectH_0\}$	$\Pr\{rejectH_0\}$	$\Pr\{rejectH_0\}$	$\Pr\{rejectH_0\}$	$\Pr\{rejectH_0\}$
0.0	0.0	0.0128	0.0128	0.0128	0.0128	0.0128
	1.0	0.1614	0.1614	0.1614	0.1614	0.1614
	2.0	0.551	0.551	0.551	0.551	0.551
	3.0	0.8746	0.8746	0.8746	0.8746	0.8746
0.2	0.0	0.0136	0.0136	0.0136	0.0136	0.0136
	1.0	0.1794	0.1794	0.1794	0.1794	0.1794
	2.0	0.5958	0.5958	0.5958	0.5958	0.5958
	3.0	0.8966	0.8966	0.8966	0.8966	0.8966
0.5	0.0	0.0122	0.0122	0.0122	0.0122	0.0122
	1.0	0.221	0.221	0.221	0.221	0.221
	2.0	0.6668	0.6668	0.6668	0.6668	0.6668
	3.0	0.9188	0.9188	0.9188	0.9188	0.9188

Table 15. 95% confidence intervals for a sample size of 5000 simulations with regard to varied Type I errors and test powers.

Using the formula $\alpha \pm 1.96 * \sqrt{\frac{\alpha * (1 - \alpha)}{5000}}$ for calculating 95% confidence intervals for a Type I error of 0.05, 0.025 or 0.0125;

Using the formula $\beta \pm 1.96 * \sqrt{\frac{\beta * (1 - \beta)}{5000}}$ for calculating 95% confidence intervals for a test power of 0.30, 0.50, 0.80 or 0.90;

Where α is the Type I error of 0.05, 0.025 or 0.0125; β is the power of 0.30, 0.50, 0.80, or 0.90.

Type I error	95% Confidence Interval
0.05	(0.0440, 0.0560)
0.025	(0.0207, 0.0293)
0.0125	(0.0094, 0.0156)
Test Power	95% Confidence Interval
0.30	(0.2873, 0.3127)
0.50	(0.4861, 0.5139)
0.80	(0.7889, 0.8111)
0.90	(0.8917, 0.9083)

4.0 CONCLUSIONS AND DISCUSSION

This paper mainly talked about the nonparametric methods for comparing medians of paired data with unpaired values on both responses or, to put it differently, nonparametric test statistics for comparing medians with incomplete paired data. First of all, four existing nonparametric test statistics were discussed. Then a new test statistic based on a Mann-Whitney U test making comparisons across complete and incomplete pairs was being proposed. The dataset being analyzed by all five test statistics must satisfy the assumption of missing completely at random (MCAR) and without tied values.

Secondly, the null hypothesis distribution for each test statistic was introduced and derived by a permutation approach, since bivariate normality cannot be assessed with the small number of pairs in this data set. All the possible values of test statistic were calculated and the probability of each value occurring was derived using the statistical package SAS (Statistical Analysis System, v9.1, SAS institute, Cary, NC). There followed an analysis of a real dataset comparing the proportions of certain T cell receptor gene families, to which all five test statistics with their respective null hypothesis distribution applied.

The main purpose of this paper was to compare the power of those five test statistics under different Type I errors, which was done using a Monte Carlo simulation study. Datasets with certain missing data configuration (3-2-3 in this paper) were generated from a bivariate normal distribution with different sets of parameter combinations in order to find out the effect

of different set of parameter combinations on the result of the power comparisons. At the Type I error of 0.05 or 0.025, most tests under different conditions showed to be conservative, with actual Type I error less than the level of 0.05 or 0.025. Even for an α value of 0.0125, although most actual significance level exceeded 0.0125, it is still quite acceptable since those levels were quite close to 0.0125. Besides, the closer the Type I error is to the alpha value being set, the more powerful the test is under the same correlation correlation.

Generally speaking, under different α levels and various sets of parameter combinations, the newly proposed test statistic T_5 is fairly powerful especially under an α value of 0.05. T_2 and T_3 also have a favorable power under an α of 0.05, while T_4 did its best under an α value of 0.025. When an α value was decreased to 0.0125, there exits no difference in power for all five test statistics. Moreover, at each alpha level, the power for each test statistic increased dramatically with the increase in mean differences. For example, the power jumped by approximately 0.60 from a mean difference of 1.0 to a mean difference of 3.0 for almost all the tests, leading to a high power of above 0.90 for all the tests with a mean difference of 3.0. With correlation coefficient increased from 0.0, 0.2 to 0.5, the power for each test also showed slight improvement, say increased by approximately 10%.

So far only a certain missing data configuration (3-2-3) has been concerned in this paper. However, further studies could be done to change the missing data configurations, say 3-2-3, 6-4-6, 12-8-12, 6-1-1, 14-1-1, or even 30-1-1. Especially study done on the 6-1-1 or 14-1-1 might mean a lot for clinical trial data analyses, since it may often happen that one observation might be missing in a fairly small sample size. People can also discuss the question of whether adding unpaired observations help to improve the power under various missing data configurations. Another possibility for further studies will be to check the results of this paper to see if they also

hold up under different bivariate population distributions. The introduction of this new nonparametric test statistic is of public health importance because it is a powerful statistical method for dealing with a pattern of missing data that may be encountered in clinical and public health research.

APPENDIX A

NULL HYPOTHESIS DISTRIBUTION FOR T_1

Possible values of T_1	Probability under H_0
6	1/80
7	4/80
8	8/80
9	12/80
10	15/80
11	15/80
12	12/80
13	8/80
14	4/80
15	1/80

APPENDIX B

NULL HYPOTHESIS DISTRIBUTION FOR T_3

Possible values of T_3	Probability under H_0
9	1/80
10	2/80
11	3/80
12	5/80
13	6/80
14	7/80
15	8/80
16	8/80
17	8/80
18	8/80
19	7/80
20	6/80
21	5/80
22	3/80
23	2/80
24	1/80

APPENDIX C

NULL HYPOTHESIS DISTRIBUTION FOR T_4

Possible values of T_4	Probability under H_0
-9	1/80
-8	1/80
-7	3/80
-6	3/80
-5	5/80
-4	4/80
-3	7/80
-2	5/80
-1	8/80
0	6/80
1	8/80
2	5/80
3	7/80
4	4/80
5	5/80
6	3/80
7	3/80
8	1/80
9	1/80

BIBLIOGRAPHY

- Ekbohm G. (1976) On Comparing Means in the Paired Case with Incomplete data on both responses. *Biometrika*, 63(2): p299-304.
- Hollander M., Wolfe D.A. (1999) *Nonparametric Statistical Methods*. 2nd Ed. New York : J. Wiley, c1999.
- Im, K.A. (2002) *Modified Signed Rank Tests to Account for Missing-ness in a Small Sample Paired Data*. MS Thesis, Department of Biostatistics, Graduate School of Public Health, University of Pittsburgh, Pittsburgh, PA, USA.
- Lehmann E.L. (1975) *Nonparametrics: Statistical methods based on ranks*. Holden-Day, Inc. San Francisco, California.
- Lin P. Stivers L.E. (1974) On difference of means with incomplete data. *Biometrika*, 61(2): p325-334.
- Little Roderick J. A. (1988) A Test of Missing Completely at Random for Multivariate Data with Missing Values, *Journal of the American Statistical Association*, 83, 1198-1202.
- Opdyke J.D. (2002) *Fast Two-Sample Permutation Tests, Even When One Sample is Large, that Efficiently Maximize Power Under Crude Monte Carlo Sampling*. The Pharmaceutical Industry SAS Users Group.
- Wilson J.W. (1992) *Nonparametric tests for the paired two-sample problem with missing data* [Unpublished manuscript].