STATISTICS IN MEDICINE, VOL. 8, 1401–1411 (1989)

A MEASURE OF SIMILARITY FOR RESPONSE CURVES BASED ON RANKS

GABI SCHULGEN

Institute of Medical Biometry and Informatics, University of Freiburg, Stefan-Meier-Str. 26, D-7800 Freiburg, F.R.G.

SUMMARY

A measure of similarity for response curves is presented and its potential use is discussed. The distribution of a suitable test statistic for testing the independence of the course of two curves is derived. The method proposed is compared with other proposals in the literature for the analysis of paired response curves. A study on the quality of life of patients with acute heart failure is used as an example.

KEY WORDS Quality of life Heart disease Response curve Kappa Similarity Ranked observations

1. THE STUDY

The study that will be used as an example is an observational study on the subjective well-being of patients with acute heart failure at the Gemeinschaftskrankenhaus Herdecke.[1,2] The aim of the study was to evaluate whether the subjective well-being, measured by a questionnaire, could serve as a criterion for determining the optimal dose of digitalis for an individual patient. The study has to be regarded as a first attempt by the hospital to include subjective assessment of well-being into treatment decisions in a quantitative way.

Besides the evaluation of other clinical criteria, the patients' intensity of dyspnea, level of oedema and degree of being bedridden (all measured on a seven-point ordinal scale) were assessed. In addition, the patients were asked to fill in a questionnaire comprising twelve questions about different aspects of health. Simultaneously, the physicians had to assess the patients' well-being by means of the reduced questionnaire with six questions shown in Table I. The last question will not be considered further in this paper.

Both patients and physicians were interviewed the day before the start of treatment ('day zero') and four, eight and fourteen days after the start of treatment.

Of the fifty patients included in the study, only twenty participated in the first interview, whereas the rating by the physicians was available before the start of the treatment for all but one patient. The main reason for non-response was the patients' bad physical condition before the start of treatment. The non-responders differed significantly from the responders with respect to their scores for dyspnea. Therefore, it was decided to perform the analysis mainly on the basis of the physician's rating of the patient's well-being, knowing that the agreement between these two assessments was relatively low and that, in general, one should rather rely on the patient's self-rating which clearly is a more valid measure of their individual subjective well-being.

To analyse the structure of the questionnaire and to reduce the dimension of the data, factor analytic techniques were applied.[3] The resulting one-factor model for the physician's rating at 'day

0277–6715/89/111401–11$05.50 *Received September* 1988

Table I. Physicians' questionnaire

1.	Today the patient seems		
	very energetic	– – – – – – –	very lethargic
		7 6 5 4 3 2 1	
2.	Today the patient seems		
	very fresh	– – – – – – –	very tired
		7 6 5 4 3 2 1	
3.	Today the patient's physical condition is		
	very good	– – – – – – –	very bad
		7 6 5 4 3 2 1	
4.	Today the patient's ability to concentrate is		
	very good	– – – – – – –	very bad
		7 6 5 4 3 2 1	
5.	The general impression is: today the patient feels		
	very fine	– – – – – – –	very bad
		7 6 5 4 3 2 1	
6.	Compared to the time before the start of the treatment the patient feels now		
	much better	– – – – – – –	much worse
		7 6 5 4 3 2 1	
	with respect to the physical symptoms		

zero' with all five items having high loadings on this factor was used for the creation of a quality of life index (QI). The weights of the different items were the standardized factor beta loadings of the items. Under the assumption that the constellation of items and factor does not change completely over time, the estimated weights from 'day zero' can be used to calculate QI for all time points. This assumption was checked by performing a factor analysis for the rating at 'day fourteen', in which a similar constellation of the items and the factor was observed.

2. A MEASURE OF SIMILARITY FOR INDIVIDUAL RESPONSE CURVES

An important aspect in analysing quality of life data is the interrelationship between the measure of the subjective well-being and the clinical course, that is the extent to which individual well-being is influenced by the intensity of special physical symptoms of the disease. In addition, the quality of life measure can be validated on criteria like the physical symptoms which are supposed closely related to the patient's well-being. In most situations, especially in quality of life studies, it is indicated not to measure these variables once only, but to consider their behaviour over time. To meet these requirements in the analysis of the relationships among the variables, the course of the variables over time should be compared in a comprehensive way and not only at distinct time points. Therefore a measure is needed which describes the similarity of two response curves on the individual level for a sample of bivariate response curves.

In the present study, the relationship between the clinical course on the one hand, represented by the variables 'dyspnea', 'oedema' and 'degree of being bedridden', and subjective well-being on the other, measured by QI, should be analysed. This situation can be summarized formally in the following way.

For each of the N patients in the study and any one clinical variable, two response curves are measured; that is a sample of N bivariate response curves (X_i, Y_i) $(i=1, 2, \ldots, N)$ is obtained. In this notation X_i denotes the course of the clinical variable for the ith patient with

measurements at the T distinct time points, these being denoted by x_{it} $(t=1, 2, \ldots, T)$; that is, $X_i=(x_{i1}, x_{i2}, \ldots, x_{iT})$. Y_i denotes the course of QI for the ith patient with the single measurements denoted by y_{it}, that is $Y_i=(y_{i1}, y_{i2}, \ldots, y_{iT})$.

In the present study all the variables under consideration are measured on an ordinal scale only; therefore parametric methods for the analysis of repeated measurements could not be considered and a non-parametric approach is chosen. Because the level of the curve is of limited information when different psychological scales are compared, the rank transformation is used. Rank transformation in this context means that each response curve is characterized by its rank pattern, separately for each of the two variables:

$$(x_{i1}, \ldots, x_{iT}) \rightarrow [R(x_{i1}), \ldots, R(x_{iT})]$$
$$(y_{i1}, \ldots, y_{iT}) \rightarrow [R(y_{i1}), \ldots, R(y_{iT})].$$

The ranks are assigned within the individual response curves.[4,5] $R(x_{it})$ denotes the rank of the observation at time point t among all observations of variable X at the distinct time points for the ith patient, and $R(y_{it})$ is defined similarly for variable Y. In the presence of ties, midranks are assigned.

This transformation preserves information about the shape of the response curve, with the restriction that one can no longer distinguish between linear and higher order changes. This is not really a limitation in a situation with few categories and few time points.

A measure of similarity between two response curves is then derived by considering the difference of the two rank patterns,

$$D=\sum_{t=1}^{T} |R(x_t)-R(y_t)|.$$

That is, for each time point t the absolute difference in the ranks is calculated and this is then summed over all time points.

If there are no tied values within the curves, the random variable D can take on only even numbers between 0 and $\sum_{t=1}^{T} |2t-T-1|$. In the presence of ties in both variables, all integers between 0 and $\sum_{t=1}^{T} |2t-T-1|$ are possible. In our situation of four time points and ties present, D can take on all integer values between 0 and 8.

As is illustrated in Figure 1(a), with parallel response curves the absolute differences in ranks at the distinct time points are all zero; therefore D takes on the value 0, which means maximum similarity or 'positive dependence' of the variables. If the curves are opposed to each other, as is illustrated in Figure 1(b), with four time points the absolute differences in ranks at the distinct time points are 3, 1, 1, 3, and D takes on the value 8, its maximum value with four time points, which means 'negative dependence' of the variables. In general, as D increases the similarity of the curves decreases.

This measure could be standardized to lie between -1 and $+1$ by using

$$1-\left[2D \Big/ \left(\sum_{t=1}^{T} |2t-T-1|\right)\right],$$

where -1 indicates minimum similarity and $+1$ indicates maximum similarity. One can think of other measures of similarity for two response curves, for example a measure which is based on the

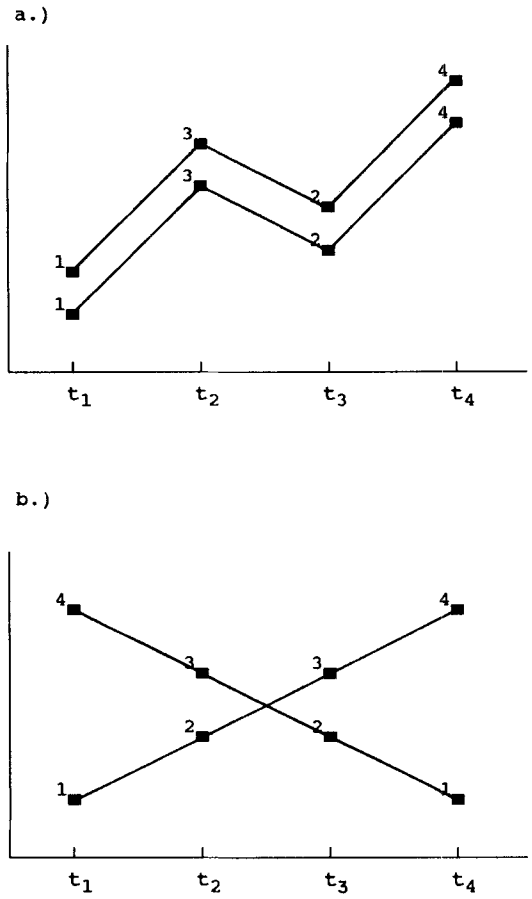

Figure 1. The measure of similarity with (a) parallel and (b) opposed response curves: the numbers beside the squares indicate the ranks of the observations within the individual response curve

sum of the quadratic differences in ranks:

$$S=\sum_{t=1}^{T}[R(x_t)-R(y_t)]^2.$$

The standardization

$$1-[6S/(T^3-T)]$$

would lead to the well-known Spearman rank correlation coefficient. Another example of a measure which uses a sign transformation instead of the rank transformation will be given in Section 5.

3. A TEST ON INDEPENDENCE AND A GLOBAL MEASURE OF SIMILARITY

So far a measure which characterizes the similarity of two response curves has been developed. The starting-point for these considerations was the question of how to judge the relationship

between two variables measured at a few prespecified time points in a sample of patients. The measure of similarity can be used to construct a formal test for the hypothesis of independence of the two variables. A suitable test criterion is the mean observed similarity among the response curves of the sample:

$$\bar{D}. = \sum_{i=1}^{N} D_i/N,$$

where

$$D_i = \sum_{t=1}^{T} |R(x_{it}) - R(y_{it})|$$

indicates the similarity of the curves in patient i.

In the test problem the hypothesis of independence will be rejected if the mean observed similarity is different from the similarity that would have been expected under the hypothesis of independence. To specify a critical value the distribution of the test statistic under this hypothesis is needed. For the derivation of the distribution of the test statistic the classical idea of a permutation test (randomization test) can be applied.[6] The particular sample of ranked bivariate response curves obtained is considered as just one of all possible samples having the same rank pattern of response curves within each of the two variables but not necessarily the same assignment of the curves to the patients. Therefore the test is conditional on the observed response curves within each of the two variables. Note that by the conditioning the assumption that all different response curves are equally likely is avoided. This would be a rather unrealistic assumption since there may be a time trend in one or both variables.

Since there are $N!$ possible ways to rearrange the N (not necessarily all different) observed response curves within one of the variables among the N individuals, there are $N!$ possible samples which are all equally likely under the hypothesis of independence. For each possible sample the value of the test statistic can be calculated. The distribution of this statistic conditional on the observed response curves can then be calculated. This is the permutation distribution of the test statistic. From this distribution a p-value can be derived and the test can be performed in the usual way.

When the number of observations N is large (or even only moderately large) the computation of the exact permutation distribution may be impractical. In such a case there are at least two possible approaches.[6] One possibility is to sample from the permutation distribution a sufficiently large number of possible values using random digits generated by a computer to determine random permutations of the N ranked response curves. From this 'simulated' permutation distribution the p-value of the test can be estimated. Another way to reach a decision in the test problem is to approximate the permutation distribution by a normal distribution. For this approximation the expectation and variance of the test statistic under the hypothesis are required; the formulae for this are given in the appendix.

In addition to this formal test of independence, it is of interest to describe the similarity between the response curves observed in the sample by a measure. A suitable global measure of similarity is a generalization of Cohen's weighted kappa coefficient,[7-9] where the mean observed similarity $\bar{D}.$ is adjusted for its expected value under the hypothesis of independence, denoted by $E_{H_0}(\bar{D}.)$. This leads to

$$\kappa = 1 - [\bar{D}./E_{H_0}(\bar{D}.)].$$

(For the explicit formula see the appendix.) The weights in κ depend on the special measure of

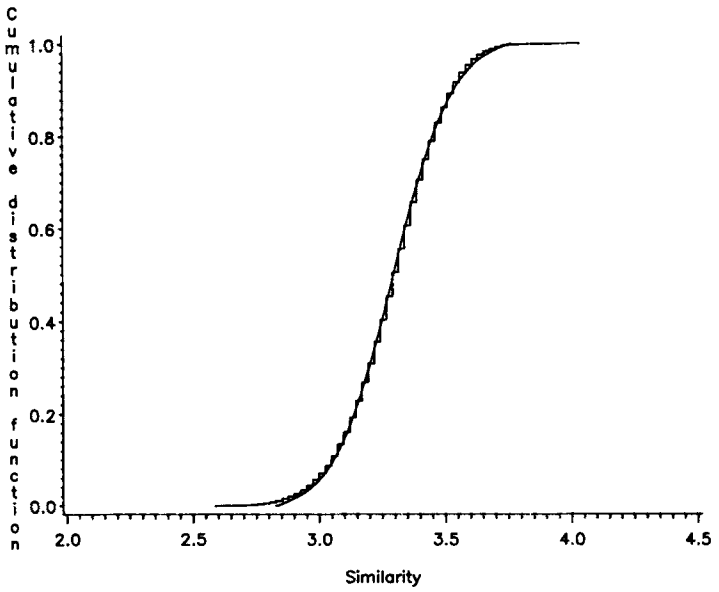

Figure 2. Simulated (step) and asymptotic (continuous) distribution function of the test statistic for the comparison of 'degree of being bedridden' and the quality of life-index: absolute differences in ranks are used as a measure of similarity

similarity chosen for the individual response curves. These could be either the 'absolute weights' $\sum_{t=1}^{T} |R(x_t)-R(y_t)|$ or the 'Spearman weights' $\sum_{t=1}^{T} [R(x_t)-R(y_t)]^2$.

The measure κ adjusts the overall observed mean similarity for the expected chance similarity conditional on the observed rank pattern within each variable, thereby adjusting also for common time effects in the variables. κ can vary between -1 and 1. A positive value of κ indicates stronger similarity than expected by chance; a value of 0 indicates that the observed similarity can be explained by chance alone, and a negative value means less than chance similarity. It should be noted that the possible values of κ might lie inside the interval from -1 to 1 for a given sample.

4. RESULTS

In the present study, three pairwise comparisons of variables 'dyspnea' 'oedema' and 'degree of being bedridden' with QI could be made. Because the number of completely observed curves in the variables is quite large ($N=41$) it is impracticable to calculate the exact permutation distribution of the test statistic. To perform the tests both the 'simulated' and the asymptotic version of the distribution are computed. Figures 2 and 3 show the results of these computations for the variables 'degree of being bedridden' and 'oedema', respectively.

For the simulation a sample size of 20,000 (without replacement) was assumed appropriate. The results of both methods correspond quite well, indicating the appropriateness of the normal assumption, at least for 'degree of being bedridden'. The poorer approximation for 'oedema' by the normal distribution can be explained by the fact that many patients did not have oedema during the course of the study, and therefore special combinations of response curves occurred much more often than others.

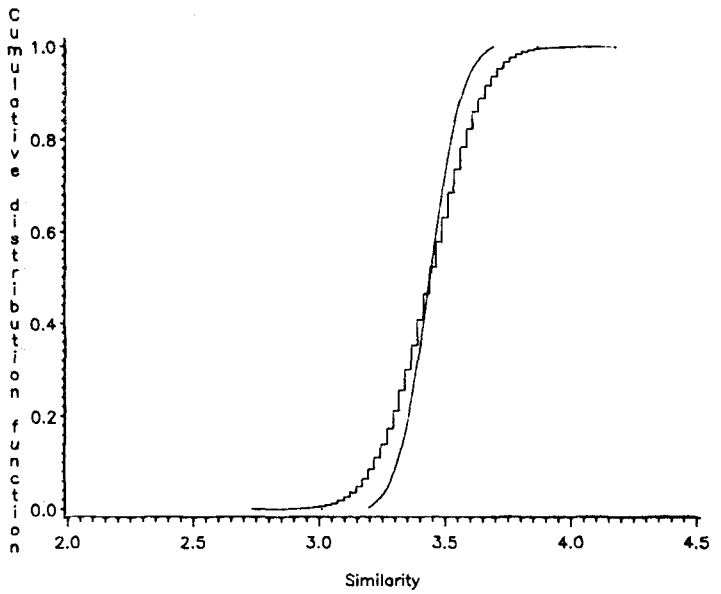

Figure 3. Simulated (step) and asymptotic (continuous) distribution function of the test statistic for the comparison of 'oedema' and the quality of life-index: absolute differences in ranks are used as a measure of similarity

Table II. Results of the test on independence (simulated and asymptotic) for the three test problems based on the absolute differences in ranks D. The global measure of similarity κ is computed with absolute weights.

Test problem	Mean similarity $\bar{D}.$	$E_{H_0}(\bar{D}.)$	$\mathrm{var}_{H_0}(\bar{D}.)$	κ	p-value
Dyspnea–QI	2·51				
Simulated		3·065	0·038		0·0051
Exact/asymptotic		3·065	0·020	0·18	<0·0001
Oedema–QI	3·59				
Simulated		3·440	0·028		0.823
Exact/asymptotic		3·440	0·010	−0·04	>0·925
Degree-of being bedridden-QI	2·59				
Simulated		3·296	0·035		0·00005
Exact/asymptotic		3·296	0·035	0·22	<0·0001

Table II summarizes the results of the study for the relationships among the three clinical variables and QI. The table shows the mean observed similarity among the curves in the sample, the conditional expectation and variance under the hypothesis of independence (simulated and exact), the overall observed similarity measured by κ, and the corresponding (estimated and asymptotic) one-sided p-values. Table III presents the results when 'Spearman weights' (the quadratic differences in ranks) are used; the results are quite similar.

Table III. Results of the test on independence (simulated and asymptotic) for the three test problems based on the quadratic differences in ranks S. The global measure of similarity κ is computed with 'Spearman weights': $\bar{S}$. is defined as the mean observed similarity when Spearman weights are used

Test problem	Mean Similarity $\bar{S}$.	$E_{H_0}(\bar{S}.)$	$\mathrm{var}_{H_0}(\bar{S}.)$	κ	p-value
Dyspnea-QI	3·72				
Simulated		4·899	0·214		0·0080
Exact/asymptotic		4·901	0·178	0·24	<0·0025
Dyspnea-QI	5·32				
Simulated		4·926	0·164		0·82
Exact/asymptotic		4·925	0·136	−0·08	>0·85
Degree of being bedridden-QI	3·76				
Simulated		5·159	0·218		0·0020
Exact/asymptotic		5·162	0·217	0·27	<0·0013

Because of the dependence of the single tests a multiple test procedure, the Bonferroni–Holm method, was applied for adjusting the level of significance. At a multiple 5 per cent level a significant association between QI and 'degree of being bedridden' and between QI and 'dyspnea' could be ascertained. The global measure of similarity κ indicates a stronger similarity than expected by chance among the observed response curves. No relationship could be found between QI and 'oedema': the global measure of similarity is even negative. The results show that the quality of life of the patients is adversely affected by symptoms of dyspnea and by being confined to bed. Having oedema seems not to influence markedly the well-being of the patients.

5. COMPARISON WITH OTHER PROPOSALS

Several approaches are mentioned in the literature dealing with the detection of a relationship in paired response curves. A comparison should demonstrate whether they are applicable in a given situation. It is necessary to distinguish approaches which use the transformation of the curves into their sign pattern and those which use the rank transformation. Within these approaches there are possibilities of first characterizing the similarity of the curves at the individual level or of characterizing the average similarity of the curves at the distinct time points.

The approach of Krauth[10] uses the sign transformation, assigning a plus when the curve is increasing between two time points and a minus when the curve is decreasing. Each subject is then characterized with respect to the variables X and Y. Subjects are arranged in a contingency table with the categories being the different sign pattern. Krauth proposes the application of Bowker's test of symmetry[11] for contingency tables for testing the hypothesis of symmetry of the two variables.

This approach implies a 'measure of similarity' for the individual response curves which is only dichotomous: the curves are either equal or different. Furthermore, in analysing the dependence of two variables measured at a few time points, the proposed test of symmetry has no power to detect the alternative, in the context of independence, that in half of the sample the observed curves in X are strictly increasing and in Y are strictly decreasing while in the other half of the sample the variables behave the opposite way round.

Table IV. Results of Lehmacher's test for profile parallelism for the three test problems

Test problem	Test statistic	p-value
Dyspnea–QI	1·18	0·80
Oedema–QI	7·68	0·05
Degree of being bedridden–QI	5·68	0·10

In their non-parametric approach for the comparison of two or more time series, Yassouridis and Hansert[12] proposed a 'measure of equidirection' based on the pattern of signs of the time series. The number of transitions ($t \rightarrow t+1$) at which not all time series change in the same direction is counted. This measure Z lies between 0 and $T-1$ and can be standardized for two time series in the following way:

$$g = 1 - 2Z/(T-1), \qquad -1 \leqslant g \leqslant 1.$$

Translated to the context of a sample of bivariate response curves, this measure can serve the same purpose as the other measures of similarity mentioned above, with the same considerations in deriving the distribution of the test statistic and a global measure.

The approach chosen by Lehmacher[13] for testing the hypothesis of profile parallelism uses the characterization of the similarity of the curves at the distinct time points. Ranks are assigned within each response curve and the vector of the differences in ranks at each time point for each pair of curves is calculated. Then a test for profile parallelism is applied to the vector of the mean observed differences at each time point. The hypothesis of parallel mean curves is rejected if the mean observed differences differ significantly from zero. Table IV shows the results of this test for our study. In interpreting the result one has to keep in mind that the hypothesis is reversed compared with the hypothesis of independence. The low p-value for 'oedema' indicates that the curves might not be parallel, that is the variables might be independent. Therefore the tests leads to similar, albeit non-significant, results as above.

Like Krauth's[10] proposal this test has no power to detect the 'symmetrical' alternative described above in a test of independence. This situation may not occur when two different treatments are compared within one patient. In quality of life studies this would indicate a strong negative correlation of the two variables.

6. DISCUSSION

In this paper a method is proposed for the analysis of the relationship between two variables measured several times within one individual. The comparison with other proposals in the literature developed for related problems shows that these approaches are only partly applicable in the present situation. The approaches which use the sign transformation retain less information about the original shape of the response curves and are therefore not as powerful as the proposed method. Furthermore, no measure is proposed for judging the overall observed similarity.

By first characterizing the similarity of two curves at the individual level the time effect is eliminated, resulting in a one dimensional measure. This measure can be used for defining the weights in Cohen's weighted kappa coefficient, which can thereby be generalized as a measure of association between the components of a sample of bivariate response curves.

Table V. Contingency table for ranked response curves in two variables

		Observed rank pattern in X					
		1	2	3	...	n	$\sum$
	1			.			.
observed	2			.			.
rank pattern	3			.			.
in Y	.			... N_{ij} ...			$N_{i.}$
	.			.			.
	.			.			.
	m			.			.
	$\sum$			... $N_{.j}$ ...			$N_{..}$

Although the measure of similarity which uses the quadratic differences in ranks may be more popular because of its connection with Spearman's rank correlation coefficient, there is no special argument for its preference. It distinguishes more levels of similarity and therefore is less robust for the whole sample compared with the measure which is the sum of the absolute differences in ranks.

By using the idea of the permutation tests, the distribution of the test statistic can be derived and no special assumption about the shape of the curves is required. Although the computation of the exact permutation distribution of the test statistic may be impracticable, the distribution can be derived by using simulation or, if indicated, by approximation by a normal distribution.

The method represents a suitable tool for detecting a relationship between two variables measured at a number of prespecified time points and could prove useful for the evaluation of quality of life.

APPENDIX

To derive the conditional expectation and conditional variance of the test statistic under the hypothesis of independence (H_0), the random variable $\bar{D}.$ is represented as a weighted sum of the observed frequencies of the ranked pairs of response curves.

$$\bar{D}. = \sum_{i=1}^{n} \sum_{j=1}^{m} d_{ij} N_{ij}/N.$$

The weights d_{ij} denote the similarity between two curves, and N_{ij} denotes the frequency of occurrence of such a special pair of ranked curves. Indexes n and m denote the number of different ranked response curves in the two variables X and Y, respectively. To illustrate this notation, the different ranked response curves in the two variables can be arranged in a contingency table (Table V). For the calculation of the conditional expectation and variance, the marginal sums $N_{i.}$ and $N_{.j}$ in the contingency table are assumed fixed.

With this notation,

$$E_{H_0}(\bar{D}.) = \sum_i \sum_j d_{ij} E_{H_0}(N_{ij})/N,$$

$$\operatorname{var}_{H_0}(\bar{D}.) = \left[\sum_j \sum_j d_{ij}^2 \operatorname{var}(N_{ij}) + \sum_i \sum_j \sum_k \sum_l d_{ij} d_{kl} \operatorname{cov}_{H_0}(N_{ij}, N_{kl}) \right] \Big/ N^2,$$

$$i \neq k \text{ or } j \neq l$$

with

$$E_{H_0}(N_{ij}) = \frac{N_{i.}\, N_{.j}}{N_{..}}$$

$$\mathrm{var}_{H_0}(N_{ij}) = \frac{N_{i.}\,(N_{..} - N_{i.})\, N_{.j}\,(N_{..} - N_{.j})}{N_{..}^{2}\,(N_{..} - 1)}$$

$$\mathrm{cov}_{H_0}(N_{ij}, N_{kl}) = \frac{N_{i.}\,(N_{..} - N_{i.})\, N_{.j}\,(N_{..} - N_{.j})}{N_{..}^{2}\,(N_{..} - 1)}, \qquad i = k \text{ and } j = l$$

$$- \frac{N_{i.}\,(N_{..} - N_{i.})\, N_{.j}\, N_{.l}}{N_{..}^{2}\,(N_{..} - 1)}, \qquad i = k \text{ and } j \neq l$$

$$- \frac{N_{i.}\, N_{k.}\, N_{.j}\,(N_{..} - N_{.j})}{N_{..}^{2}\,(N_{..} - 1)}, \qquad i \neq k \text{ and } j = l$$

$$\frac{N_{i.}\, N_{k.}\, N_{.j}\, N_{.l}}{N_{..}^{2}\,(N_{..} - 1)}, \qquad i \neq k \text{ and } j \neq l.$$

The transformed variable $[\bar{D}_{.} - E_{H_0}(\bar{D}_{.})]/\sqrt{[\mathrm{var}_{H_0}(\bar{D}_{.})]}$ can be approximated by a standard normal distribution, with the restriction that the table should not be too 'extreme', that is one or two special combinations of two ranked curves should not occur much more often than other combinations.

ACKNOWLEDGEMENT

I thank Martin Schumacher, Juergen Schulte Moenting and Walter Lehmacher for helpful discussions and suggestions on this work and Roland Bersdorf for permission to use the data.

REFERENCES

1. Bersdorf, R., Kümmell, H. Chr. and Scholz, G. 'Optimierte Digitalistherapie', *Therapiewoche*, **37**, 401–412 (1987).
2. Bersdorf, R. 'Prüfparameter in der Digitalistherapie: Klinik, Herzdynamik und Befinden', PhD thesis, Witten-Herdecke, 1987.
3. Arminger, G. *Faktorenanalyse*, Teubner, Stuttgart, 1979.
4. Immich, H. and Sonnemann, E. 'Which statistical methods can be used in practice for the comparison of curves over a few time-dependent measure points?', *Biometrie-praximetrie*, **15**, 43–52 (1974).
5. Lehmann, E. L. and D'Abrera, H. J. M. *Nonparametrics: Statistical Methods Based on Ranks*, Holden-Day, San Francisco, 1975.
6. Pratt, J. W. and Gibbons, J. D. *Concepts of Nonparametric Theory*, Springer, New York, 1981.
7. Cohen, J. 'Weighted kappa: nominal scale agreement with provision for scaled disagreement or partial credit', *Psychological Bulletin*, **70**, 213–220 (1968).
8. Fleiss, J. L. and Cohen, J. 'The equivalence of weighted kappa and the intraclass correlation coefficient as measures of reliability', *Educational and Psychological Measurements*, **33**, 613–619 (1973).
9. Lienert, G. A. *Verteilungsfreie Methoden in der Biostatistik*, vols I and II, Hain, Meisenheim, 1978.
10. Krauth, J. 'Nichtparametrische Ansätze zur Auswertung von Verlaufskurven', *Biometrische Zeitschrift*, **15**, 557–566 (1973).
11. Bowker, A. H. 'A test for symmetry in contingency tables', *Journal of the American Statistical Association*, **43**, 571–574 (1948).
12. Yassouridis, A. and Hansert, E. 'Equidirection: a measure of similarity among time series', *Biometrical Journal*, **28**(6), 747–758 (1986).
13. Lehmacher, W. 'Tests for profile analysis of paired curves based on Friedman ranking methods', *Biometrical Journal*, **22**(2), 141–152 (1980).