*Brit. J. Psychiat.* (1969), 115, 1379–82

# A Method for Analysing Assessments of Symptom Change

By ALISTAIR E. PHILIP

In studies which attempt to assess the efficacy of some new treatment it is customary to make a formal assessment of the relevant behaviour or symptomatology using standardized inventories, checklists, symptom rating scales or *ad hoc* ratings of variables considered to be important by the clinician-researcher. Ratings made at the beginning and end of treatment are compared for groups of patients using improvement scores, arbitrary cut-off points and other devices aimed at circumventing the statistical problems arising from the non-normal distributions of most rating scales. Present practice favours the use of some nonparametric statistic in these comparisons. The aim of this paper is to present a method which facilitates the analysis of ratings made on more than two occasions, allowing a trend analysis to be carried out without making assumptions about the distribution of scores. The method also allows the clinician-researcher to make a statistically-based decision regarding the efficacy of the treatment in question for individual patients.

A basic assumption made here is that the items comprising a rating scale have been selected because of their assumed relevance to the behaviour under study so that it is legitimate to consider each such item in its own right rather than to sum responses into a total score. If a patient has been rated on ten items on five occasions it is possible to consider each item in turn and transform the ratings given on that item into ranks. The occasion with the highest rating score is ranked 1, the occasion with the next highest rating score is ranked 2 and so forth; where two or more occasions have the same rating score they are given the same tied ranks. Transforming the data into ranks permits the use of one of the useful but little known methods of nonparametric trend analysis devised by Ferguson (1965).

Ferguson's method makes use of the statistic S which is employed in the calculation of Kendall's *tau* (Kendall, 1948) and can be regarded as a nonparametric analogue of analysis of variance methods of trend analysis using orthogonal polynomials. The method used here is superior to the nonparametric analysis of variance methods of Friedman (Siegel, 1956) and Page (1963) in two respects. Account is taken of the non-independence of scores in the situation where an individual is tested a number of times using the same measures, and non-linear trends in the data can be evaluated. Not all patients show an orderly drop in symptoms from one rating occasion to the next; some patients show marked improvement at first but are rated less well on later occasions, others show a zig-zag course during treatment, and a few show worsening of their condition. Ferguson's technique allows the clinician-researcher to test the significance of these monotonic, bitonic and other trends, thus adding to his knowledge of the treatment being evaluated.

*Example*

Table I shows the ratings made on an individual on the ten items of a scale on four occasions. For each item of the scale ratings can range from 0 to 4, high ratings indicating much pathology.

The right hand side of Table I shows the rating scores transformed into ranks, each item being ranked in turn so that the occasion having the highest score is ranked 1 and that with the lowest score ranked 4. In the example four pairs of occasions have the same score and are given tied ranks.

An inspection of the total rating scores and the corresponding sums of ranks suggests that while symptoms have shown an overall decrease there is a possibility that some non-linear effect is present (shown by the slight zig-zag of the scores).

TABLE I

*Rating scores and ranks for items on each occasion*

| | Rating scores | | | | | Ranks | | | |
|---|---|---|---|---|---|---|---|---|---|
| | Occasion | | | | | Occasion | | | |
| Item | I | II | III | IV | | I | II | III | IV |
| 1 | 4 | 3 | 1 | 0 | | 1 | 2 | 3 | 4 |
| 2 | 0 | 2 | 0 | 1 | | 3·5 | 1 | 3·5 | 2 |
| 3 | 0 | 0 | 2 | 1 | | 3·5 | 3·5 | 1 | 2 |
| 4 | 4 | 1 | 0 | 0 | | 1 | 2 | 3·5 | 3·5 |
| 5 | 3 | 4 | 2 | 2 | | 2 | 1 | 3·5 | 3·5 |
| 6 | 4 | 2 | 0 | 1 | | 1 | 2 | 4 | 3 |
| 7 | 3 | 4 | 0 | 1 | | 2 | 1 | 4 | 3 |
| 8 | 4 | 2 | 0 | 1 | | 1 | 2 | 4 | 3 |
| 9 | 4 | 3 | 2 | 0 | | 1 | 2 | 3 | 4 |
| 10 | 3 | 4 | 0 | 1 | | 2 | 1 | 4 | 3 |
| Total | 29 | 25 | 7 | 8 | Sum of ranks | 18 | 17·5 | 33·5 | 31 |

TABLE II

*Steps in the calculation of probability values*

| Occasions | S | $S_1$ | $S_2$ | $S_3$ |
|---|---|---|---|---|
| I and II .. .. .. | 1 | 1 | −1 | 1 |
| I and III .. .. .. | 7 | 1 | −1 | −1 |
| I and IV .. .. .. | 6 | 1 | 0 | 1 |
| II and III .. .. .. | 8 | 1 | 0 | −1 |
| II and IV .. .. .. | 8 | 1 | 1 | −1 |
| III and IV .. .. .. | −2 | 1 | 1 | 1 |
| | | $\Sigma S_1$ | $\Sigma S_2$ | $\Sigma S_3$ |
| | | 28 | −2 | −18 |
| | $\lvert \Sigma S \rvert - 1$ | 27 | −1 | −17 |
| | $(\sigma \Sigma S)^2$ | 82·67 | 64·00 | 82·67 |
| | $\sigma \Sigma S$ | 9·09 | 8·00 | 9·09 |
| | z | 2·97 | 0·37 | 1·87 |
| | p | 0·003 | 0·711 | 0·061 |

To arrive at probability values for each trend from the ranks in Table I requires three steps, (a) the calculation of the statistic S, (b) the calculation of its standard deviation $\sigma\Sigma S$ and (c) the calculation of the normal deviate z. Table II shows the results of these calculations, which are now described in detail.

*Step* 1

Compare the ranks of each occasion with every other occasion two at a time, giving a weight of +1 for pairs which occur in their natural order, a weight of 0 for pairs which tie and a weight of −1 for pairs which occur in inverse order. When occasions (columns) I and II are compared, the following weights are obtained; +1, −1, 0, +1, −1, +1, −1, +1, +1, −1. When summed, these weights give for occasions I and II a value which is the first entry in column S. Similar calculations for each pair of occasions yield the S scores given in Table II.

A similar procedure is carried out to find the values $S_1$, $S_2$ and $S_3$ which are weights calculated from comparison of the ranks of orthogonal polynomials of the first (linear or monotonic), second (quartic or bitonic) and third (cubic or tritonic) order. These ranks are

https://doi.org/10.1192/bjp.115.529.1379 Published online by Cambridge University Press

derived from the tables of polynomials found in *Biometrika Tables for Statisticians, Vol.* 1 (Pearson and Hartley, 1966). The first order polynomial has for four occasions the ranks 1, 2, 3 and 4 so that when ranks for pairs of occasions are compared weights of +1 are found in each case. These weights form column $S_1$ of Table II. The second order polynomial has ranks of 3·5, 1·5, 1·5 and 3·5; when the ranks for pairs of occasions are compared the weights are as shown in column $S_2$ of Table II. The third order polynomial has ranks of 2, 4, 1 and 3 and the weights produced by comparison of pairs of occasions form column $S_3$ of Table II.

The values in column S are multiplied by the weights given in columns $S_1$, $S_2$ and $S_3$ to give the values $\Sigma S_1$, $\Sigma S_2$ and $\Sigma S_3$. These values are then corrected for continuity by subtracting unity from each, ignoring algebraic sign.

*Step* 2

The sampling variance of S is defined by the formula

$$(\sigma \Sigma S)^2 = \frac{Nk(k-1)(2k+5)}{18}$$

where N is the number of items and k is the number of occasions. This value must be corrected when ties occur in the experimental variable or in the ranks of the polynomials. Correction for the first situation is brought about by subtracting from the original formula the value $\frac{\Sigma t\,(t-1)\,(2t+5)}{18}$, where t is the number of cases in each tie. For tied pairs, each $t=2$, for triplets of ties, each $t = 3$ and so on. When ties occur only in the ranks of the polynomials, the original formula is corrected by subtracting the value rN, where r is the number of tied pairs of occasions in the ranks for polynomials. Where both the experimental and polynomial ranks have tied values the formula reads

$$(\sigma \Sigma S)^2 = \frac{Nk(k-1)(2k+5)}{18} - \frac{\Sigma t(t-1)(2t+5)}{18} - rN + \frac{2r\Sigma t\,(t-1)}{2k(k-1)}.$$

The square roots of these variances yield the standard deviations for the values $\Sigma S_1$, $\Sigma S_2$ and $\Sigma S_3$.

In the present example the experimental data have four tied pairs while the ranks of the second order polynomial show two tied pairs indicated by weights of 0 in column $S_2$. The variances for $\Sigma S_1$ and $\Sigma S_3$ need to be corrected only for ties in the experimental variable while for $\Sigma S_2$ the complete correction formula must be used.

*Step* 3

The values $\Sigma S_1$, $\Sigma S_2$ and $\Sigma S_3$, corrected for continuity, are divided by their respective standard deviations to give values of the normal deviate z. The probability value associated with each z can then be determined. In the present case it can be seen that there is a highly significant monotonic trend in the data ($p = 0{\cdot}003$) but none of the higher order trends reach significance.

When using standard rating scales it is frequently found that some aspects of behaviour covered by the scales are not applicable to some patients. Similarly other aspects of behaviour remain unchanged throughout all rating occasions. These situations produce numerous ties in the ranked data but it can be shown that these do not affect the results.

In the case of items which are never endorsed as being present or are given the same rating throughout, the magnitude of the tie (t) which occurs (triplet, quartet, etc.) will equal the number of test occasions (k). Thus $\frac{\Sigma t(t-1)(2t+5)}{18} = \frac{\Sigma k(k-1)(2k+5)}{18}$, so that in calculating $(\sigma \Sigma S)^2$, ties of magnitude $t = k$ contribute to the first term and the second term in equal amounts and therefore do not influence the size of $(\sigma \Sigma S)^2$. Similarly for such ties the fourth term, $\frac{2r\Sigma t(t-1)}{2k(k-1)}$ can be expressed as $r\,\frac{2\Sigma t(t-1)}{k(k-12)} = r\,\frac{2\Sigma k(k-1)}{2k(k-1)}$ so that the contribution given by these items to the third term (rN) is removed by the contribution to the fourth term. In short, items which are never endorsed and items which have the same

endorsement throughout do not contribute to $(\sigma \Sigma S)^2$ and no correction or special calculation is needed.

## Discussion

Since clinicians are interested in individual cases, studies on the efficacy of some treatment which report findings in terms of groups means are of less value than studies in which the amount and pattern of response to treatment can be objectively stated for each individual. The present method allows such statements about individuals to be made.

Most studies of treatment report differences between initial and final testing occasions, although assessments may have been carried out on several occasions during the trial period. It is hard to analyse repeated measurements of variables which meet few of the assumptions necessary for the application of parametric techniques, and Ferguson's monograph is the first presentation of a nonparametric method which is adequate to the task. The technique is not brief, but it makes use of all the ratings gathered on a patient and is therefore more efficient than other methods which use only part of the data.

The method is very flexible, since the number of items or rating occasions can be increased at will and if preferred a trend analysis of groups can be carried out by substituting person scores for item scores. In the individual case it might be that the customary levels of significance are too stringent. Since the method yields exact probabilities, the individual clinician-researcher is at liberty to choose his own cut-off point in accepting the efficacy of a form of treatment. Useful results could not be expected in a situation where an individual had been assessed on many occasions using a rating scale with only one or two rating points. Similarly, little could be expected if the test instrument used was not appropriate to the behaviour being studied. Like any statistical procedure, the method will not make a silk purse out of a sow's ear.

## References


Ferguson, G. A. (1965). *Nonparametric Trend Analysis*. Montreal: McGill University Press.

Kendall, M. G. (1948). *Rank Correlation Methods*. London: Griffin.

Page, E. B. (1963). 'Ordered hypotheses for multiple treatments: a significance test for linear ranks.' *J. Amer. statist. Ass.*, **58**, 216–30.

Pearson, E. S., and Hartley, H. O. (1966). *Biometrika Tables for Statisticians, Vol. 1*. Cambridge: Cambridge University Press.

Siegel, S. (1956). *Nonparametric Statistics for the Behavioural Sciences*. New York: McGraw-Hill.



Alistair E. Philip, M.A., Ph.D., Dip. Clin. Psychol., *Scientific Staff, Medical Research Council, Unit for Research on the Epidemiology of Psychiatric Illness, Edinburgh University Department of Psychiatry, Edinburgh, 10*



(*Received 20 September 1968*)