This research was supported by National Institutes of Health, Institute of General Medical Sciences Grants GM-70004-01 and GM-12868-08.

THE USE OF NON-PARAMETRIC METHODS IN THE STATISTICAL ANALYSIS OF THE TWO-PERIOD CHANGE-OVER DESIGN

by

Gary G. Koch

Department of Biostatistics
University of North Carolina at Chapel Hill

Institute of Statistics Mimeo Series No. 787

November 1971

The Use Of Non-Parametric Methods In The Statistical Analysis Of The Two-Period Change-Over Design

by

Gary G. Koch

Department of Biostatistics, University of North Carolina

Chapel Hill, North Carolina, 27514, U.S.A.

Summary

The two-period change-over design is often used in clinical trials in which subjects serve as their own controls. This paper is concerned with the statistical analysis of data arising from such subjects when assumptions like variance homogeneity and normality do not necessarily apply. Test procedures for hypotheses concerning direct effects and residual effects of treatments and period effects are formulated in terms of Wilcoxon statistics as calculated on appropriate within subject linear functions of the observations. Thus, they may be readily applied to small sample data.

1. A TWO-PERIOD CHANGE-OVER EXPERIMENT

Let us consider the data from an experiment undertaken at the Dental Research Center, University of North Carolina and reproduced here with the kind permission of Dr. William J. Waddell and Dr. Eugene Howden. The design was as follows: n=10 children were randomly assigned to two groups, each of size $n_1=n_2=5$. All of the children were observed at two different time periods which were separated by one week. The two treatments which were applied were

1. The child first drank 100 ml. of grapefruit juice followed by an elixir of Pentobarbital; this treatment has been called G.

2. The child first drank 100 ml. of water followed by an elixir of Pentobarbital; this treatment has been called H.

For both G and H, the amount of the Pentobarbital elixir given to a specific subject was proportional to the child's body weight. The subjects in Group 1 received G at the first time period and H at the second time period while those in Group 2 received H first and then G. The data resulting from the experiment were measurements on the amount of drug in a 10 ml. sample of blood taken 15 minutes after the elixir was administered and were expressed in μg/ml. The data are displayed in the 3rd and 4th columns of Table 1.

Table 1

Sequence	Subject	Day 1	Day 2	Sum	Difference	Cross-over Difference
G:H	C11	1.75	.55	2.30	1.20	1.20
G:H	C12	.30	1.05	1.35	- .75	- .75
G:H	C13	.35	.63	.98	- .28	- .28
G:H	C14	.20	1.55	1.75	-1.35	-1.35
G:H	C15	.30	8.20	8.50	-7.90	-7.90
H:G	C21	7.20	.35	7.55	6.85	-6.85
H:G	C22	7.10	1.55	8.65	5.55	-5.55
H:G	C23	.75	.25	1.00	.50	- .50
H:G	C24	2.15	.35	2.50	1.80	-1.80
H:G	C25	3.35	1.50	4.85	1.85	-1.85

2. THE MODEL

Following Grizzle [1965], we will assume that an adequate model for the data is

$$y_{ijk} = \mu + b_{ij} + \pi_k + \phi_{\ell} + \lambda_{\ell} + e_{ijk} \qquad (1)$$

$$j=1,2,\ldots,n_i;\quad i=1,2;\quad k=1,2;\quad \ell=1,2;$$

where μ is a general mean, b_{ij} is the random effect for the j-th subject within the i-th sequence, π_k is the effect of the k-th period, ϕ_{ℓ} is the direct effect of the ℓ-th drug, λ_{ℓ} is the residual effect of the ℓ-th drug, and e_{ijk} reflects random error in the measurement of the response. For the case in which the $\{b_{ij}\}$ and the $\{e_{ijk}\}$ are each normally distributed as $N(0,\sigma_b^2)$ and $N(0,\sigma_e^2)$ respectively and are mutually independent, Grizzle discusses tests of hypotheses pertaining to the direct effects, residual effects, and period effects. In this paper, we shall be concerned with these same questions except normality will not be assumed.

3. THE ANALYSIS

In this section, we shall describe certain non-parametric methods based on ranks which can be applied to the data in Table 1. These procedures are similar to those described in Koch [1970] for a complex split-plot experiment; hence, they are formulated by identifying within subject functions of the data to which the null distribution theory of an appropriate non-parametric rank test is applicable; here, suitable modifications of the Wilcoxon statistic are used. These tests are pertinint to the data in Table 1 because of the different patterns of variability at different factor combinations and because of the presence of certain extreme valued responses.

3.1. Testing the equality of residual effects

From (1), it follows that the sum of the two observations on the same subject is given by

$$y_{ij1} + y_{ij2} = 2(\mu + b_{ij}) + (\pi_1 + \pi_2) + (\phi_1 + \phi_2) + \lambda_i + (e_{ij1} + e_{ij2}) \qquad (2)$$

where λ_i represents the residual effect of the i-th drug in the sequence ii'. Hence, under the hypothesis of no residual effects

$$H_{o\lambda}: \lambda_1 = \lambda_2 \quad , \qquad (3)$$

the within subject sums satisfy the same model for the subjects in the two different sequences. As a result, a non-parametric statistic for testing (3) is obtained by ranking the 10 sums and adding the ranks in the smaller sample; i.e., applying the Wilcoxon test to the sums. The appropriate sums are displayed in Column 5 of Table 1 and the appropriate ranks of sums are displayed in Column 5 of Table 2. The statistical significance of the resulting rank sum T=22 may be determined by referring to exact tables of the Wilcoxon statistic; see Owen [1962]. Hence, residual effects are not significant.

Table 2

Sequence	Subject	Day 1	Day 2	Sum	Difference	Cross-over Difference
G:H	C11	6	4	5	6	10
G:H	C12	2.5	6	3	3	7
G:H	C13	4	5	1	4	9
G:H	C14	1	8.5	4	2	6
G:H	C15	2.5	10	9	1	1
H:G	C21	10	2.5	8	10	2
H:G	C22	9	8.5	10	9	3
H:G	C23	5	1	2	5	8
H:G	C24	7	2.5	6	7	5
H:G	C25	8	7	7	8	4

3.2. Testing the equality of direct effects when residual effects are absent.

If there are no residual effects in the sense of (3), then (1) implies that the difference between the two observations on the same subject satisfy the model

$$y_{ij1} - y_{ij2} = (\pi_1 - \pi_2) + (-1)^{i+1}(\phi_1 - \phi_2) + (e_{ij1} - e_{ij2}) \qquad (4)$$

where $(-1)^{i+1}=1$ for the sequence G:H and $(-1)^{i+1}= -1$ for the sequence H:G. Hence, under the hypothesis of no direct effects

$$H_{o\phi}: \phi_1 = \phi_2 , \qquad (5)$$

the within subject differences satisfy the same model for subjects in the two respective sequences. As a result, a non-parametric statistic for testing (5) is obtained by ranking the 10 differences and adding the ranks in the smaller sample; i.e., applying the Wilcoxon test to the differences. The appropriate differences are displayed in Column 6 of Table 1 and the appropriate ranks of differences are displayed in Column 6 of Table 2. The statistical significance of the resulting rank sum T=16 may be determined by referring to exact tables of the Wilcoxon statistic; see Owens [1962]. Hence, direct effects are significant at the α=.05 level.

3.3. Testing the equality of period effects when residual effects are absent.

If there are no residual effects in the sense of (3), then differences between the observations on the same subject satisfy the model (4). Let us consider the set of cross-over differences obtained by using $(y_{ij1} - y_{ij2})$ for the sequence G:H, but the differences $(y_{ij2} - y_{ij1})$ for the sequence H:G. Under the hypothesis of no period effects

$$H_{o\pi}: \pi_1 = \pi_2 \qquad (6)$$

the within subject cross-over differences satisfy the same model for subjects in the two respective sequences. As a result, a non-parametric statistic for testing (6) is obtained by ranking the 10 cross-over differences and adding the ranks in the smaller sample; i.e., applying the Wilcoxon test to the cross-over differences. The appropriate cross-over differences are displayed in Column 7 of Table 1 and the corresponding ranks are in Column 7 of Table 2. The statistical significance of the resulting rank sum T=33 may be determined by referring to exact tables of the Wilcoxon statistic. Hence, period effects are not significant.

3.4. Testing the equality of direct effects and residual effects simultaneously.

Another hypothesis of interest is

$$H_{o\phi\lambda}: \phi_1 = \phi_2, \lambda_1 = \lambda_2. \qquad (7)$$

When (7) holds the bivariate response (y_{ij1}, y_{ij2}) has the same distribution for subjects in the two sequences. Hence, an appropriate non-parametric test is the bivariate Wilcoxon statistic (see Chatterjee and Sen [1964], or Koch [1970]). The appropriate ranks for the responses within each period are displayed in Columns 3 and 4 of Table 2 (where mid-ranks have been used in case of ties). The resulting normalized statistic L=5.97; under the hypothesis (7), L has approximately the chi-square distribution with D.F. = 2. In very small samples, the statistical significance of L can be determined by working out the permutation distribution of L under the randomization model associated with all possible assignments of subjects to the sequence groups. In this case, there are $\binom{10}{5} = 252$ such assignments of which the one observed here is the most extreme. Since the two-tailed conditional randomization model probability for an event at least as extreme as observed here is $(2/252) < (0.01)$, the data are not consistent with (7). Other aspects of this test have been recently discussed by Bhattacharyya, Johnson, and Neave [1971].

It is appropriate to note here that if the residual effects cannot be deleted from the model, then the test procedures described in sections 3.2 and 3.3 are no longer valid. The hypothesis (5) for equality of direct effects can be tested by applying the Wilcoxon statistic to the data observed during the first period and ignoring the data observed during the second period. For the example here, the resulting Wilcoxon statistic

T=16, which is significant at the $\alpha = .05$ level. This result is analogous to that quoted by Grizzle [1965] for the case of unequal residual effects when the $\{b_{ij}\}$ and $\{e_{ijk}\}$ were normally distributed. He also noted that the difference between the two period effects $\pi_1 - \pi_2$ is not estimable in this situation. Hence, no non-parametric test is given for $H_{0\pi}$ under these conditions.

Finally, if one can assume that both residual effects are equal and both periods' effects are equal, then the hypothesis (5) of equality of direct effects can be investigated by applying a sign test (or Wilcoxon signed rank test) to the cross-over differences. Hence, since H exceeds G for 9 out of 10 subjects, we conclude again that direct effects are significant at the $\alpha=.05$ level. This test is more general than the previous ones described since it requires only an ordinal scale for data obtained from the same subject whereas the other tests considered require a measurement scale in which sums and differences (or alternatively products and quotients if the logarithms of the observations are analyzed) have a meaningful interpretation and can be ranked. On the other hand, this approach is not valid with respect to (5) if either residual effects or period effects are present. Since the bivariate Wilcoxon test only requires an ordinal scale for ranking across subjects, it should be used instead of the sign test for general situations in which residual and/or period effects are unequal and within subject linear functions are invalid. The principal disadvantage which this procedure has is that the tests of direct and residual effects are confounded. However, if only one of them is significant, L can be partitioned into components which reflect this. Thus,

non-parametric methods can be used to interpret data from two-period change-over designs under a broad range of conditions.

References

1. Bhattacharyya, G. K., Johnson, R. A. and Neave, H. R. [1971]. A comparative power study of the bivariate rank sum test and T^2, Technometrics, 13, 191-198.

2. Chatterjee, S. K. and Sen, P. K. [1964]. Non-parametric tests for the bivariate two-sample location problem, Calcutta Stat. Assoc. Bulletin, 13, 18-58.

3. Grizzle, J. E. [1965]. The two-period change-over design and its uses in clinical trials. Biometrics, 21, 467-480.

4. Koch, G. G. [1970]. The use of non-parametric methods in the statistical analysis of a complex split plot experiment. Biometrics, 26, 105-128.

5. Koch, G. G. and Sen, P. K. [1968]. Some aspects of the statistical analysis of the mixed model, Biometrics, 24, 27-48.

6. Owen, D. B. [1962]. Handbook of Statistical Tables. Addison-Wesley, Reading, Massachusetts.

7. Wilcoxon, F. [1945]. Individual comparisons by rank methods. Biometrics, 1, 80-83.

8. Wilcoxon, F. [1949]. Some Rapid Approximate Statistical Procedures. American Cyanamid Co., Stamford, Connecticut.