Statistical Method for Analysis of Responses in Control Critical Trials with Three Outcomes

Oyeka I. C. A. and Nwankwo Chike H.

Abstract— **This paper proposes a statistical method for the analysis of multiple responses or outcome data in case control studies including situations in which the observations are either continuous or frequency data. Test statistics are proposed for assessing the statistical significance of differences between case-control response score. The proposed methods are illustrated with some sample data. When there only three possible response options in which the proposed method and the Stuart- Maxwell test can be equally used to analyse the data, the proposed test statistic is show to be at least as powerful as the Stuart-Maxwell test statistic.**

Index Terms— **Multiple Response, Case Control, Scores, Test Statistic, Treatment, Prospective, Retrospective**

I. INTRODUCTION

Often in controlled comparative prospective or retrospective studies involving matched samples of subjects or patients, the response of a subject to a predisposing factor in a retrospective study or to a condition or treatment in a prospective study may be dichotomous with only two possible naturally exclusive outcomes and appropriate for analysis using the McNemar Test (Gibbons1973). But the responses may be much finer than simply dichotomous, assuming several possible values. For example in a retrospective study where the predisposing factor may be a subject's employment status, a subject may be classified as unemployed, self employed, public servant, student, housewife etc. In a prospective study involving some conditions or tests, subjects or patients may be classified as recovered, much improved, improved, no change, worse or dead. A treatment or drug may be graded as very effective, effective, ineffective etc

If there are only three possible response options or categories, then the Stuart-Maxwell test (Fleiss, 1981; Robertson et al, 1974; Schlesselman, 1992; Zhao and Kolonel, 1992; Box and Cox, 1964; Maxwell, 1970; Stuart, 1955; Fleiss, 1981; Everitt, 1977) may be used to analyse the data. We here propose an alternative and easier to use method that is often more powerful than the usual Stuart/Maxwell test for three outcomes in a clinical trial and which is easily generalisable when there are more than three outcomes.

II. THE PROPOSED METHOD

Suppose we have a random sample of n pairs of patients or subjects matched on a number of characteristics to be exposed to two experimental conditions, treatments, drugs or tests. Suppose further that the responses of these pairs of subjects are more than dichotomous but numbering $c(c \geq 3)$ possible response options. Suppose further that the i_{th} pair of patients is selected, for i =1, 2….., n and one member of the pair is randomly assigned to one of the treatments T_1 (standard drug; control), say, and the remaining member of the pair is assigned to the second treatment T_2 (new drug; case) say, and the various c possible responses are recorded for each subject. If in particular the responses of each matched pair of subjects are classified into c = 3 mutually exclusive categories or classes, the data presentation format is as in Table 1 below.

Table 1: Format For Presentation Of Data On 'C' = 3 Outcomes in a Clinical Trial of Matched Pairs

	Outcome Category for Control (Standard T_1)			
Outcome Category for Cases (Experimental Condition T_2)	1	2	3	Total ($n_{i.}$)
1	n_{11}	n_{12}	n_{13}	$n_{1.}$
2	n_{21}	n_{22}	n_{23}	$n_{2.}$
3	n_{31}	n_{32}	n_{33}	$n_{3.}$
Total ($n_{.j}$)	$n_{.1}$	$n_{.2}$	$n_{.3}$	n..(=n)

Each entry in Table 1 consists of a matched pair of case and control subjects. For example n_{ij} is the number of pairs in which the case is in category 'i' response while the corresponding control subject is in outcome or response category j for $i, j = 1,2,3.\ n_{i.} and\ n_{.j}$ are respectively the total number of pairs in which the case is in category 'i' response and the control is in category 'j' response for $i, j = 1,2,3.$

In all, there are a total of $n = n.. = \sum_{i=1}^{3}\sum_{j=1}^{3} n_{ij} = \sum_{i=1}^{3} n_{i.} = \sum_{j=1}^{3} n_{.j}$ pairs of subjects studied. A null hypothesis usually tested using the Stuart-Maxwell test is that case and control subjects or patients do not differ in their response to the treatments. The corresponding star-Maxwell test statistic for this purpose is

$$x^2 = \frac{\bar{n}_{23}.d_1^{\,2} + \bar{n}_{13}.d_2^{\,2} + \bar{n}_{12}.d_3^{\,2}}{2(\bar{n}_{12}.\bar{n}_{13} + \bar{n}_{12}.\bar{n}_{23} + \bar{n}_{13}.\bar{n}_{23})} \quad \ldots (1)$$

which under H_o has approximately the chi-square distribution with 2 degrees of freedom, for sufficiently large $n = n..where$

Manuscript received January 30, 2015.

Oyeka I. C. A. and Nwankwo Chike H., Department of Statistics, Nnamdi Azikiwe University, Awka Nigeria.

$$d_i = n_{i.} - n_{.i} \quad \ldots (2)$$

and

$$\bar{n}_{ij} = \frac{n_{ij} + n_{ji}}{2} \quad \ldots (3)$$

for $i, j = 1,2,3; i \neq j$

Now to develop the proposed method let, as in Stuart-Maxwell method, the difference between the number of pairs of respondents in the i_{th} category of responses for case and j_{th} category of responses for control (Miettinen, 1969; Maxwell, 1970; Everitt, 1977; Stuart, 1955) be d_i (equation 2) which is independent of n_{ii}, i = 1, 2, 3 the number of pairs in which both case and control subjects have the same response or outcome. Also let

$$d_{ij} = n_{ij} - n_{ji} \quad \ldots (4)$$

which is the difference between the number of pairs in which the case is in the response category *i* and the control is in the response category j and the number of pairs in which the case is in response category j and the control is in the response category *i; i,j* = 1, 2, 3 $i \neq j$.

Now having selected our random sample of n matched pairs, let x_{i1} be the response by a member of the randomly selected i_{th} pair of patients or subjects randomly assigned treatment T_1 (control, standard drug) and x_{i2} be the response by the other member of the pair of patients or subjects assigned treatment T_2 (case, new drug) for i = 1, 2, ….., n. We here assume for ease of presentation, but without loss of generality, that the three mutually exclusive possible response categories have been ordered from the highest or most serious (lowest or least serious) level of response to the lowest or least serious (highest or most serious) level of response. For example, a patient's response to a treatment for an illness or disease may range from recovered, through no change to dead; a subject's response to a screening test may range variously from definitely positive, no change, to definitely negative. A candidate's or student's performance in a job interview or examination may range from very poor, good, to excellent.

We here assume that those responses have been appropriately arranged either in increasing or decreasing order of seriousness.

Now let

U_i {

1, If x_{i2}, i.e the response by the member in the i_{th} pair of patients or subjects assigned treatment T_2 (case) is a higher or more serious (lower or less serious) level of response than x_{i1}, the response by the other member of the pair assigned treatment T_1 (control) for all the 3 response categories.

… 5

0, if x_{i1} and x_{i2}, are the same level of response for the two patients or subjects in the i_{th} pair for all the 3 response categories

-1, if x_{i1}, the response by the member in the i_{th} pair of patients or subjects assigned treatment T_2 (case) is a lower or less serious (higher or more serious) level of response than x_{i1}, the response by the other member of the pair assigned treatment T_1 (control) for all the 3 response categories

For i =1, 2,…,n

This means that u_i assumes the value 1, if the response of the member of the i_{th} pair of patients administered treatment T_2 (case) is a higher or more serious (lower or less serious) level or response than the response of the other member of the pair administered treatment T_i (control); 0, if the response of the two members of the pair are the same, and -1, if treatment T_2 (case) the response of the pair administered is a lower or less serious (higher or more serious) level of response than the response of the other member of the pair administered treatment T_1 (control) for all the 3 response categories.

Now let

$$\pi^+ = P(u_i = 1); \pi^0 = P(u_i = 0); \pi^- = P(u_i = -1) \quad \ldots 6$$

Where

$$\pi^+ + \pi^0 + \pi^- = 1 \quad \ldots 7$$

Let

$$W = \sum_{i=1}^{n} u_i \quad \ldots 8$$

Now

$$E(u_i) = \pi^+ - \pi^- \quad \ldots 9$$

And

$$Var(u_i) = \pi^+ + \pi^- - (\pi^+ - \pi^-)^2 \quad \ldots 10$$

Also

$$E(W) = \sum_{i=1}^{n} Eu_i = n(\pi^+ - \pi^-) \quad \ldots 11$$

Note that $\pi^+ - \pi^-$ is the differential response rate between the sub-populations administered treatments T_2 (case) and T_1 (control) respectively in the paired population of patients or subjects for all the c = 3 response categories and is estimated by

$$\hat{\pi}^+ - \hat{\pi}^- = \frac{W}{n} \quad \ldots 12$$

Note also that π^+, π^0 and π^- which are respectively the probabilities that a randomly selected case is at a higher (or more serious) level, the same or lower (or less serious) level of response than the corresponding control subject in the pair for all the three response categories are estimated using the frequencies in Table 1 and following the specification in Equation 5 as

$$\hat{\pi}^+ = P^+ = \frac{f^+}{n} = \frac{\sum_{i=1}^{2}\sum_{\substack{j=2 \\ i<j}}^{3} n_{ij}}{n} = \frac{n_{12}+n_{13}+n_{23}}{n} \quad \ldots 13$$

$$\hat{\pi}^0 = P^0 = \frac{f^0}{n} = \frac{\sum_{i=1}^{3} n_{ii}}{n} = \frac{n_{11}+n_{22}+n_{33}}{n} \quad \ldots 14$$

And

$$\hat{\pi}^- = P^- = \frac{f^-}{n} = \frac{\sum_{i=1}^{3}\sum_{\substack{j=1 \\ i>j}}^{2} n_{ij}}{n} = \frac{n_{21}+n_{31}+n_{32}}{n} \quad \ldots 15$$

Where f^+, f^0 and f^- are respectively the number of 1's, 0's and -1's in the frequency distribution of the n values of these numbers in u_i in accordance with Equation 5, $i = 1, 2\ldots\ldots n$

Hence using these results in equation12, we have that

$$W = f^+ - f^- = \sum_{\substack{i=1 \\ i<j}}^{2}\sum_{j=2}^{3} n_{ij} - \sum_{\substack{i=2 \\ i>j}}^{3}\sum_{j=1}^{2} n_{ij} = (n_{12}-n_{21}) + (n_{13}-n_{31}) + (n_{23}-n_{32}) \quad \ldots 16$$

Now from equations 8 and 10 we have that

$$Var(W) = \sum_{i=1}^{n} Var(u_i) = n((\pi^+ + \pi^-) - (\pi^+ - \pi^-)^2) \quad \ldots 17$$

Whose sample estimate is from equations 13 and 15 as

$$Var(W) = n((\hat{\pi}^+ + \hat{\pi}^-) - (\hat{\pi}^+ - \hat{\pi}^-)^2) \quad \ldots 18$$

As noted above π^+ is the proportion of pairs of case and control subjects in which on the average the response rate by the sub-population of patients or subjects administered treatment T_2 (experimental, case) is greater (less) than the rate by the sub-population of patients or subjects administered treatment T_1 (standard, control); while π^- is the proportion of pairs in which on the average the response rate by the sub-population of patients or subjects administered treatment T_1 (standard, control) is greater (less) than the response rate by the sub-population of patients administered treatment T_2 (experimental, case) in the paired population of patients or subjects for all the three response categories. Hence the null hypothesis that there exists no difference between the response rates by the sub-population of patients administered treatment T_2 (experimental, case) and the sub-population of patients administered treatment T_1 (standard, control) in the paired population of patients for all response categories is equivalent to the null hypothesis

$$H_0: \pi^+ - \pi^- = 0 \; versus \; H_1: \pi^+ - \pi^- \neq 0 \quad \ldots 19$$

To test this null hypothesis, we may use the test statistic

$$x^2 = \frac{W^2}{Var(W)} = \frac{W^2}{n(\hat{\pi}^+ + \hat{\pi}^- - (\hat{\pi}^+ + \hat{\pi}^-)^2)} = \frac{(f^+ - f^-)^2}{f^+ + f^- - \frac{(f^+ - f^-)^2}{n}} \quad \ldots 20$$

Which under H_0 has approximately a chi-square distribution with 1 degree of freedom for sufficiently large n. Although strictly speaking, the test statistic in Equation 20 has a Chi-square distribution with 1 degree of freedom, however because its construction in equation 5 involves a combination of some c = 3 response categories, to help increase its power and reduce the chances of erroneously accepting a false null hypothesis (Type II error), it is here recommended that all comparisons should be made against critical Chi-square values with 3 -1 =2 degrees of freedom, instead of 1 degree of freedom. Hence here H_0 is rejected at the α level of significance if

$$x^2 \geq x^2_{(1-\alpha;2)} \quad \ldots 21$$

Otherwise H_0 is accepted.

Note that

$$n(\hat{\pi}^+ + \hat{\pi}^-) = \sum_{\substack{i=1 \\ i<j}}^{2}\sum_{j=2}^{3} n_{ij} + \sum_{i=2}^{3}\sum_{j=1}^{2} n_{ij} = n\big((n_{12}+n_{13}+n_{23}) + (n_{21}+n_{31}+n_{32})\big) \quad \ldots 22$$

Hence using equation 16 and 22 in equation 20 the test statistic becomes

$$\chi^2 = \frac{\big((n_{12}-n_{21})+(n_{13}-n_{31})+(n_{23}-n_{32})\big)^2}{n_{12}+n_{13}+n_{23}+n_{21}+n_{31}+n_{32} - \left(\frac{(n_{12}-n_{21})+(n_{13}-n_{31})+(n_{23}-n_{32})}{n}\right)^2} \quad \ldots 23$$

If there are only two possible outcomes or responses, that is c = 2, equation 20 under H_0 reduces to a modified version of the McNemar test statistic which is

$$\chi^2 = \frac{(n_{12}-n_{21})^2}{n_{12}+n_{21}-\frac{(n_{12}-n_{21})^2}{n}} \quad \ldots 24$$

This has a chi-square distribution with 1 degree of freedom.

Note that equation 24 has smaller variance than the usual McNemar test because of its modification to provide for possible ties between case and control subject pairs in their responses.

If c = 3, equation 20, under H_0, reduces to

$$\chi^2 = \frac{\left((n_{12}-n_{21})+(n_{13}-n_{31})+(n_{23}-n_{32})\right)^2}{n_{12}+n_{13}+n_{21}+n_{23}+n_{31}+n_{32}-\frac{\left((n_{12}-n_{21})+(n_{13}-n_{31})+(n_{23}-n_{32})\right)^2}{n}} \quad \ldots 25$$

And this has a chi-square distribution with c – 1 = 3 – 1 = 2 degrees of freedom.

Finally, note that if we let

$$m_{ij} = n_{ij} + n_{ji}, \quad i,j = 1,2,3\,; \quad i \neq j \quad \ldots 26$$

Then the test statistic of equation 20 can be written in an easier and more compact form using equations 4 and 26 as

$$x^2 = \frac{\left(\sum_{i=1}^{s}\sum_{j=1}^{s} d_{ij}\right)^2}{\sum_{i=1}^{s}\sum_{j=1}^{s} m_{ij} - \frac{\left(\sum_{i=1}^{s}\sum_{j=1}^{s} d_{ij}\right)^2}{n}} \quad \ldots 27$$

If equation 20 leads to a rejection of the null hypothesis of equal response rates then one may wish to proceed to identify the response categories or combination of categories that may have led to the rejection of H_0. This is done by appropriately pooling or combining the response options into (2) groups and apply the McNemar test (McNemar 1983, Somes 1983, Sheskin 2000) to each of the groups. In all cases comparisons are made using critical chi-square values with 2 degrees of freedom to again avoid erroneous conclusions.

A. Illustrative Example 1

We here use data on matched pairs of 151 patients from a controlled comparative clinical trial who manifest three possible responses to illustrate the proposed method. Suppose the data in Table 2 are obtained by assigning a standard treatment T_1 (control) and a new treatment T_2 (case) at random to members of each pair of a random sample of 151 pairs of HIV patients matched on age, gender and body weight used in a controlled clinical trial to compare the effectiveness of two HIV drugs.

Table 2: Data from Controlled Comparative Clinical Trial Using Matched Pairs with Three Responses

	Standard Treatment T_1 (control)			
New reatment T_2 (case)	Improved	No Change	Dead	Total ($n_{i.}$)
Improved	60	31	4	95 (=$n_{1.}$)
No Change	16	24	6	46 (=$n_{2.}$)
Dead	3	4	3	10 (=$n_{3.}$)
Total ($n_{.j}$)	79 ($n_{.1}$)	59 ($n_{.2}$)	13 ($n_{.3}$)	151 (=$n_{..}$)

To test the null hypothesis that case and control do not differ in their response to the treatments (Equation 19,) we have from equation 13 that

$$\hat{\pi}^+ = \frac{31+4+6}{151} = \frac{41}{151} = 0.271$$

And from Equation 15, we have that

$$\hat{\pi}^- = \frac{16+3+4}{151} = \frac{23}{151} = 0.152$$

Note that $\hat{\pi}^0 = 1 - (0.271 + 0.152) = 0.577$

Also from Equation 11, we have that

$$E(W) = n(\hat{\pi}^+ - \hat{\pi}^-) = 151(0.271 + 0.152) = 151(0.119) = 17.969$$

From Equation 17, we have that

$$Var(W) = 151[(0.271 + 0.152) - (0.271 - 0.152)^2] = 61.735$$

Hence from Equation 23, we have that

$$x^2 = \frac{\left((31-16)+(4-3)+(6-4)\right)^2}{31+4+6+16+3+4-\frac{\left((31-16)+(4-3)+(6-4)\right)^2}{151}} = \frac{324}{63.986} = 5.064 \; (p-value = 0.008$$

Which with c – 1 = 3 -1 = 2 degrees of freedom is highly statistically significant at α = 0.01.

We may therefore conclude at the 1 percent significance level that the treatments have differential effects on the patients.

If we had used the Stuart/Maxwell method to analyse the data we would have from Equation 2 that

$$d_1 = 95 - 79 = 16, d_2 = 46 - 59 = -13, d_3 = 10 - 13 = -3$$

Also letting $\bar{n}_{ij} = \frac{n_{ij}+n_{ji}}{2}, i = 1,2,3; j = 1,2,3; i \neq j$, we have

$$\bar{n}_{12} = \frac{31+16}{2} = \frac{47}{2} = 23.5; \; \bar{n}_{13} = \frac{4+3}{2} = \frac{7}{2} = 3.5; \; \bar{n}_{23} = \frac{6+4}{2} = \frac{10}{2} = 5$$

Hence using the Stuart Maxwell test, we have

$$x^2 = \frac{5(16)^2+3.5(-13)^2+23.5(-3)^2}{2\left((23.5)(3.5)+(23.5)(5)+(3.5)(5)\right)} = \frac{1894}{434.5} = 4.359 \; (p-value = 0.12)$$

Which, with 2 degrees of freedom, is statistically significant at the 2 percent level of significance but not statistically significant at the 1 percent level of significance, the usually used norm in medical research?

Thus the present (extended) method leads to a rejection of the null hypothesis H_0 while the Stuart/Maxwell test statistic

leads to an acceptance of the null hypothesis at the 1 percent significance level. Hence the Stuart/Maxwell Test is likely to lead to an acceptance of a false null hypothesis (Type II error). This means that the present test statistic is likely to be more efficient and more powerful than the Stuart/Maxwell test statistic.

As noted above, the present method may also be used to analyse quantitative or numeric data obtained in matched controlled studies. Often, responses from controlled experiments are reported as numeric scores assuming all possible values on the real line such that scores in the interval (c_1, c_2) where c_1 and c_2 are real. For example, these responses may be values on the real line such that scores in the interval (c_1, c_2) where c_1 and c_2 are real numbers ($c_1 < c_2$), indicate that the responses by the subject concerned are normal, negative, condition absent, etc; values less than c_1 indicate that the subjects have abnormally low scores; and values above c_2 indicate that the subjects have abnormally high scores. It is also possible to have situations in which subjects have scores that are either some c_3 units below c_1 or some c_4 units above c_2. These subjects may be concerned to have non specific or non definitive manifestations. Subjects whose scores are below c_3 and above c_4 may be considered to have critically abnormal manifestations, one below the critical minimum and the other above the critical maximum normal scores. If these results are considered important manifestations, then the first set of subjects may be grouped into three response categories, while the second set of subjects may be grouped into five response categories for policy and management purposes.

To illustrate the use of the present method when the case and control subjects in matched controlled studies have quantitative scores with three possible outcomes for instance, we would proceed as follows:

Suppose as above, a random sample of n pairs of case and control subjects are used in a controlled experiment on two procedures T_1 (control, standard) and T_2 (case, experimental procedure). Suppose as before, one member of each pair is randomly assigned treatment T_1 (control, standard) and the remaining member assigned treatment T_2 (case, experimental procedure).

Let y_{i1} and y_{i2} be respectively the responses or scores with real values, quantitatively measured, by the subjects assigned treatment T_1 (control) and T_2 (case) for the i_{th} pair of subject for $i = 1,2,\dots,n.$

Then u_i of Equation 3 may now be defined as

$$u_i = \begin{cases} 1, & \text{if either } y_{i2} < c_1 \text{ and } c_1 \le y_{i1} \le c_2 \text{ or } y_{i2} < c_1 \\ & \text{and } y_{i1} > c_2 \text{ or } c_1 \le y_{i2} \le c_2 \text{ and } y_{i1} > c_2 \\ 0, & \text{if either } y_{i2} < c_1 \text{and } y_{i1} < c_1 \text{ or } c_1 \le y_{i2} \le c_2 \\ & \text{and } c_1 \le y_{i1} \le c_2 \text{ or } y_{i2} > c_2 \text{ and } y_{i1} > c_2 \\ -1, & \text{if either } c_1 \le y_{i2} \le c_2 \text{ and } y_{i1} < c_1 \text{or } y_{i2} > c \\ & \text{and } y_{i1} < c_1 \text{or } y_{i2} > c_2 \text{ and } c_1 \le y_{i1} \le c_2 \end{cases} \quad \dots 28$$

For $i = 1,2,\dots,n.$

Note that this specification may be depicted in a 3 x 3 table if we let n_{ij} be the number of paired case and control subjects in the (i, j)th case – control response classification for $i = 1,2,3$ and $j = 1,2,3.$

Specifications similar to Equation 28 can also be easily developed for more than three quantitative response categories if of interest. Now to use Equation 20 to analyse these data, we would again simply define π^+, π^0, π^- and W as in Equations 6 – 8. Then data analysis proceeds as usual.

B. *Illustrative Example 2*

A medical researcher is interested in knowing the relationship between heart disease and low density Lipo-Protein Levels (LPL). Using a random sample of 36 non-heart disease patients and another random sample of 36 heart disease patients, she paired each non heart disease patient with a heart disease patient matched in age, gender, body weight and occupation and then measured the LPL of each subject in the pair. The results are presented in Table 3

Table 3: LPL levels of Paired Samples of Patients in a Clinical Trial

S/N	Paired LPL levels	Scores (u_i)
1	(1.97,4.14)	0
2	(3.70,1.57)	-1
3	(5.40,5.60)	0
4	(2.60,5.10)	1
5	(3.10,1.50)	-1
6	(1.48,4.56)	1
7	(1.69,1.70)	0
8	(4.97,1.21)	-1
9	(2.34,2.51)	0
10	(3.95,1.55)	-1
11	(4.84,1.25)	-1
12	(4.65,4.59)	0
13	(1.29,1.37)	0
14	(1.15,6.24)	1
15	(5.41,1.20)	-1
16	(4.62,1.25)	-1
17	(2.02,1.53)	-1
18	(1.45,1.30)	0
19	(5.31,1.07)	-1
20	(5.18,4.37)	-1
21	(4.52,5.38)	1
22	(5.03,3.34)	-1
23	(5.21,4.55)	0
24	(4.74,5.59)	0
25	(3.76,3.96)	0
26	(5.21,3.50)	-1
27	(5.09,4.66)	0
28	(1.97,4.14)	0
29	(2.60,5.10)	1
30	(1.69,1.70)	0
31	(3.95,1.55)	-1
32	(1.29,1.37)	0
33	(4.62,1.25)	-1
34	(5.31,1.07)	-1
35	(5.03,3.34)	-1
36	(3.76,3.96)	0

LPL Normal range (1.68, 4.53)

Applying the specification of Equation 28 to the LPL levels of Table 3 with $c_1 = 1.68$, the lowest and 4.53 the highest normal values respectively we obtain the corresponding scores $u_{i\,of}$ 1_s, 0_s and -1_s shown in the 3rd column of this table.

Thus we have $f^+ = 5$, $f^0 = 15$ and $f^- = 16$. Hence, we have from Equations 12 – 15 that

$$\hat{\pi}^+ = \frac{5}{36} = 0.139;\ \hat{\pi}^0 = \frac{15}{36} = 0.417\ and\ \hat{\pi}^- = \frac{16}{36} = 0.444$$

From Equation 18, we have that the estimated variance of W is

$$Var(W) = 36(0.139 + 0.444) - \frac{(-11)^2}{36} = 20.988 - 3.361 = 17.627$$

The null hypothesis to be tested is that heart disease patients and non-heart disease patients do not differ in their LPL which is equivalent to testing

$$H_0 : \hat{\pi}^+ - \hat{\pi}^- = 0\ versus\ H_1 : \hat{\pi}^+ - \hat{\pi}^- \neq 0$$

Using the test statistic of equation 20 or 23, we have that

$$x^2 = \frac{(-11)^2}{17.627} = \frac{121}{17.627} = 6.864\ (p\ value = 0.0391)$$

which with c - 1 = 3 - 1 = 2 degrees of freedom is statistically significant at the 5 percent level $\left(x^2_{0.95,2} = 5.991\right)$

We may therefore conclude that heart disease patients and non- heart disease patients do infact differ in their LPL.

The data of Table 2 may infact be represented by a 3 x 3 table and following the specifications of Equation 28 with $c_1 = 1.68$ *and* $c_2 = 4.53$ to aid in clearer analysis as in Table 3.

Table 4: Distribution of Scores u_i of Matched pairs of case and control subjects of Table 3

	Control (T_1) *Scores*			
Case(T_2) *Scores*	Below Normal ($y_{i1} < 1.68$)	Normal ($1.68 \leq y_{i1} \leq .53$)	Above Normal ($y_{i1} >$ 4.53)	Total
Below Normal ($y_{i2} < 1.68$)	4	0	2	6
Normal ($1.68 \leq y_{i2} \leq 4.53$)	5	6	3	14
Above Normal ($y_{i2} > 4.53$)	7	4	5	16
Total	16	10	10	36

To re-analyse these data consistent with the generalized method, we have from Equation 13 that

$$\hat{\pi}^+ = \frac{0+2+3}{36} = \frac{5}{36} = 0.139$$

From Equation 14, we have that

$$\hat{\pi}^0 = \frac{4+6+5}{36} = \frac{15}{36} = 0.417$$

And from Equation 15, we have that

$$\hat{\pi}^- = \frac{5+7+4}{36} = \frac{16}{36} = 0.444$$

These are the same results obtained earlier using the scores in Table 3. We would therefore obtain the same values of W (-11) and chi-square (6.864) and arrive at the same conclusions. Hence, the present example illustrates how to analyse matched quantitative test scores without first converting them into frequency data.

The data of Example 2 as presented in Table 4 may also be analysed using the Stuart-Maxwell test. However as already pointed out, the Stuart/Maxwell test statistic is almost as powerful as the test statistic used in the proposed method presented here when the two methods are used with data of equal sample sizes

III. Summary and Conclusion

We have in this paper presented and discussed a generalisable statistical method for the analysis of three responses or outcomes in case - control studies, including situations in which the data being analysed are either quantitative or qualitative frequency data.

Test statistics are developed for testing the statistical significance of differences between responses.

The proposed methods are illustrated with sample data and shown to be more powerful than the usual Stuart/Maxwell test when the two methods are equally applicable to a set of data.

REFERENCES

[1] Box G. E. P. and Cox D. R: "An analysis of transformations. Journal of the Royal Statistical Society (B), 1964, 26, 211 – 252.

[2] Everitt H. B. S: The Analysis of Contingency Tables. London: Chapman and Hall, 1977.

[3] Fleiss, J. L. (1981): Statistical Methods for Rates and Proportions (Second Edition), New York. Wiley 1981.

[4] Gibbons, J. D. (1971): Nonparametric Statistical Inference. McGraw-Hill Book Company, New York.

[5] Maxwell, A. E: Comparing the classification of subjects by two independent judges. British Journal of Psychiatry, 1970, 116, 651 – 655.

[6] McNemar Q: Note on the sampling error of the difference between correlated proportions or percentages. Psychometrika 1947, 12, 153 – 157.

[7] Miettinen O. S: Individual Matching with multiple controls in the case of all – or – more response. Biometrics, 1969, 22, 339 – 355.

[8] Robertson C, Boyle P, Hseih C, Macfarlane, G. J, and Maisonnesuve P: Some Statistical considerations in the analysis of case-control studies when the exposure variables are continuous measurements. Epidemiology, 1974, 5, 164-170.

[9] Schlesselman J. J. (1992): Case-Control studies. Oxford University Press, New York.

[10] Sheskin D. J. (2000): Handbook of Parametric and Non-parametric Statistical procedures (Second Edition). Boca Raton Chapman and Hall, 491 – 508.

[11] Somes, G: McNemar Test Encyclopaedia of Statistical Sciences, S. Kotz and N. Johnson eds, 1983, 5, 361 – 363. New York.Wiley.

[12] Stuart, A. A: A test for homogeneity of the marginal distributions in a two – way classification. Biometrika, 1955, 42, 412 – 416.
[13] Zhao L. P and Kolonel L. N: "Efficiency Loss from Categorizing quantitative exposure into qualitative exposures in case-control studies". American Journal of Epidemiology, 136, 464 – 474. 1992.