

## Computer aided EEG analysis: Evaluation of changes during cognitive processes and topographic mapping

P. Rappelsberger, H. Pockberger, H. Petsche

### Summary

*In the past decade, computer aided EEG analyses have proved to be an efficient means in studying cognitive processes. Our approach is based on simultaneous recordings with 19 EEG channels in order to consider as many as possible topographic properties. Spectral parameters are extracted from the EEG of control recordings and recordings during the performance of cognitive tasks and are compared to evaluate significant differences. The used statistical procedures aim at an explorative statistic and are not to confirm but only to describe possible changes during cognitive processing. The results are presented using appropriate mapping techniques. The applied methods are described in detail and are demonstrated by some examples.*

### Zusammenfassung

*Computerunterstützte EEG-Analysen haben sich in den letzten Jahren als wirksames Mittel beim Studium kognitiver Prozesse erwiesen. Unser Ansatz geht von der gleichzeitigen Ableitung mit 19 EEG-Kanälen zur Erfassung möglichst vieler topographischer Eigenschaften aus. Aus den EEG von Kontrollableitungen und Ableitungen während der Durchführung kognitiver Aufgaben werden Spektralparameter extrahiert und miteinander zur Bestimmung von signifikanten Unterschieden verglichen. Die eingesetzten statistischen Verfahren dienen einer explorativen Statistik und nicht dem Nachweis, sondern lediglich der Beschreibung wahrscheinlicher Veränderungen während kognitiver Prozesse. Die Darstellung der Ergebnisse erfolgt mit Mapping-Techniken. In der Arbeit wird die gesamte Methodik beschrieben und durch Beispiele aus verschiedenen Studien erläutert.*

### 1. Introduction

The conjecture that cognitive processes are reflected in the electrical brain activity goes back till the early days of the discovery of the EEG by Hans Berger. However, in those days, the technical equipment was far from sufficient to proof this conjecture. Only in the last decade computer aided EEG

analysis has turned out to be a powerful tool for the study of cognitive processes.

Nowadays, the number of papers dealing with EEG related aspects of cognitive performance is considerably increasing, mainly in research concerning hemispheric dominance (LEDOUX, 1983) and animated by cognitive studies in split brain patients. Besides psychophysiological studies on normal subjects in recent years also studies on depressive and schizophrenic patients (FLOR-HENRY, 1983) have been performed aiming at lateralisation effects due to specific tasks. Hitherto, however, the results have been far from consistent. One major reason for the often divergent results seems to be the different psychological paradigms and the different methods of EEG recording and processing. In many studies the recording sites were restricted to the temporal and parietal regions and hardly more than four electrodes are used for each hemisphere. Moreover, in many studies only the EEG in the alpha-frequency range was examined and other frequency bands were not taken into consideration.

Principally, two strategies may be followed in studying mental activities: the first strategy uses event related potentials, as e.g. many psychologists, while the second approach deals with the analysis of the ongoing EEG. Following the second strategy, preferably spectral analytical methods are applied to reduce the data and to obtain frequency dependent parameters, which are taken as a measure of neocortical activity. Most authors applying spectral analytical methods are only interested in, or have only the possibilities to compute, power spectral parameters. But there are also a few publications reporting about significant changes in coherence parameters related to different cognitive activities (SHAW et al., 1976; SKLAR et al., 1972; BUSK and GALBRAITH, 1975). The coherence function is a measure of the correlation of two signals in the frequency domain and may be used as a measure for functional relationships between brain areas.

However, most of the results obtained thus far by power and coherence analysis in the various studies are hardly comparable, since different modes of EEG recording, diverse recording conditions (too few electrodes) and various frequency bands were used.

In our studies both power- and coherence analyses are made in attempting at evaluating EEG changes during cognitive

tasks. The studies are done with healthy volunteers and psychiatric patients and demonstrate that changes of the EEG due to different conditions effect not only a single but various frequency bands and that EEG changes are not restricted to a single location. The studies hitherto clearly show that power and coherence analyses are complementary to each other thus revealing more insight into the dynamics of the EEG.

The use of more than a few recording electrodes, preferably 19 according to the international 10–20 system or even more, requires appropriate methods for the presentation of the multitude of parameters. In the past, several topographic mapping methods have been introduced, e.g. REMOND and OFFNER (1952), LEHMANN (1971), CHILDERS et al. (1972), BUCHSBAUM et al. (1982) and others. DUFFY et al. (1981) introduced a topographic method called »significant probability mapping«. By this method significant changes in either amplitude of evoked potentials or power of spontaneous EEG activity related to different cognitive tasks or psychopathological syndroms are plotted according to their topographic distribution. On our studies we apply a modified version of Duffy's approach and the results hitherto indicate how much more information can be drawn from the EEG by topographic mapping.

## 2. Methods

### 2.1 EEG recording and parameter selection

After carefully measuring and marking of the electrode positions according to the international 10–20 system 19 gold-disc electrodes are glued to the scalp. Additionally, reference electrodes are applied to both ear lobes. Recordings are made against the linked ear electrodes. The filters of the EEG amplifiers are set to 30 Hz cut-off frequency and 0,1 s time constant. The EEG is written out on paper (Fig. 1) and stored on analogue tape for further processing.

The subjects are sitting comfortably in an arm chair. The test situation is as follows: each task lasts 1 minute; the tasks are separated one from another by a period of rest while the subject is allowed to move his head and limbs. During this resting interval, the subject is interrogated about his impressions of the just performed task. This not only yields additional personal data on performance strategies for each task but also has the purpose of maintaining the person in an approximately constant state of vigilance. Before and after each task 1 min control records are made with eyes opened or closed according

to the special task, while the subjects are asked to relax. Each experiment starts with two control records, one with eyes opened and one with eyes closed, to enable each subject to acquaint himself with the situation. In the eyes-opened, relaxed condition the subjects are asked to fixate a black dot on a white wall 2 m in front of them. This is to reduce eye movements.

Depending on the aim of a study the following mental tasks have been performed with eyes opened: silently reading a newspaper text which is thought to be of interest to the subject, viewing a painting with many details, spatial imagination tests and others. Corresponding tasks with eyes closed have been: doing mental arithmetic, listening to a story, listening to different kinds of music, and others. The sequence of the tasks in a study is changed at random from subject to subject. Each session takes approximately half an hour to record the EEG data.

Data analysis is done off-line. The 19 EEG traces are digitized at 256/s and up to 15 artifactfree 2 s epochs are selected from each record and Fourier transformed. Conspicuous artifacts are eliminated from computation by visual inspection. After Fourier transformation of the 2 s epochs averaged power spectra and averaged cross-power spectra are computed. This yields 19 power spectra for each recording.

Averaged cross-power spectra are calculated between adjacent electrodes along the transversal and longitudinal electrode rows and also between electrodes on homologous sites of both hemispheres. This yields 38 cross-power spectra for each recording.

Besides the visual elimination of conspicuous artifacts a quality control program is run. This is based on the computation of various relations of the amount of power (GEVINS et al., 1977; HINRICHS et al., 1984) between the different frequency bands as given below. The result yields the basis for an artifact classification scheme used in the further processing to eliminate artifacts restricted only to individual frequency bands like e.g. eye movement artifacts in the theta band or muscle artifacts in the beta band.

Since 19 power spectra and 38 cross-power spectra each with a frequency resolution of 0.5 Hz according to the 2 s epochs and calculated up to 31.5 Hz yield an enormous amount of data, data reduction is performed. This data reduction is done by extracting broad band parameters for five frequency bands: theta (4–7.5 Hz), alpha (8–12.5 Hz), beta1 (13–18 Hz), beta2 (18.5–24 Hz) and beta3 (24.5–31.5 Hz). The broad band parameters are absolute power and coherence. Thus the

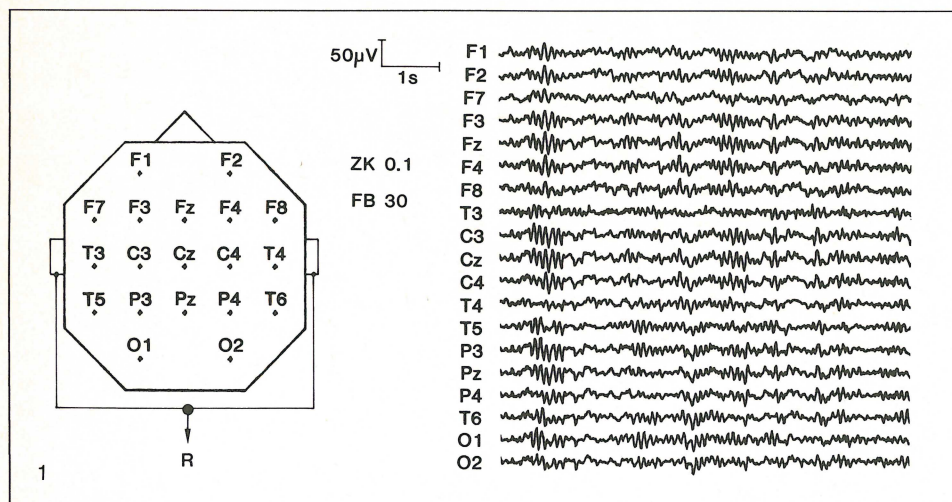


Fig. 1: Electrode positions according to the international 10–20 system (left) and example of an EEG record with eyes closed against linked ear electrodes (right).

number of data per power spectrum is reduced to 5 power values according to the five frequency bands. Those 5 parameters per spectrum, in total  $5 \times 19 = 95$  parameters per recording, are stored in a data base. The five coherence values according to the five broad bands are also stored in the data base. In total,  $38 \times 5 = 190$  coherence parameters are obtained per recording, either a control or a task recording.

Personal data of each proband like sex, handedness, age, task relevant data like training in music etc., are also entered in the data base and serve as discrimination criteria for the selection of data. Finally, the above-mentioned artifact classification scheme of each recording is stored.

## 2.2 Statistical Procedures

The applied statistical procedures aim at the evaluation of significant differences between control and task situations within subjects (patients) or groups of subjects, or between two groups of subjects with different properties but having performed the same task (male-female, training in music-no training in music, healthy subject-patient etc.).

Since with respect to a special task two control records are made both comparisons can be performed: »control before: task« and »control after:task«. If both comparisons show a significant change the probability is high that this change is task dependent. If e.g. the alpha power at electrode  $P_3$  significantly increases in the comparison »control before:task« and also in the comparison »control after:task« the result can be considered as significant reversible. A significant increase in the comparison »control after:task« corresponds to a significant decrease in the comparison »task:control after« and therefore, the power increase in the comparison »control before: task« is reversible. This is a strong criterion for changes dependent on a special task (see also discussion).

### 2.2.1 Confidence intervals

In individual subjects, for the evaluation of significant differences between the parameters of the control and task situations, 95 % confidence intervals are computed (RAPPELSBERGER, 1977). Power values are known to be chi-square distributed (JENKINS and WATTS, 1968) and the equivalent degrees of freedom are obtained by the averaging procedures both over the number of epochs and the number of frequency bands.

The distribution of coherence values is unknown. However, since coherence values are correlation coefficients per frequency band the z-transformation towards normal distribution as proposed by Fisher is applied. Confidence intervals are estimated in the z-plan (RAPPELSBERGER, 1977) and bias corrected according to BENIGNUS (1969).

### 2.2.2 Fisher permutation test

For group comparisons randomization tests according to Fisher (EDGINGTON, 1980; LEBART et al., 1982) are applied. The first test concerns paired groups (»control before:task« or »control after:task«). In this case differences between corresponding parameters are computed and the distribution of the differences is tested by the permutation paired Fisher test. This non-parametric test yields critical probabilities or estimations of them with confidence limits. This is dependent on the number of subjects within one group. Since  $N$  subjects require  $2^N$  systematic permutations per test computation time is increasing exponentially with increasing number of subjects.

Therefore, if  $2^N$  exceeds 5000 the systematic permutation is replaced by a random permutation.

For the comparison of the parameters of two independent groups (e.g. males:females in the same task situation) a similar permutation test is applied. In this case it is tested if the distributions of both parameter groups are identical. The results are also critical probabilities or estimates of them with confidence limits.

According to the artifact classification scheme for frequency dependent artifacts in individual leads artifact contaminated data are omitted from computation.

## 2.3 Topographic mapping

The results are presented in topographic maps. Significant power changes are indicated by squares at the corresponding electrode position. Empty squares indicate a power decrease and full squares a power increase.

A significant transverse or longitudinal coherence change is presented by a square drawn between the two electrodes involved. At least, a significant interhemispheric coherence change is marked by squares at the two electrode locations of homologous sites of both hemispheres connected by a line. As with power changes, full squares indicate a coherence increase and empty squares a coherence decrease.

In group comparisons three different sizes of squares can be drawn according to the obtained critical probability. A small square relates to a critical probability of  $2P = 0.10$ , a middle sized square to a critical probability of  $2P = 0.05$  and a large square to a critical probability of  $2P = 0.01$ , respectively.

The results based on confidence intervals are presented only by middle sized squares according to  $2P = 0.05$ . Significant results obtained in frequency bands and electrodes with artifacts are not presented.

## 3. Examples

### 3.1 Single subject studies

Fig. 2 shows an example out of a study aimed at evaluating EEG changes during mental arithmetic. Only the results obtained in the alpha band are demonstrated. In the uppermost row significant changes are presented based on the comparison »control before:mental arithmetic« (KV-KR). The second row shows the results of the comparison »control after:mental arithmetic« (KN-KR) and in the third row the significant reversible changes are demonstrated (KV-KR & KN-KR). In the left column the significant power changes can be seen. Coherence changes are presented in the middle and right columns.

During mental arithmetic in almost all brain areas a significant alpha power decrease is found which is reversible with one exception. In electrode  $O_2$  (see Fig. 1) only in the comparison »control before:mental arithmetic« (KV-KR) a significant power decrease is seen.

Power decrease can be accompanied by coherence decrease or coherence increase, as the middle row of Fig. 2 demonstrates. This clearly demonstrates that coherence is an additional parameter to characterize the electrophysiological properties during cognitive processes. Coherence may be interpreted as a measure of the functional relationship between different areas of the brain and is independent of the absolute amount of EEG power.

The coherence increase in the middle column of Fig. 2 during mental arithmetic concerns the central regions respon-

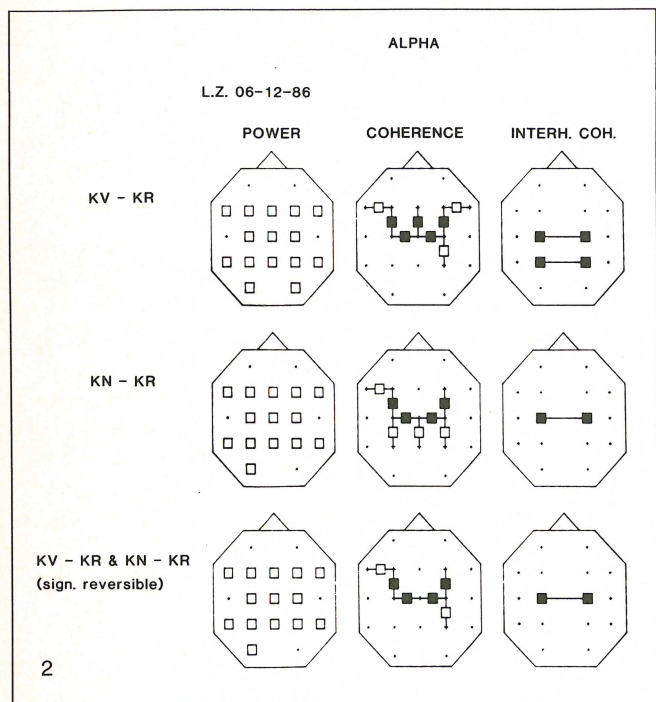
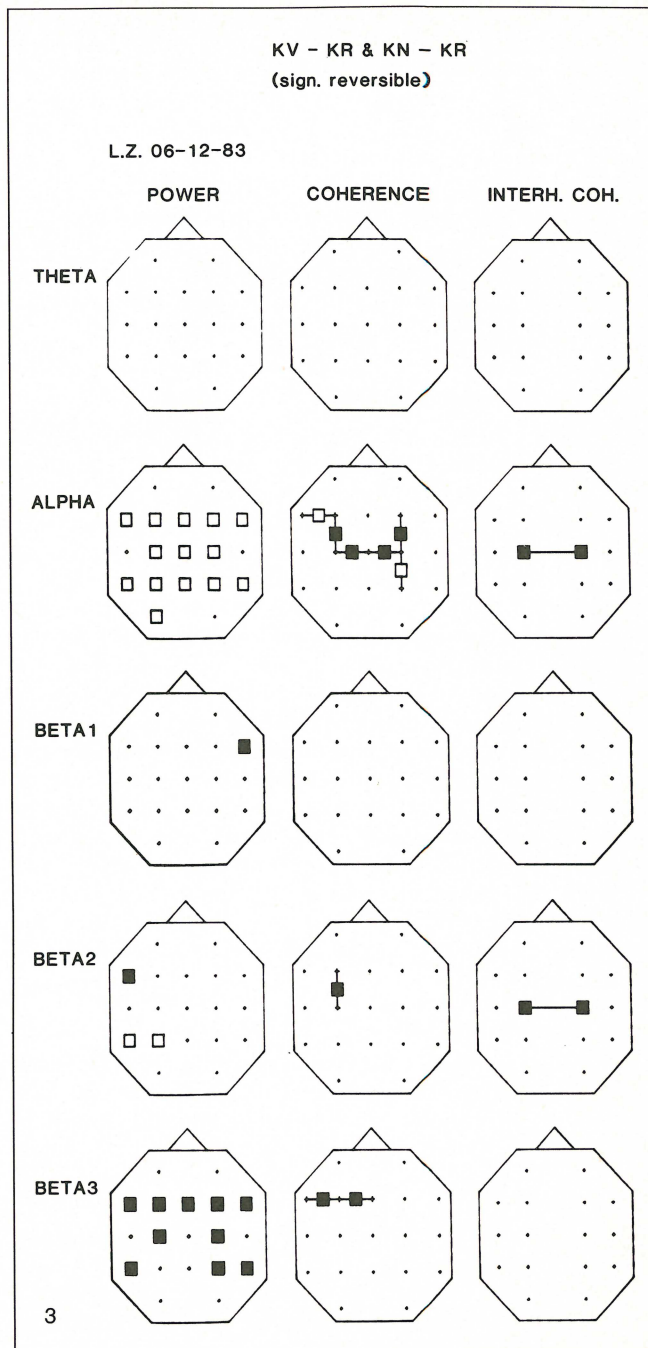


Fig. 2: Topographic maps showing the changes in the alpha band in a single subject during mental arithmetic. Row by row the significant changes based on the comparisons »control before (KV):mental arithmetic (KR)« and »control after (KN):mental arithmetic (KR)« and the significant reversible changes (KV-KR & KN-KR) are plotted. The three columns relate to power-, coherence- and interhemispheric coherence changes. Significant changes found by means of 95% confidence intervals are indicated by squares. Empty squares indicate a decrease of the corresponding parameter during KR as compared with a control situation (KV or KN or both) and full squares indicate an increase. Power changes are plotted at the corresponding electrode locations. Longitudinal or transversal coherence changes are indicated by a square between both electrodes involved. An interhemispheric coherence change is shown by two squares at the corresponding electrode positions connected by a line. For the interpretation of the results see text.

Fig. 3: Significant reversible changes (KV-KR & KN-KR) during mental arithmetic of the same subject as in Fig. 2. The figure shows the results obtained in the 5 frequency bands examined. The arrangement of the columns is as in Fig. 2 and the maps of the alpha band correspond to the lowermost maps of Fig. 2.

Besides the results in the alpha band an essential finding is the increase of beta3 power showing that cognitive processes are not restricted to a single EEG frequency band.



information flow during mental arithmetic. In the comparison »control before:mental arithmetic« (KV-KR) also a coherence increase between electrodes P<sub>3</sub>-P<sub>4</sub> is found indicating an increase in the functional coupling of the parieto-occipital association areas. However, this finding does not show a significant reversibility.

Fig. 3 shows the significant reversible results of all the five frequency bands examined. Besides the results in the alpha band as shown in Fig. 2 there is a significant increase of power in the beta3 band during mental arithmetic. This increase is located in the frontal, central and parieto-temporal regions and hints at the activation of high frequent activity during the task. However, coherence increase within this frequency band is only observed in the left frontal region. This example is to demonstrate that EEG changes during cognitive processes are

not restricted to a single frequency band, e.g. the alpha band, but other frequency bands may also be involved and should therefore be examined in such studies.

3.2 Group studies

3.2.1 Paired studies

Examples of paired comparisons for a group of 19 probands show Figs. 4 and 5. The group consisted of 10 female and 9 male students who had to read a text silently, which was thought to be of interest for them, during the EEG recording. In Fig. 4 the results obtained for the alpha band are demonstrated in detail. The arrangement of the plots is as in Fig. 2. However, the squares indicate critical probabilities according to the results of the paired Fisher permutation test.

During silent reading power decreases in the fronto-central, central and parietal regions and also in the left temporal and occipital areas. This indicates that almost all cortical areas are involved to a high degree. The power decrease is accompanied by coherence decrease in the frontal and in the left and right posttemporo-parieto-occipital regions and by coherence increase in the central parts. The decrease of frontal coherence is most likely due to desynchronized activities in the frontal association areas. In contrast to this, the increase of central coherence indicates increased functional couplings within and between motor cortices possibly due to the fact that the subjects had to hold a sheet of paper with the text in their hands and possibly due to the activation of the motor speech area during mental reading.

The decrease of coherence in the rear parts of the brain may be interpreted as due to desynchronization processes according to the visual information processing involving the occipital cortices and the parietal and temporal association areas. It is known that the temporal lobe may be ascribed an integration of visual informations. In this context the coherence decrease between electrodes T<sub>3</sub>-T<sub>5</sub> (see Fig. 1) in the left hemisphere seems to be meaningful since it could hint at the involvement of the sensory speech area. Furthermore, the coherence decrease in the frontal regions may be considered under the view-point of the known fiber connections between the visual association areas and the prefrontal association areas.

In the right column of Fig. 4 significant reversible interhemispheric coherence decrease can be observed in the frontal and prefrontal areas. This indicates a functional decoupling between those parts of both hemispheres. Only the motor regions show an increased functional coupling. No significant reversible interhemispheric coherence changes are found in the rear part.

Additional information is obtained by considering the other frequency bands examined (Fig. 5). It would go too far here to describe all the obtained results and to discuss their possible interpretation. However, it is worth mentioning again that the electrophysiological changes during cognitive processing are not restricted to a single frequency band but the entire spectrum of the EEG is involved more or less.

3.2.2 Independent groups

Fig. 6 shows an example of the results after comparison of the EEG parameters of two independent groups of subjects who did the same cognitive task. This task was listening to the first movement of Mozart's Jagdquartett. The first group consisted of 27 male students (22.6 a ± 1.8 a), the second group of 11 female students (24.7 a ± 4.7 a).

In this case the differences of the parameters of the recordings »control before« and »listening to music«, i.e. the changes

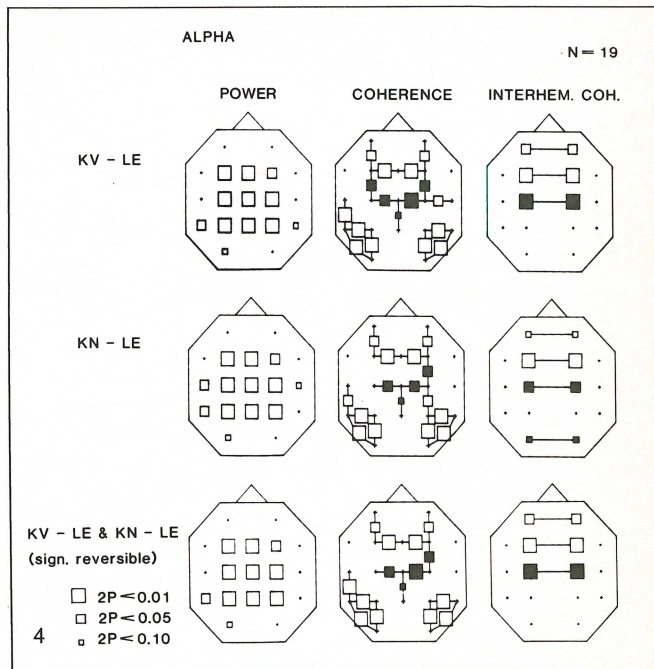


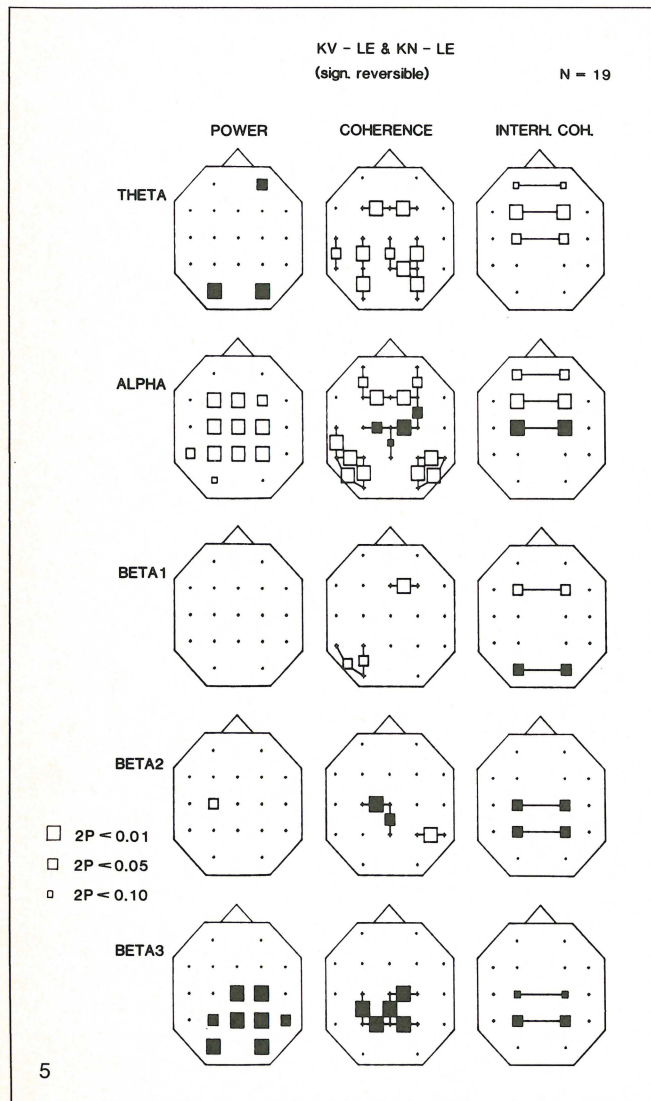
Fig. 4: Changes in the alpha band obtained in a group study with 19 subjects. In this case the subjects had to read mentally. The arrangement of the maps is as in Fig. 2. The different sizes of the squares relate to the descriptive P-values as indicated in the left lower corner. Significant changes were obtained by paired Fisher permutation procedure.

For the interpretation of the results see text.

between both situations, were computed and tested. The test aimed at the exploration of differences in the musical perception of those males and females. However, since differences were examined, the interpretation of the results is not unambiguous unless additional informations are available. Considering the empty squares between electrodes P<sub>4</sub>-T<sub>6</sub> and T<sub>6</sub>-O<sub>2</sub> in the coherence maps of the three beta bands in Fig. 6, the following interpretations can be given: first, females show less increase of beta coherence than males, or second, females show greater decrease of coherence than males, or third, if females show a decrease of coherence the coherence changes in males are positive. This additional information is yielded by printing the means of the tested parameters.

With this background, the above mentioned results are read: in the beta1 and beta3 band both males and females show a decrease of coherence in the right posttemporo-parietal-occipital region whereby the decrease in females is much greater yielding the indicated significant differences. In contrast to this, in the beta2 band males show an increase of coherence while females, as in the other beta bands, show a decrease. From this result it may be speculated that in females during listening to music the right rear part of the temporal lobe is much more involved, and therefore generating desynchronized electrical activity, than in males. This seems to be supported by the empty squares at electrode T<sub>6</sub> in the power maps of the beta1 and beta2 band obtained according to a decrease of power in females with a simultaneous increase of power in males.

One further result shall be mentioned here: the interhemispheric increase of coherence frontally (right column of Fig. 6). The indicated significant increase of coherence results due to an increase of coherence in females whereas in males a



decrease is found (theta, alpha and beta2 band). This hints at an increased functional coupling in females during listening to music.

#### 4. Discussion

Since in this paper main emphasis is laid on the methods applied a discussion of the results of our studies of EEG changes during cognitive processing is set aside. The figures shown are only to demonstrate the methods, the results of our studies will be published elsewhere.

When recording an EEG one has to take into account that it is largely dependent on the state of vigilance of the person being recorded. Especially when examining cognitive processes, changes of vigilance may distort the results and therefore one has to try to maintain the subject in a constant state of alertness. For this reason only relative short periods of about 1 min are recorded, separated by intervals during which the subject can move his head and stretch his limb. This is to avoid the subject getting tired and drowsy during a session lasting between 30 to 40 minutes according to the number of tasks which have to be performed. From subject to subject the sequence of the tasks is not constant but in random order. This serves to eliminate biasing effects or trends which may occur if

Fig. 5: Significant reversible changes obtained in a group study with 19 subjects reading mentally. The maps of the alpha band correspond to the lowermost maps of Fig. 4. See text.

the tasks are performed in a constant sequence. Since the EEG may also be influenced by the daytime of recording, all recordings are made at about the same daytime, i.e. 10 o'clock in the morning.

Whether the individual tasks are appropriate for examining a special question does not concern the methods itself and will not be discussed here for reasons already mentioned above.

The hitherto results show that during the different mental tasks many parts of the brain may be involved according to the spatial organization of the various brain functions. Therefore, topographic aspects should not be ignored in such studies. Neglecting topographical features may be the reason why many studies of EEG and higher brain functions have not been successful. In our approach we follow the recommendations of the IFSECN for clinical EEG recordings and use the full electrode set according to the international 10-20 system.

An important point is the choice of derivation technique in topographical studies. Some authors prefer common average reference recordings (e.g. LEHMANN, 1981), some use reference recordings (e.g. DUFFY et al., 1981) but also bipolar techniques are applied (PFURTSCHELLER et al., 1984). WALTER et al. (1984) point out the difficulties concerning the recording techniques. Common to all are localization problems which may occur if the reference potential is not small enough as compared with the signals at the recording sites. It is difficult to find a reference point on the head with negligible cerebral potential. This may be the case with a reference electrode placed e.g. on chin but then muscle and breathing artifacts frequently occur and obscure the recordings. A small reference signal is also necessary for an unambiguous interpretation of coherence functions (FRENCH and BEAUMONT, 1983). As with different reference points also with the common average reference one cannot be sure that the reference potential is negligible, especially in the case of highly synchronized activities. Therefore, a compromise has to be found. Linked ear electrodes as reference turned out to be the best compromise. We found that in our subjects, in which only normal EEG patterns are studied, the ear electrodes are not seriously affected by temporal EEG activity. Furthermore, the linkage of both ear leads results in an averaging of the ear potentials which may reduce their amplitude. Nevertheless, the presence of a certain amount of reference potential must be kept in mind in interpreting the results, especially if there is a significant amount of power in the temporal leads with respect to other sites.

Since almost no EEG recording is free of artifacts, i.e. non cerebral potentials, the quality control of the data has high priority. Visual control is used to eliminate conspicuous artifacts from the further processing. These are mainly eye movements and blinking artifacts whereas electrode artifacts occur very rarely because of the electrode technique used. Different automatic artifact recognition programs based on various algorithms checking different parameters are described (EARLE and PIKUS, 1982; JOHN et al., 1983, HINRICHS et al., 1984). However, none of these algorithms can be thoroughly trusted so that automatic artifact recognition programs cannot yet fully replace the visual control.

After Fourier transformation of the chosen epochs the transformed data are submitted in a second control. This is done by a program which checks different power parameters.

Fig. 6: Example of the comparison of 27 male students and 11 female students listening to music. In this case the differences between »control before (KV)« and »listening to music (MU)« were compared by means of the Fisher permutation procedure for independent samples. The different sizes of squares correspond to the descriptive P-values as shown in Figs. 4 and 5. See text.

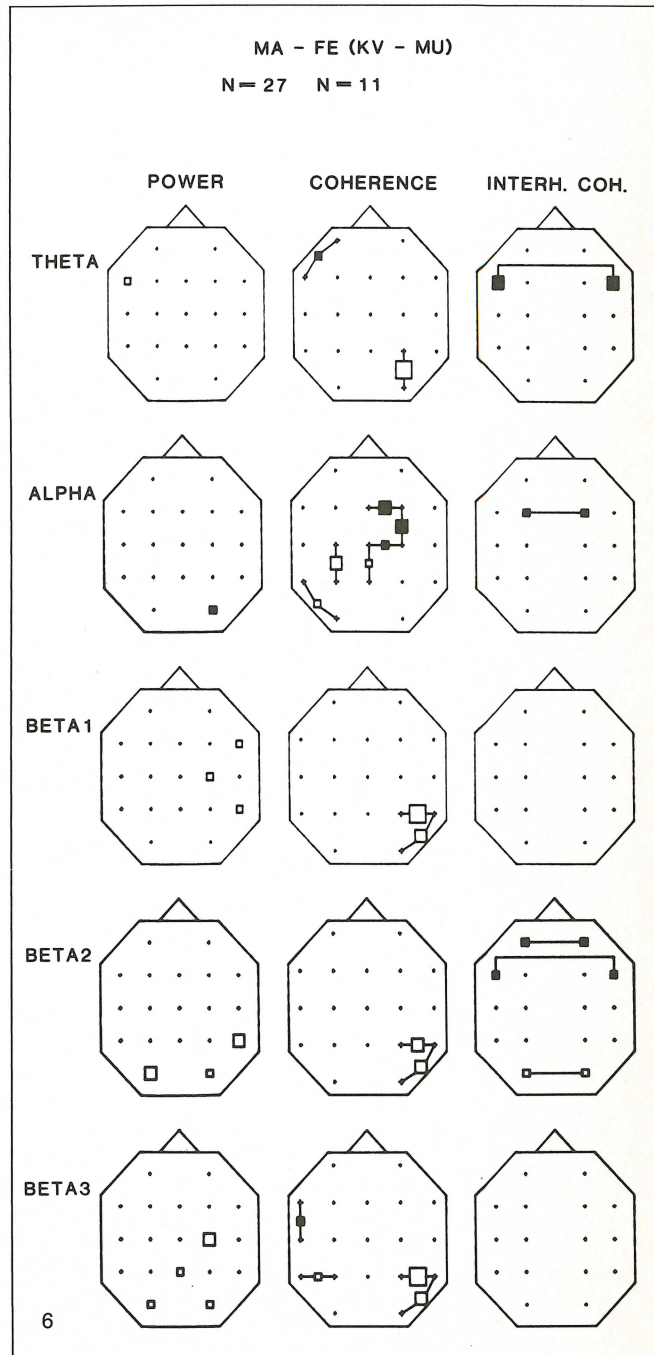
Its aim is the indication of electrode dependent and/or frequency dependent artifacts since some artifacts may be restricted only to individual electrodes and/or certain frequency bands. However, the verification of the artifacts is done visually and yields a frequency and electrode dependent artifact classification scheme. This is used for the elimination of contaminated data of individual electrodes and frequency bands prior to the applied statistical procedures.

Spectral analytic methods are widely used in EEG analyses and there are several papers dealing with power spectral analysis of ongoing EEG records during cognitive processes. In most of them the increase or decrease of power within certain frequency bands, preferably the alpha band, is related to different conditions or cognitive tasks (e.g. DUFFY et al., 1981). However, there are only a few papers in which coherence analyses were employed in studying cognitive processes (SKLAR et al., 1972; BUSK and GALBRAITH, 1975; SHAW et al., 1976, 1978; BEAUMONT et al., 1978; COLTER and SHAW, 1982). In those studies only a small number of electrodes and recording channels were used. Furthermore, different recording techniques, bipolar recordings and reference recordings with different sites of the reference electrodes, were used so that any comparison becomes difficult.

In studying functional relationships, coherence analyses yield important new aspects of brain activities which complement the data obtained by power spectral analyses, since a change of power does not necessarily mean that the functional relationship or coupling between two signals also changes. It has been argued that coherence between electrophysiological signals from different parts of the brain may depend on structural connectivities or functional couplings between these parts. For example, BUSK and GALBRAITH (1975) found EEG coherence to be higher between areas known to have relatively dense fiber connections. However, the interpretation of coherence spectra is only meaningful if a certain amount of power is present. Therefore in our studies we compute both, power and coherence spectra to obtain as many as possible information about the dynamics of the EEG (POCKBERGER et al., 1985; PETSCHKE et al., 1985a; PETSCHKE et al., 1985b; PETSCHKE et al., 1986).

As already mentioned, the reference potential can make the interpretation of coherence functions ambiguous (FRENCH and BEAUMONT, 1983) since an increase of the reference potential can yield an increase as well as a decrease of coherence. This can easily be demonstrated mathematically. The reason for this effect is the dependence of coherence on the ratios: electrode signals and reference signal (signal-to-noise-ratios). Only if these ratios are high, i.e. low reference signal, a change of coherence may be assigned unambiguously to a change in the relations of the two signals of interest.

The five frequency bands theta, alpha, beta1, beta2 and beta3 used in our analyses correspond to the usually used clinical frequency bands. However, it is questionable whether these frequency bands optimally describe the characteristic features of the EEG or whether an other division of the whole spectrum would be more appropriate. This may be a matter of future examinations. Factor analytical studies based on power



spectra (KUBICKI et al., 1980) e.g. show that the alpha band may be divided into two or three subbands; these authors also emphasize the division of the beta band into only two instead of three subbands. Their studies, however, were performed with only one or a few EEG channels and it cannot be excluded that in other brain areas different results would have been obtained.

The next point to be discussed concerns the statistics used. The procedure of computing confidence intervals of spectral parameters is well described (JENKINS and WATTS, 1968; BENIGNUS, 1969; RAPPELSBERGER, 1977) and needs no further discussion.

For the use of non-parametric tests like the Fisher permutation procedure no requirements concerning the data have to be met. In contrast to this, most parametric statistics, e.g. t-statistic, require normally distributed data. The assumption

of a normal distribution is seldom justified for neurophysiological data and t-tests cannot be assumed to be robust against all forms of violation of this assumption. This may also be true if a normal distribution is approximated by a suitable transformation of the data as proposed by GASSER et al. (1982) for power values. Therefore, in our studies the Fisher permutation procedure is used according to the suggestion of ETEVENON et al. (1985).

From the statistical point of view the significance testing of many variables like theta, alpha and beta power or at many locations like 19 electrode positions, 30 transversal or longitudinal coherences or 8 interhemispheric coherences causes problems because of the so-called  $\alpha$ -inflation, i.e. the inflation of the error probability to reject a null hypothesis although it is valid (ABT, 1983). To counteract this effect various procedures have been proposed involving so-called  $\alpha$ -adjustments (e.g. HOLM, 1979; RÜGER, 1978). However, besides a data analysis to confirm differences between true effects with control of the error probability of a false decision an entirely different approach can be made: an exploratory data analysis (ABT, 1983). Following this approach the permutation test is used to estimate critical or tail probabilities under each corresponding null hypothesis, i.e. no change of power within an individual frequency band or at a certain electrode location. This approach is a descriptive one without the aim of rejecting null hypotheses but rather to yield hints at those of the various comparisons for which differences possibly exist. For this reason, the P-values in the figures have to be considered as purely descriptive.

Significant reversible changes are obtained by a simple »AND.« comparison and can be considered as largely task dependent changes according to our experimental design. Significant changes which are obtained in only one of both comparisons are cancelled by this procedure. However, as found by PETSCHÉ et al., 1985b, informations obtained in the comparison »control after:task« and which are not yielded in the comparison »control before:task« can be very meaningful since they can hint at »echo phenomena« due to the performed task.

The value of topographic mapping has been recognized since the toposcans of WALTER and SHIPTON (1951). With the development of computer capabilities in the past decade topographic mapping of cerebral activity, as means of enhancing the diagnostic capabilities of neurophysiological data, has rapidly advanced in interest and activity. This emerging technology has proved to have great value as a complement to the usual EEG recording and to other brain imaging techniques, as e.g. CT (computer tomography) or NMR (nuclear magnetic resonance), in identifying and characterizing some types of organic neurological abnormalities. As shown in this paper, EEG mapping techniques have also proved to be an efficient tool in studying cognitive processes in the field of psychophysiology and pathophysiology (POCKBERGER et al., 1985).

## Literature

- ABT, K.: Significance Testing of Many Variables. *Neuropsychobiology*, 1983, **9**: 47-51.
- BENIGNUS, V. A.: Estimation of the coherence spectrum and its confidence interval using the Fast Fourier Transform. *IEEE Trans. on Audio and Electroacoustics*, Vol. AU-17, No 3, Sept. 1969.
- BEAUMONT, J. G., A. R. MAYER, and M. D. RUGG: Asymmetry in EEG alpha coherence and power: effects of task and sex. *Electroencephal. clin. Neurophysiol.*, 1978, **45**: 393-401.
- BUCHSBAUM, M. S., F. RIGAL, R. COPPOLA, J. C. CAPPELETTI, C. KING, and J. JOHNSON: A new system for grey-level surface distribution maps of electrical activity. *Electroencephal. clin. Neurophysiol.*, 1982, **53**: 237-242.
- BUSK, J., and G. C. GALBRAITH: EEG correlates of visual motor practice in man. *Electroencephal. clin. Neurophysiol.*, 1975, **38**: 415-422.
- COLTER, N., and J. C. SHAW: EEG coherence analysis and field dependence. *Biol. Psychol.*, 1982, **15**: 215-228.
- CHILDERS, D. G., N. W. PERRY JR., O. S. HALPENY, and J. R. BOURNE: Spatio-temporal measures of cortical functioning in normal and abnormal vision. *Comp. biomed. Res.*, 1972, **5**: 114-130.
- DUFFY, F. H., P. H. BARTELS, and J. L. BURCHFIELD: Significant probability mapping: and aid in the topographic analysis of brain electrical activity. *Electroencephal. clin. Neurophysiol.*, 1981, **51**: 455-462.
- EARLE, B. B., and A. A. PIKUS: The effect of sex and task difficulty on EEG alpha activity in association with arithmetic. *Biological Psychology*, 1982, **15**: 1-14.
- EDGINGTON, E. S.: *Randomization Tests*. Marcel Dekker, Inc., New York and Basel, 1980.
- ETEVENON, P., D. TORTRAT, and C. BENKELFAT: Electroencephalographic Cartography. *Neuropsychobiology*, 1985, **13**: 141-146.
- FLOR-HENRY, P.: Commentary and Synthesis. In: *Laterality and Psychopathology* (P. Flor-Henry and J. Gruzeliér, Eds.), Elsevier, Amsterdam, 1983, 1-18.
- FRENCH, C. CH., and J. G. BEAUMONT: A critical review of EEG coherence studies of hemispheric function. *Int. J. of Psychophysiology*, 1984, **1**: 241-254.
- GASSER, T., P. BÄCHER, and J. MÖCKS: Transformations towards the normal distribution of broad band spectral parameters of the EEG. *Electroencephal. clin. Neurophysiol.*, 1982, **53**: 119-124.
- GEVINS, A. S., C. L. YEAGER, G. M. ZEITLIN, S. ANCOLI, and M. F. DEDON: On-line computer rejection of EEG artifact. *Electroencephal. clin. Neurophysiol.*, 1977, **42**: 267-274.
- HINRICHS, H., G. FERBER, and U. SANDERS: EMG-Artefakt-Erkennung für die Verwendung bei der automatischen Routine-Analyse des klinischen EEG. *Z. EEG. EMG*, 1984, **15**: 94-98.
- HOLM, S.: A simple sequentially rejective multiple test procedure. *Scand. J. Stat.*, 1979, **6**: 65-70.
- JENKINS, G. M., and D. G. WATTS: *Spectral Analysis and Its Application*, 1968, Holden Day, San Francisco.
- JOHN, E. R., L. PRICHEP, H. AHN, P. EASTON, J. FRIDMAN, and H. KAYE: Neurometric evaluation of cognitive dysfunctions and neurological disorders in children. *Progress in Neurobiology*, 1983, **21**: 239-290.
- KUBICKI, ST., W. M. HERRMANN, and G. LANDAHN (Hrsg.): *Faktorenanalyse und Variablenbildung aus dem Elektroenzephalogramm*. Gustav Fischer Verlag, 1980.
- LEBART, L., A. MORINEAU, and J.-P. FENELON (Eds.): *Traitement des données statistiques: méthodes et programmes*. Dunod, Paris, 1982.
- LEDoux, J. E.: Cerebral asymmetry and the integrated function of the brain. In: *Functions of the Right Hemisphere* (A. W. Young, Ed.), Academic Press, 1983, 203-216.
- LEHMANN, D.: Multichannel topography of human alpha EEG fields. *Electroencephal. clin. Neurophysiol.*, 1971, **31**: 439-460.
- LEHMANN, D.: Spatial analysis of evoked and spontaneous EEG potential field. In: *Recent advances in EEG and EMG data processing* (Yamaguchi, Fujisawa, Eds.), Elsevier, Amsterdam, 1981, 117-132.
- PETSCHÉ, H., H. POCKBERGER, and P. RAPPESBERGER: EEG topography and mental performance. In: F. Duffy (Ed.): *Topographic Mapping of Brain Electrical Activity*, Butterworths, 1986, pp. 63-98.
- PETSCHÉ, H., H. POCKBERGER, and P. RAPPESBERGER: EEG studies in musical perception and performance. In: R. Spintger, R. Droh (Eds.): *Musik in der Medizin*. Editiones Roche, Basel, pp. 31-60, 1985a.
- PETSCHÉ, H., H. POCKBERGER, and P. RAPPESBERGER: Musikrezeption, EEG und musikalische Vorbildung. *Z. EEG-EMG* **16**: 183-190, 1985b.



- PFURTSCHELLER, G., G. LADURNER, H. MARESCH, and R. VOLLMER: Brain electrical activity mapping in normal and ischemic brain. In: Quantitative EEG and Imaging Techniques (Pfurtscheller, G., J. Jonkman and F. Lopes da Silva, Eds.), Elsevier, Amsterdam, 1984, 287-302.
- POCKBERGER, H., H. PETSCHKE, P. RAPPESBERGER, B. ZIDEK, and H. G. ZAPOTOCZKY: On-going EEG in depression: a topographic spectral analytic pilot study. *Electroencephal. clin. Neurophysiol.* **61**: 349-358, 1985.
- RAPPESBERGER, P.: Spektrale Kurzzeitanalyse stochastischer Signale zur Auswertung von Elektroenzephalogrammen. Dissertation, Technische Universität Wien, 1977.
- REMOND, A., and F. OFFNER: A new method for EEG display. *Electroencephal. clin. Neurophysiol.*, 1952, **7**: 453-460.
- RÜGER, B.: Das maximale Signifikanzniveau des Tests: »Lehne  $H_0$  ab, wenn  $k$  unter  $n$  gegebenen Tests zur Ablehnung führen.« *Metrika*, 1978, **25**: 171-178.
- SHAW, J. C., K. P. O'CONNOR, and O. C. ONGLEY: The EEG as a measure of cerebral functional organization. *Br. J. Psychiatry*, 1976, **130**: 260-264.
- SHAW, J. C., K. P. O'CONNOR, and O. C. ONGLEY: EEG coherence as a measure of cerebral functional organization. In: *Architectonics of the Cerebral Cortex* (M. A. B. Brazier and H. Petsche, Eds.), Raven Press, New York, 1978: 245-256.
- SKLAR, R. L., J. HANLEY, and W. W. SIMMONS: An EEG experiment aimed toward identifying dyslexic children. *Nature (London)*, 1972, **240**: 414-416.
- WALTER, D. O., P. ETEVENON, B. PIDOUX, D. TORTRAT, and S. GUILLOU: Computerized Topo-EEG Spectral Maps: Difficulties and perspectives. *Neuropsychobiology*, 1984, **11**: 264-272.
- WALTER, W. G., and H. W. SHIPTON: A new topographic display system. *Electroencephal. clin. Neurophysiol.*, 1951, **3**: 281-292.

Eingegangen am 23. Juni. 1986.

Adresse der Verfasser: Doz. Dr. P. Rappelsberger, H. Pockberger, H. Petsche, Neurophysiologisches Institut der Universität, and Brain Research Institute of the Austrian Academy of Sciences, Währinger Straße 17, A-1090 Vienna.

EDV in Medizin und Biologie 17 (3), 53-59, ISSN 0300-8282

© Verlag Eugen Ulmer GmbH & Co., Stuttgart; Gustav Fischer Verlag KG, Stuttgart

## Erfahrungen mit einem rechnergestützten »Diagnoseprogramm« am Beispiel der Histopathologie der Mamma

H. J. Hahn, J. R. Iglesias, Ch. Brinzel und F. Pfannkuch

### Zusammenfassung

Ein histologisches Lernprogramm für »Personalcomputer« wurde überprüft. 239 herkömmlich befundete Mammahistologien wurden anhand von 50 morphologischen Kriterien nachbefundet. Die Berechnung der 17 Diagnosen erfolgte anhand der Diskriminierungsfunktion und der »a posteriori« Wahrscheinlichkeit auf der Grundlage des BAYESSchen Theorems.

Mit diesem »Diagnoseprogramm« konnte eine direkte Übereinstimmung der konventionell erstellten mit der errechneten Diagnose in 77,4 % erzielt werden. In weiteren 12,6 % der Fälle bestand eine große histomorphologische Ähnlichkeit der Diagnosen. Dies wurde als »Treffer 2. Ordnung« bezeichnet; Gesamttrefferquote: 90 %). Das vorgestellte »Diagnoseprogramm« ist ein leicht erlernbares Hilfsmittel zum Erarbeiten der histologischen Diagnostik. Dies wird am Beispiel der Mammahistologie dargelegt.

### Summary

A histological teaching program for personalcomputers has been studied. 239 traditional diagnosed breast tumors were re-

evaluated by calculating the frequency of 50 histological criteria.

Using the BAYES' theory of decision the discriminant function and the probability »a posteriori« of 17 diagnosis were calculated. With this »diagnostic program« a direct conclusion between the traditional and the mathematical diagnosis was attained in 77.4 %. In 12.6 % further cases the calculated and the traditional diagnoses displayed a rather similar histomorphology. They are called right classified in 2nd degree; (overall accuracy: 90 %). The introduced »computer-program« is a study aid for histological diagnostics, which is simple to learn. This is illustrated in the breast-histology.

### Einführung

Die Beurteilung histologischer Präparate durch einen Pathologen basiert auf dem Erkennen und Bewerten morphologischer Kriterien. Durch diesen Erfahrungsgewinn ergibt sich eine zunehmende intuitive Komponente. Die resultierende, individuelle Kriterienbewertung macht es schwierig, exakt reproduzierbare Diagnosen zu erstellen. Um eine bessere Reproduzierbarkeit einer histopathologischen Diagnose zu erzielen,

bieten sich die Möglichkeiten eines Rechners an. Über den erfolgreichen Einsatz rechnergestützter mathematisch-statistischer Verfahren in der klinischen Diagnostik wird in jüngerer Zeit mehrfach berichtet (FATTU 1982, FOTTER 1982, IGLESIAS u. Mitarb. 1983, 1985, PFANNKUCH u. Mitarb. 1983).

Nach der Analyse eines histologischen Präparates anhand einfach zu erkennender zyto- und histomorphologischer Kriterien und anschließender Verschlüsselung sollte es möglich sein, unter Anwendung einer statistischen Rechenoperation zu einem Diagnosevorschlag zu gelangen. Vorteile eines solchen Verfahrens wären: a) die höhere Reproduzierbarkeit der

Diagnose, b) Datenspeicherung zur Bearbeitung wissenschaftlicher Fragestellungen sowie c) der Einsatz in der Ausbildung der histologischen Diagnostik.

Im Vergleich zu den recht zeitaufwendigen Lehrverfahren, die auf gemeinsamer Beurteilung am Diskussionsmikroskop oder der Kontrolle nach Vormustern beruhen, könnte mit einem solchen kriterienorientierten, automatisierten Verfahren dem auszubildenden Diagnostiker ein brauchbares Lern-

Abb. 1. Fragebogen – Histologische Diagnose der Mammatumoren

		Nr.: ( : : : : : / )	
		51 53 55 57/58)	
Name		59:60	
		Alter:( : )	
Geschlecht: m = 1		* 0 1 2 3	
	f = 2 (61) ( )	* 1) : : : Knotig abgrenzb. Tumorwachstum	
	(62) (6) ( )	* 2) : : : Zellverb. mit Läppchenstruktur	
Untersuchungsart:		* 3) : : : Zellverb. mit Gangstruktur	
Biopsie = 1		* 4) : : : intraduct. & intrazyst. Zellverb.	
Sektionsmaterial = 2 (63) ( )		* 5) : : : siebförmige Auflockerung	
		* 6) : : : papilläre Zellverbände	
Makroskopischer Befund:		* 7) : : : atyp. tubul. Proliferation	
Max. 2 cm = 1		* 8) : : : Verhornungen	
Max. 2-5 cm = 2		* 9) : : : fischzugart. & wirbelf. Zellanordn.	
Max. 5 cm = 3	64. ( )	* 10) : : : Mikrozysten	
Lymphknotenbeteilig.	65. ( )	* 11) : : : Makrozysten	
schleimig	66. ( )	* 12) : : : Epithelprolif. in Gängen & Läppch.	
hart	67. ( )	* 13) : : : Myoepithelproliferation	
gut abgegrenzt	68. ( )	* 14) : : : intakte Basalmembran	
schlecht abgegrenzt	69. ( )	* 15) : : : herdförm. Tumorinfiltrate	
strahlige Wuchsform	70. ( )	* 16) : : : diffuse Tm-Zellinfiltration	
Zysten	71. ( )	* 17) : : : Tm-Infiltr. v. Mamille od. Haut	
Nekrosen	72. ( )	* 18) : : : Tm-Einbruch in Gefäße	
Blutungen	73. ( )	* 19) : : : Tm-Infiltr. von Lymphspalten	
Mamillen/Hautinfiltr.	74. ( )	* 20) : : : Tm-Infiltr. v. Drüsenläppchen	
(ekzematöse Veränd.)		* 21) : : : Tm-Infiltr. v. term. Milchgängen	
		* 22) : : : Anisozytose	
		* 23) : : : transparentes Zytoplasma	
Histologische Diagnose	( : : )	* 24) : : : grobschollige Hyperchromasie	
	75 76 77	* 25) : : : Zellgröße (3 = üb. 3 Lymphozyten)	
-----			
<b>Erläuterungen:</b>			
Zu 25:	0 = kleinzellig = 1 Lymphozyt	* 26) : : : Siegelringzellen	
zu 31:	0 = vakuolisiert, 2 = wie Lymphozyt	* 27) : : : Spindelzellen	
allgem.	0 = nicht vorhanden	* 28) : : : Kernpolymorphie	
wenig	1 = mehr als 30 % Anteil	* 29) : : : Mitosen (-10, -30, üb. 30/Ges.-Feld)	
mittel	2 = mehr als 75 % Anteil	* 30) : : : Path. Kern-Plasma-Relation	
viel	3 = über 80 Anteil	* 31) : : : Chromatingehalt	
		* 32) : : : Nekrosen	
		* 33) : : : Gefäßreichtum	
		* 34) : : : Einblutungen	
		* 35) : : : Schleimsekretion	
		* 36) : : : Mikroverkalkungen	
		* 37) : : : Lymphozyt. Infiltrate	
		* 38) : : : Plasmazellul. Infiltrate	
		* 39) : : : Granulozytäre Infiltrate	
		* 40) : : : BGW-Septen mit 1-3 Tm-Zellreihen	
		* 41) : : : BGW-Septen mit 4-6 Tm-Zellreihen	
		* 42) : : : solid. Tm-Nester mit üb. 6 Zellr.	
		* 43) : : : strahlige Tm-Ausdehnung	
		* 44) : : : Fibrose	
		* 45) : : : einheitl. Tm-Differenzierung	
		* 46) : : : Dos-a-Dos-Phänomene	
		* 47) : : : spaltf. komprim. Gänge	
		* 48) : : : konzent. periduct. Strom. prolif.	
		* 49) : : : Knorpel- od. Knochenanteile	
		* 50) : : : ringf. periduct. Tm-Wachstum	
Fall Nr.: ( : : )			
	78 79 80		

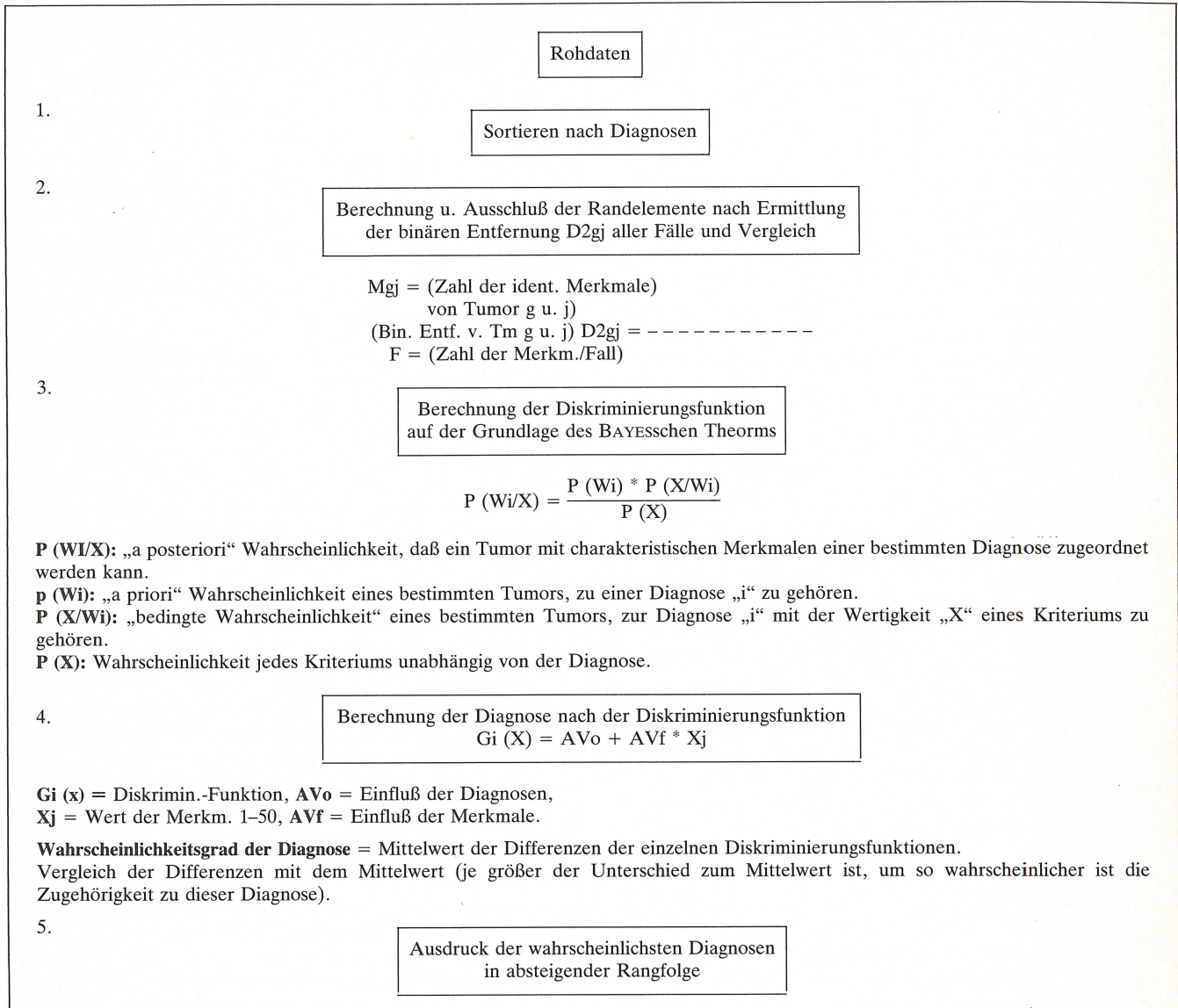


Abb. 2. Schematische Darstellung des Programmablaufes

mittel zur Verfügung gestellt werden. In Anlehnung an die Erfahrungen von IGLESIAS (1983) und PFANNKUCH (1983) sollte die praktische Anwendung eines solchen Computerprogramms am Beispiel der Mammahistologie untersucht werden. Von besonderem Interesse war die Frage, ob gutartige von bösartigen Tumoren gut zu unterscheiden sind und ob eine histologische Subklassifikation mit hinreichender diagnostischer Sicherheit möglich ist.

**Material und Methoden**

239 Mammabiopsate wurden histologisch aufgearbeitet und mit Hämatoxilin-Eosin in herkömmlicher Technik gefärbt. Die Präparate mußten zur Computerauswertung nach 50 vorher festgelegten Kriterien durchgemustert werden. (Abb. 1 Fragebogen, die zusätzlich aufgeführten makroskopischen Kriterien sind späteren Programmweiterungen vorbehalten.) Die Kriterien wurden semiquantifiziert bewertet in: wenig, mittel und viel. Die so gewonnenen Daten erhielten

einen Zahlenschlüssel von 0–3, der anschließend über eine Tastatur in einen Kleinrechner eingegeben wurde (Commodore 64 mit Diskettenlaufwerk 1541).<sup>1)</sup>

Das von IGLESIAS und Mitarb. (1983) bereits ausgiebig überprüfte und in der Computersprache »Microsoft-Basic« verfaßte Diagnoseprogramm wurde für den verwendeten Rechnertyp umgeschrieben und an die spezielle diagnostische Problematik angepaßt. Hinsichtlich der mathematischen Grundlagen sei auf die Ausführungen von IGLESIAS u. Mitarb. (1983 u. 1985) verwiesen.

Der Dateneingabe folgte ein Sortierdurchgang nach Diagnosen. Im Anschluß wurden sogenannte Randelemente der Normalverteilung eliminiert und die Diskriminierungsfunktion sowie die »a posteriori« Wahrscheinlichkeit für jede Einzeldiagnose berechnet. Aufgrund der nur geringen Speicherkapazität von 64 KB war eine Auswertung des statistischen Mehrgruppenfalls nicht ohne weiteres möglich. Für den praktischen Einsatz wurden lediglich 2 Ausprägungen, »vorhanden« und »nicht vorhanden« verrechnet.

<sup>1)</sup> Das Programm ist z. Zt. auf folgenden Rechnern lauffähig: Commodore 64, Commodore 8096, Apple IIe, Wang PC.

Die vom Untersucher gestellte Diagnose für Fall 5 lautete: 1	
# 1. (1) Ca. solid. scirrhosum	= 92,5 %
# 2. (2) Ca. solid. simplex	= 7,2 %
# 3. (20) Invasives lobuläres Ca.	= 0,1 %
Die vom Untersucher gestellte Diagnose für Fall 27 lautete: 11	
# 1. (11) Muzinöses Carcinom	= 100 %
Die vom Untersucher gestellte Diagnose für Fall 81 lautete: 7	
# 1. (7) Komedocarcinom	= 99,2 %
# 2. (14) Tubuläres Ca.	= 0,4 %
# 3. (1) Ca. solid. scirrhosum	= 0,3 %
Die vom Untersucher gestellte Diagnose für Fall 135 lautete: 22	
# 1. (22) Skleros. Adenose	= 60,7 %
# 2. (21) Fibrozyst. Mastopathie	= 39,3 %

Abb. 3. Beispiel eines Diagnosenausdruckes

Das vorliegende Datenmaterial war in 17 histomorphologische Einzeldiagnosen unterteilt. Die histologische Subklassifikation orientierte sich an den Vorschlägen von BÄSSLER (1978). Die ärztlichen Diagnosen dienten als Einteilungskriterien für die Berechnung der Diskriminierungsfunktion. Sie gingen jedoch nicht in die Berechnung der Computerdiagnose ein. Der Arzt, der den Erhebungsbogen bearbeitete, hatte auf die Diagnosenstellung keinen Einfluß.

Der Berechnungsvorgang endete in einem Ausdruck der konventionell erstellten Diagnose und der errechneten Diagnosen (Abb. 3). Um eine zuverlässige mathematisch-statistische Auswertung zu erzielen, waren von jeder Diagnose mindestens 10 Fälle zu fordern. Da von einigen Diagnosen weniger Fälle vorlagen, wurde zur Berechnung der Diskriminierungsfunktion eine Mindestanzahl von 3 Fällen festgelegt. Eine geringere statistische Zuverlässigkeit mußte hier in Kauf genommen werden.

Zur Überprüfung der Zuverlässigkeit dieser Ergebnisse erfolgte eine Kontrolle in einem Großrechner mit dem SPSS-Programm »Diskriminant« Version 9,0 (BEUTEL u. Mitarb. 1980).

### Ergebnisse

Bei der Auswertung des Verfahrens wurden die Ergebnisse in 3 Kategorien eingeteilt:

1. Übereinstimmung: Die Arzt diagnose stimmt mit der »Computerdiagnose« überein.
2. Richtige Diagnosen 2. Ordnung: Arzt diagnose und »Computerdiagnose« stimmen nicht überein, entsprechen jedoch morphologisch ähnlichen Bildern oder sind Mischformen, die Strukturen des 1. errechneten Diagnosevorschlages enthalten.
3. Keine Übereinstimmung: Originaldiagnose und Computerdiagnose sind völlig verschieden.

Die Auswertung nach den oben angeführten Kriterien ergab eine direkte Übereinstimmung der Diagnosen in 77,4 %. Richtigkeit 2. Ordnung konnte in 12,6 % erzielt werden. Keine Übereinstimmung war bei 10 % der untersuchten Präparate festzustellen.

Bei den unterschiedlichen Diagnosen waren mehr oder weniger differente Kriterienkombinationen festzustellen.

Abb. 4 u. 5 zeigen Beispiele von Histogrammen der Häufigkeitsverteilung der 50 Kriterien.

### Diskussion

Bei der rechnergestützten Diagnostik auf der Basis des BAYESschen Theorems wird über die prozentuale Häufigkeit von Merkmalen eine Wertigkeitsberechnung (Diskriminanzkoeffizient) einer bestimmten Diagnose vorgenommen.

Die verwendeten Bewertungskriterien entsprechen den allgemein anerkannten und in der konventionellen Diagnostik üblichen Merkmalen. Die auf Erfahrungswerten beruhende Kriterienkombination wird durch den Rechenvorgang übernommen. Die errechnete Diskriminierungsfunktion bei neu zu analysierenden Gewebsschnitten orientierte sich an den Diagnosen erfahrener Histopathologen. Systematische Fehler bei der Diagnosestellung durch falsche oder subjektive Kriterienkombinationen bei gleichen Untersuchern sind hierbei nicht zu befürchten. (PFANNKUCH 1986, pers. Mitt.)

Tab. 1. Statistische Auswertung der Ergebnisse

Nr.	Diagnose	. Übereinstimmung	: Richtig : 2. Ordnung	: keine : Übereinst.	: Summe
1.	Carcinoma sol. scirrhosum	. 38 (84,4 %)	: 3 (6,6 %)	: 2 (8,8 %)	: 45
2.	Carcinoma sol. simplex	. 12 (35,3 %)	: 15 (44,1 %)	: 7 (20,6 %)	: 34
3.	Carcinoma sol. medullare	. 17 (77,2 %)	: 4 (18,2 %)	: 1 (4,5 %)	: 22
4.	Adenoides Carcinom	. 5 (100 %)	:	:	: 5
5.	Anaplastisches Carcinom	. 3 (60 %)	: 2 (40 %)	:	: 5
6.	Inflammatorisches Ca.	. 2 (100 %)	:	:	: 2
7.	Komedocarcinom	. 10 (71,4 %)	: 1 (7,1 %)	: 3 (21,4 %)	: 14
8.	Solides & cribrif. Ca.	. 10 (76,9 %)	:	: 3 (23,1 %)	: 13
9.	Papilläres & intrazyst. Ca.	. 7 (87,5 %)	:	: 1 (12,5 %)	: 8
10.	Morbus Paget	. 4 (100 %)	:	:	: 4
11.	Muzinöses Carcinom	. 8 (100 %)	:	:	: 7
14.	Tubuläres Carcinom	. 12 (80 %)	: 2 (12,5 %)	: 3 (20 %)	: 15
20.	Invas. lobul. Carcinom	. 13 (81,2 %)	: 3 (17,6 %)	: 1 (6,3 %)	: 16
21.	Fibro-zystische Mastopathie	. 14 (82,3 %)	:	:	: 17
22.	Sklerosierende Adenose	. 3 (100 %)	:	:	: 3
23.	Fibroadenom	. 20 (100 %)	:	:	: 20
24.	Gutart. Papillomatose	. 7 (87,5 %)	:	: 1 (12,5 %)	: 8
Gesamtübereinst. = 215 (90 %)		: 185 (77,4 %)	: 30 (12,6 %)	: 24 (10 %)	: 239

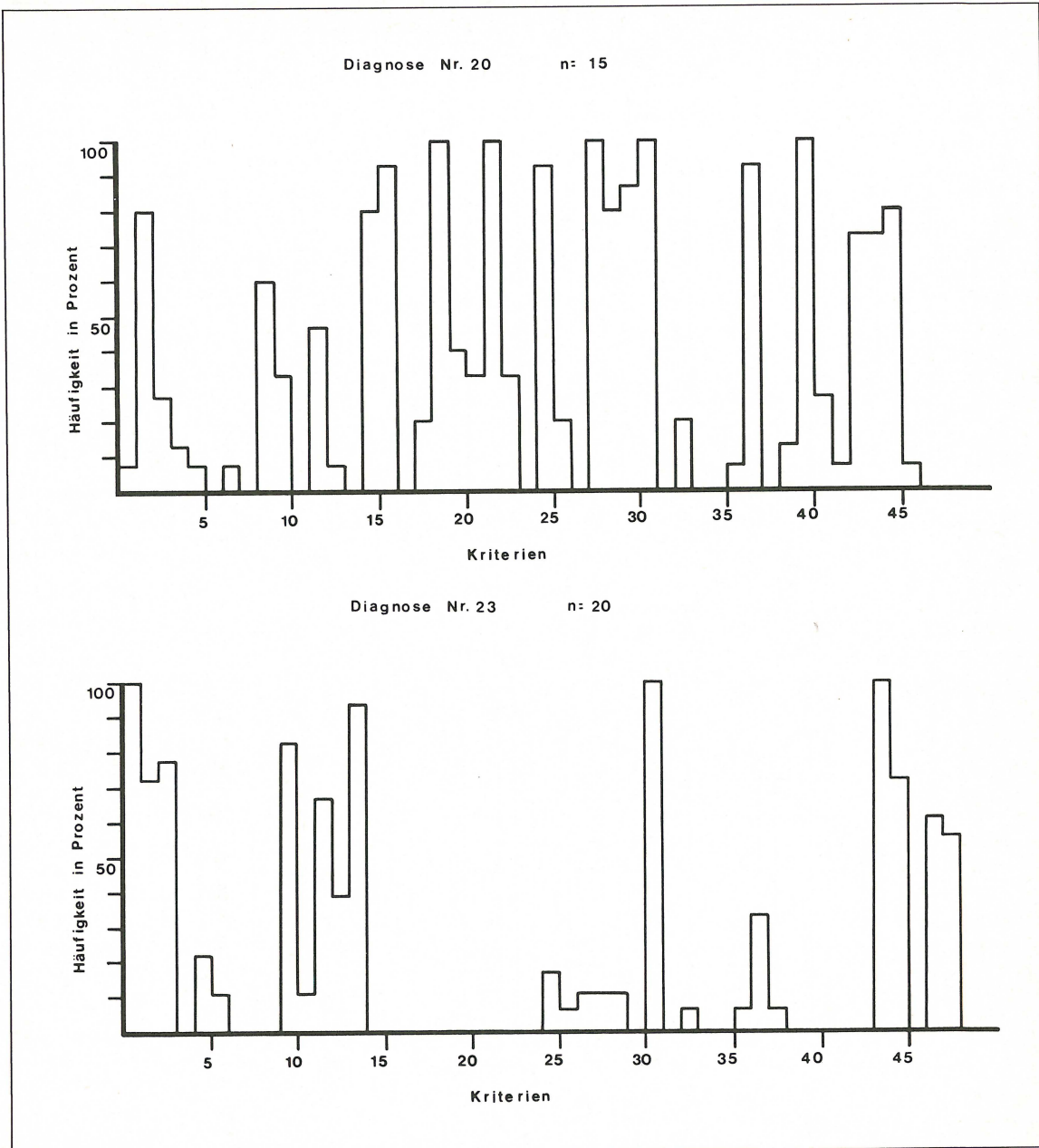


Abb. 4. Häufigkeitsverteilung der einzelnen Kriterien von Diagn. Nr. 20 = invasiv lobuläres Ca. und Diagn. Nr. 23 = Fibroadenom.

Die Häufigkeitsverteilung der Kriterien zeigt Ähnlichkeiten verschiedener Diagnosen. Dies erklärt die zum Teil schwierige Differenzierung mit dem vorliegenden Rechenprogramm (Tab. 1).

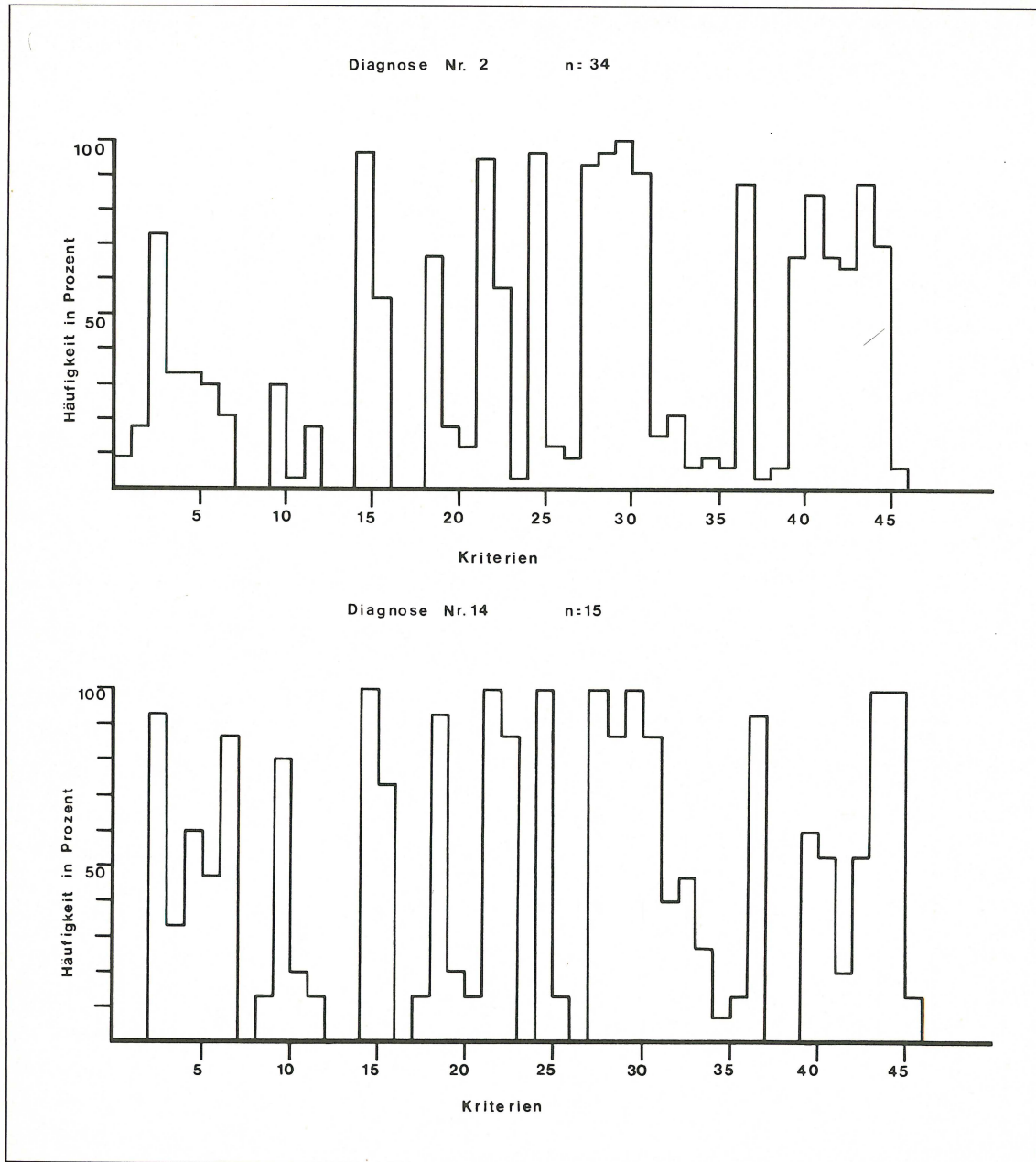
Kriterien, die bei verschiedenen Diagnosen gleichhäufig vorkommen, sind für die Diagnosenberechnung ohne Bedeutung. Ebenso werden solche Merkmale bewertet, die gleichermaßen selten sind oder fehlen.

Bei einem großen Teil der Fälle, die eine Richtigkeit 2. Ordnung aufweisen, handelt es sich um sogenannte Mischformen, die Anteile verschiedener histomorphologischer Subtypen enthalten. Unterschiede bestehen in der Festlegung der Wertigkeit der einzelnen Bestandteile, die sich in der konventionell erstellten Diagnose häufig an der klinischen Prognose orientierte. So wurde bei etwa gleichen Anteilen zweier Subty-

pen diejenige Diagnose bevorzugt, die die schlechtere Prognose besitzt. Auffällig ist die Verwechslungshäufigkeit der Diagnose 2 (Carcinoma solidum simplex) mit der Diagnose 1 (Carcinoma solid. scirrhosum) sowie Diagnose 3 (Ca. solid. medullare) mit der Diagnose 14 (tubuläres Carcinom), die in den Bewertungskriterien große Ähnlichkeiten aufweisen.

Die Verwechslungen der Diagnose 2 (Ca. solid. simplex) mit der Diagnose 14 (Tubuläres Ca.) konnte zum Teil nicht zufriedenstellend geklärt werden. Tubuläre Anteile waren nur in einigen Fällen in geringer Ausprägung vorhanden. Es lagen zumeist ähnliche Häufigkeitsmuster vor, wobei lediglich im Kriterium »atypische tubuläre Proliferationen« (Nr. 7) ein größter Unterschied bestand. Diagnosen, die sich nur in wenigen Kriterien unterscheiden, können offensichtlich nicht scharf genug getrennt werden. Trotzdem war es möglich, mit hinreichender Sicherheit (84,4 % bzw. 77,2 %) scirrhöse und medulläre Carcinome voneinander zu trennen.

Ein nur in geringer Ausprägung vorhandenes Kriterium kann zu hoch eingeschätzt werden und zu einer »Fehlberrech-



nung« Anlaß geben. In einigen Fällen führte die Veränderung nur eines Kriteriums zu einer anderen Diagnosenreihenfolge.

Eine vollständige Übereinstimmung war bei adenoiden, inflammatorischen, muzinösen Carcinomen, beim M. Paget, bei der sklerosierenden Adenose und den Fibroadenomen vorhanden. Mit Ausnahme des Carcinoma solidum simplex, das mit einer Übereinstimmung von nur 35,3% am schlechtesten einzuordnen war, und den anaplastischen Carcinomen mit 60%, lag die Trefferquote über 70%, (9/17) wobei der größte Teil davon über 80% lag (6/9).

Nur in einem Fall war die Trennung von gutartigem und bösartigem Tumor nicht möglich. Es handelte sich um einen Beurteilungsfehler beim Ausfüllen des Fragebogens.

Die mit dem SPSS-Programm »Diskriminant« Version 9 gewonnenen Vergleichsdaten bestätigen die Zuverlässigkeit des vorgestellten »Diagnoseprogramms«.

Die Anwendung des statistischen Mehrgruppenfalles (Kriterien in quantifizierter Ausprägung) ergab keine erhöhte Trefferrate wie erwartet. Mit dem SPSS-Programm konnte eine

Abb. 5. Häufigkeitsverteilung der einzelnen Kriterien von Diagn. Nr. 2 = Ca. solidum simplex und Diagn. Nr. 14 = tubuläres Ca.

Richtigkeit 1. Ordnung in 75% der Diagnosen ermittelt werden. In die Gruppe der Richtigkeit 2. Ordnung wurden 11,8% eingeordnet (Gesamttrefferquote: 86,8%). Die mit dem Personal-Computer ermittelten Resultate weichen hiervon nur unwesentlich ab.

**Schlußfolgerungen**

Nach den bisherigen Erfahrungen ist das vorliegende »Diagnoseprogramm« durchaus geeignet (bei sorgfältig ausgefülltem Fragebogen) einen zuverlässigen Diagnosevorschlag zu liefern. Es ist denkbar, dieses Verfahren bei der Ausbildung in der Histopathologie anzuwenden. Der Auszubildende kann seinen Diagnosevorschlag selbst überprüfen. Die zur Analyse

wichtigen Kriterien müssen stärker beachtet werden, was zu einem zusätzlichen Lerneffekt führt.

Eine völlig automatisierte Diagnostik, wie sie von einigen Autoren erprobt wird (PRESSMANN u. WIED 1979), erscheint nicht sinnvoll. Vollständig »maschinell« erstellte Diagnosen werden den individuellen Eigenheiten von Tumoren nicht ausreichend gerecht. Kein automatisches Verfahren ist zudem in der Lage, die Ausbildung durch einen erfahrenen Diagnostiker am Diskussionsmikroskop zu ersetzen.

Das vorgestellte Verfahren kann hier als Verbesserung und Intensivierung des zusätzlich notwendigen Eigenstudiums und als Selbstkontrolle dienen.

### Danksagung

Herrn PD Dr. GENZ und Herrn Prof. Dr. BUSCH aus der UFK Berlin-Charlottenburg danken wir vielmals für die freundliche Bereitstellung von histologischen Präparaten.

### Literatur

1. ALLER, R. D., S. J. ROBBY, J. W. POITRAS, M. CAMERON, M. C. PRIOR, S. MIAO, G. BARNET: Computer assisted pathology encoding and reporting system (CAPER). *Am. J. Clin. Path.* **68**: 715-720 (1977).
2. BABEL, R.: Das Interace Age Systemhandbuch zum Commodore 64. Interf. Age Verl. München (1983).
3. BÄSSLER: Pathologie der Brustdrüse (In Spez. Pathol. Anat. Bd. II), (1978).
4. BEUTEL, B., H. KÜFFNER, W. SCHUBÖ: SPSS, 8 Statistik-Programm-System für die Sozialwissenschaften (n. Nie H, Hull CH). G. Fischer, Stuttgart, New York, S. 129-139 (1980).
5. ENGLISCH, SZCZEPANOWSKI: Das große Floppy-Buch, Data Becker Verl. (1984).
6. FATTU, J. M., E. PATRICK, W. SUTTON: Thyroid disorders: automatic diagnosis in CONSULT I. *Comput. Biol. Med.* **12**: 285-293 (1982).
7. FOTTER, R.: Computerunterstützte Diagnostik von Knochentumoren und tumorähnlichen Skeletterkrankungen im Kindesalter. *RÖFO* **136**: 627-634 (1982).
8. IGLESIAS, J. R., M. SANCHEZ, A. SENDRA, A. MOHNHAUPT: Computer model of archive and diagnosis of brain tumors based on the WHO classification. *EDV in Med. & Biol.* **14**: 40-44 (1983).
9. IGLESIAS J. R., E. KAZNER, C. ARUFFO, J. ESPARZA: CT-Diagnosis of Brain Tumors: Semiautomatcal type-specific Diagnosis with the help of a Personal Computer. Mathematical fundaments and practical Application. In Lemke HU., Rhodes ML., Jaffee CC., Felix R. (eds) *Computerized assisted Radiology*, Springer Berlin: 443-448 (1985).
10. IGLESIAS J. R., E. KAZNER, C. ARUFFO, J. ESPARZA: A Model of Semiautomatcal Type-Specific Diagnosis of Brain Tumors. Mathematical Fundaments and Practical Application. *Br. J. Radiol.* (im Druck).
11. JACOB, W.: Elektronische Datenverarbeitung in der Pathologie. *Münch. Med. Wschr.* **114**: 737-740 (1972).
12. KOLLER S., G. WAGNER: Handbuch der med. Dokumentation und Datenverarbeitung. Schattauer Verlag, Stuttgart (1975).
13. PFANNKUCH, F., H. J. SCHOOLMANN, J. IGLESIAS, C. H. BRINNEL: Application of the personal computer in routine diagnosis and preparation of reports of thyroid cytology. (IX. Europ. Congr. Pathology, Hamburg, Sept. 1983), *Path. Res. Pract.* **178**: 155 (1983).
14. PRESSMANN N., G. WIED ed.: The Automation of Cancer Cytology and Cell Image Analysis. *Tutorials of Cytology Chicago* (1979).
15. SCHOOLMANN H. J., F. PFANNKUCH, J. IGLESIAS, C. H. BRINNEL: Mikrocomputer als Hilfsmittel in der zytologischen Diagnostik und bei Erstellung von Befundberichten, *Verh. Dtsch. Ges. Path.* **68**: 428 (1984).
16. SONNENBERG, A.: Sequential iteration of Bayesian Formula by pocket calculator and its use in clinical routine. *Comput. Biol. Med.* **12**: 357-360 (1982).
17. STEINHAUSEN D., K. LANGER: Einführung in Methoden und Verfahren der automatischen Klassifikation. Verlag W. de Gruyter Berlin (1977).
18. Commodore 64 Programmers Referenc Guide, Commodore Business Machines Wayne USA (1982).

Eingegangen am 28. Juli 1986.

Anschrift des Autors: Dr. Hans Joachim Hahn, Zentralkrankenhaus St. Jürgenstr. (Frauenklinik), St. Jürgenstr., D-2800 Bremen 1.

# Identification and interpretation of effects in two-period crossover designs

Hartmut Hecker

## Summary

*In the two-period crossover design the second treatment period generates four types of additional effects: period effect, carry-over (residual) effect, treatment by period interaction and treatment by treatment interaction. It is shown that in the presence of non-zero interaction terms the second treatment period cannot be used for the analysis of the direct treatment effect. However, the period and the carryover effect can be eliminated by using the simple, respectively the extended crossover design. Univariate test statistics are derived for the extended crossover design.*

## Zusammenfassung

*Durch die zweite Behandlungsperiode im Zwei-Perioden-Crossover-Design werden vier Arten von Effekten zusätzlich erzeugt: Der Phaseneffekt, der Carryover-(Residual-)Effekt und die Wechselwirkungen zwischen Behandlung und Periode und zwischen den Behandlungen untereinander. Es wird gezeigt: Sind die Wechselwirkungen ungleich Null, so kann die zweite Behandlungsperiode nicht zur Analyse des direkten Behandlungseffektes verwendet werden. Periodeneffekt und Carryover-Effekt hingegen werden bei der Anwendung des einfachen bzw. des erweiterten Crossover-Designs eliminiert. Für den erweiterten Crossover-Versuch werden univariate Tests der Modellparameter hergeleitet.*

## 1. Introduction

The crossover design, often used in clinical trials to compare two or more treatments, gains much of its attraction from the intuitive belief that it should be more efficient to compare the reactions of one patient under different conditions rather than to compare the reactions of different patients under different treatments. Indeed, the random variations within a subject are often smaller than those between the subjects, and if this is the case the crossover design results in a significance test with a greater power than the randomized block design. More precise relations between these two designs are given by BROWN [2]. On the other hand, using each subject »as its own control« generally bares some uncertainties about the appropriateness

of the control: after the first treatment period the patient is »no longer the same« as before. This may be due to the treatment in the first period itself, or to a general time effect, and these changes may invalidate the comparison of the treatments in the two or more periods.

The aim of this paper is to describe in detail all possible effects which may arise in the two-period crossover design and thus to clarify the conditions, under which this design is applicable to compare two treatments. Thereby, some of the ideas of HILLS and ARMITAGE [4] are formulated in a somewhat more precise way, and simple test statistics are derived to test simple hypotheses in the two-period crossover design with four treatment-sequence groups.

## 2. Parameter definition and interpretation

To understand the effects arising in a two-treatment two-period design it is helpful to compare the location parameters of all treatment groups with the no-treatment and one-treatment situation. Generally, let  $\mu_{ij}$  denote the location parameter of the  $i$ -th treatment group in period  $j$  [ $j = 1, 2$ ]. [In most applications  $\mu_{ij}$  are the expected values of normally distributed random variables, or the logits of the probabilities of success.] The two treatments are denoted by A and B and »no treatment« is indicated by »-«. The following Table 1 shows all  $3^2 = 9$  possible treatment combinations together with their location parameters, and a parameter transformation which is helpful to show all possible effects.

Table 1. The complete two-period crossover design

Treatment	Period 1	Period 2
1: AB	$\mu_{11} = b_1 + a_{A1}$	$\mu_{12} = b_2 + a_{B2} + d_A + e_{AB}$
2: BA	$\mu_{21} = b_1 + a_{B1}$	$\mu_{22} = b_2 + a_{A2} + d_B + e_{BA}$
3: AA	$\mu_{31} = b_1 + a_{A1}$	$\mu_{32} = b_2 + a_{A2} + d_A + e_{AA}$
4: BB	$\mu_{41} = b_1 + a_{B1}$	$\mu_{42} = b_2 + a_{B2} + d_B + e_{BB}$
5: A-	$\mu_{51} = b_1 + a_{A1}$	$\mu_{52} = b_2 + d_A$
6: -A	$\mu_{61} = b_1$	$\mu_{62} = b_2 + a_{A2}$
7: B-	$\mu_{71} = b_1 + a_{B1}$	$\mu_{72} = b_2 + d_B$
8: -B	$\mu_{81} = b_1$	$\mu_{82} = b_2 + a_{B2}$
9: --	$\mu_{91} = b_1$	$\mu_{92} = b_2$



The transformed parameters may be interpreted as follows:  
 $b_j$ : baseline parameter of period  $j$  ( $j = 1, 2$ )  
 $a_{Aj}$ ,  $a_{Bj}$ : effect of treatment A, B in period  $j$  ( $j = 1, 2$ )  
 $d_A$ ,  $d_B$ : carryover (residual) effect of treatment A, B  
 $e_{AB}$ : interaction of treatment sequence AB (etc.)

Note that  $\mu_{11} = \mu_{31} = \mu_{51}$ ,  $\mu_{21} = \mu_{41} = \mu_{71}$  and  $\mu_{61} = \mu_{81} = \mu_{91}$ , since in the first period the treatments are equal in the corresponding groups. Thus there are only 12 different location parameters, in concordance with the 12 transformed parameters.

The parameter interpretation is illustrated in the following schematic figures 1-3. Herein we assume that the location parameters are defined as the expected differences of a criterion variable  $Y$  between the end and the beginning of the treatment in period 1 and period 2 respectively, and that the treatment-periods are separated by a treatment-free »wash-out« period.

Note that the term »effect« is meaningful only, when two or more different location parameters are compared with each other, and their differences are due to a well defined difference in the experimental situation. Therefore, in the parameter transformation described here, the parameters  $b_j$  are not »effects« and therefore called »baseline parameters« for period 1 and 2. The effects  $a_{Aj}$  and  $a_{Bj}$  result from the comparison with the no-treatment situation in both periods. To identify the carryover effects  $d_A$  and  $d_B$  one has to regard the location parameters in the second period, in which no treatment is assigned, and to compare the treatment and no-treatment situation of treatment A (respectively B) in the first period with each other. Finally, the interaction effect  $e_{AB}$  results from the comparison of the location parameter in the second period in the treatment sequence AB with the sequences A- and -B (after elimination of all other effects and baseline parameters).

Altogether, the parameters  $b$ ,  $a$ ,  $d$  and  $e$ , implicitly defined in Table 1, may be explicitly defined as follows:

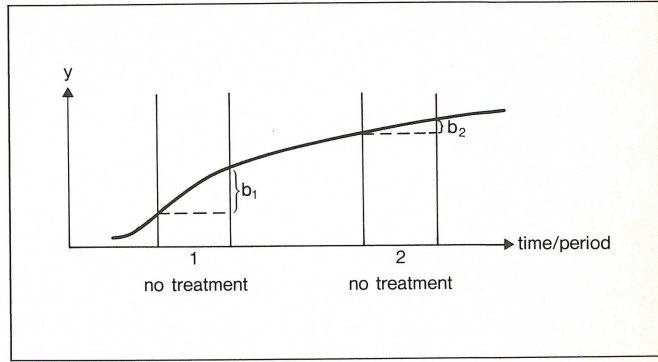


Fig. 1. Period-baseline parameters.

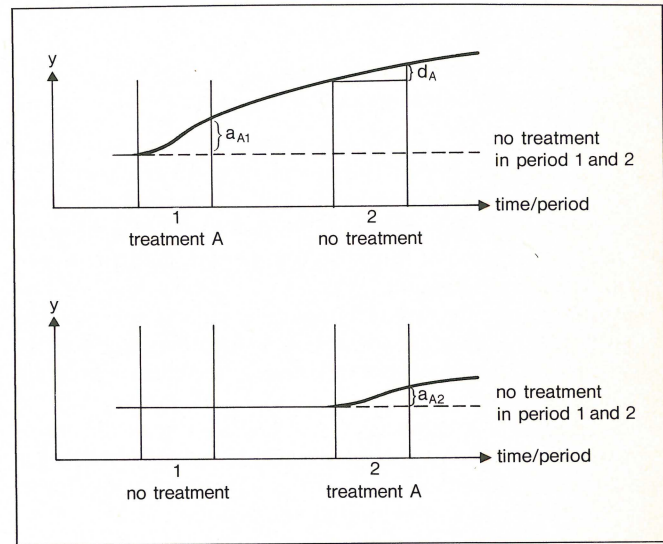


Fig. 2. Treatment- and carryover effect (in the absence of period effects).

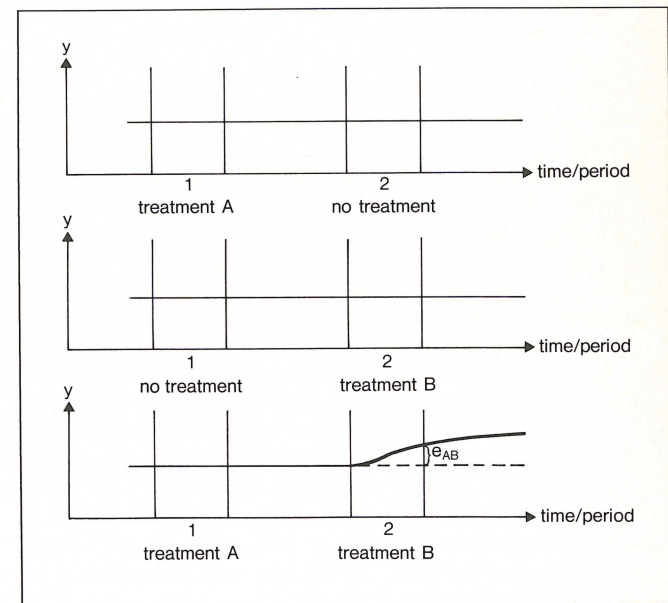


Fig. 3. Interaction of treatment sequence AB (in the absence of treatment, period and carryover-effects).

Let

$$(1) \begin{cases} \mu_{A1}: = \mu_{11} = \mu_{31} = \mu_{51} \\ \mu_{B1}: = \mu_{21} = \mu_{41} = \mu_{71} \\ \mu_{-1}: = \mu_{61} = \mu_{81} = \mu_{91} \end{cases}$$

Then

$$(2) \begin{cases} a_{A1} = \mu_{A1} - \mu_{-1} \\ a_{B1} = \mu_{B1} - \mu_{-1} \\ a_{A2} = \mu_{62} - \mu_{92} \\ a_{B2} = \mu_{82} - \mu_{92} \end{cases}$$

$$(3) \begin{cases} b_1 = \mu_{-1} \\ b_2 = \mu_{92} \end{cases}$$

$$(4) \begin{cases} d_A = \mu_{52} - \mu_{92} \\ d_B = \mu_{72} - \mu_{92} \end{cases}$$

$$(5) \begin{cases} e_{AB} = \mu_{12} - \mu_{52} - \mu_{82} + \mu_{92} \\ e_{BA} = \mu_{22} - \mu_{72} - \mu_{62} + \mu_{92} \\ e_{AA} = \mu_{32} - \mu_{52} - \mu_{62} + \mu_{92} \\ e_{BB} = \mu_{42} - \mu_{72} - \mu_{82} + \mu_{92} \end{cases}$$

Some further parameter transformations are useful for interpretation:

The baseline parameters  $b_j$  may be different for period 1 and 2, so they can be split into the »general mean«  $m$  and a »period effect«  $b$ , which measures the departure from  $m$ :

$$(6) \quad \begin{cases} b_1 = m + b \\ b_2 = m - b \end{cases}$$

Similarly, as the treatment effects may be different in the two periods, one may define the »main« treatment effects  $a_A$  and  $a_B$  and departures from the general effect in period 1 and 2:

$$(7) \quad \begin{cases} a_{A1} = a_A + c_A \\ a_{A2} = a_A - c_A \\ a_{B1} = a_B + c_B \\ a_{B2} = a_B - c_B \end{cases}$$

We call the effects  $c_A$  and  $c_B$ , implicitly defined by (7), the »interaction effects« of treatment A by period, and treatment B by period, respectively.

To summarize so far we remark that, apart from the period and the carryover effect, there may arise two types of effects in the two-period crossover design that are added to the treatment effect:

1. Treatment by treatment interaction ( $e_{AB}, e_{BA}, e_{AA}, e_{BB}$ )
2. Treatment by period interaction ( $c_A, c_B$ ).

Note that both effects are clearly separated from the usually regarded carryover effect. The treatment by treatment interaction indicates, that the treatment in the first period may have influences on the experimental unit, that result in a different effect of the treatment in the second period, compared with the no-treatment situation.

The treatment by period interaction reflects the possible time dependent effect of a treatment, which may e.g. arise in situations in which the »natural« course of an illness is progressive or regressive.

In this case, a long »wash-out« period between the treatment periods, which should eliminate possible carryover effects (and treatment by treatment interaction) may even augment the treatment by period interaction.

The treatment by period effect is incorporated in the models of KERSHNER and FEDERER [5], but is not mentioned in the important papers of GRIZZLE [3], ZIMMERMANN and RAHLFS [9] and LASKA and MEISNER [6]. HILLS and ARMITAGE [4] describe three reasons for interaction between treatment and period, that correspond to the effects d, c, and e, for which a precise definition was given here.

### 3. Analysis of the direct treatment effect

We now regard experiments in which the main aim is to analyze the treatment effect in the first period (the »direct effect«), and in which the crossover design is applied to reduce the variance for the estimation of this effect.

Clearly, the crossover design can be useful in this situation only when the observations of both periods may be used for the estimation of the direct effect. In our notation, this condition is equivalent to the existence of linear combinations  $\sum w_{ij}\mu_{ij}$  of the location parameters  $\mu_{ij}$  with non-zero weights  $w_{ij}$  for the second period, that are equal to the parameters  $a_{A1}, a_{B1}$  or their difference. We shall now derive for some special designs conditions, under which the existence of such linear combinations is given. In each case we will start with a parameter transformation of the given location parameters  $\mu_{ij}$ , that yields a set of parameters which is uniquely defined by the  $\mu_{ij}$ , so that each parameter of this set is estimable. An interpretation is then found by expressing them in terms of the parameters of the complete two-period design defined in the previous section.

### The simple two-period crossover

If the only treatment sequences applied are AB and BA, we introduce the following parameter transformation (Tab. 2):

Table 2. The simple two-period crossover

Treatment	Period 1	Period 2
1: AB	$\mu_{11} = \mu + \alpha + \beta + \gamma$	$\mu_{12} = \mu - \alpha - \beta + \gamma$
2: BA	$\mu_{21} = \mu - \alpha + \beta - \gamma$	$\mu_{22} = \mu + \alpha - \beta - \gamma$

The parameters  $\mu, \alpha, \beta$  and  $\gamma$  are related to the location parameters  $\mu_{ij}$  by

$$(8) \quad \begin{cases} \mu = \frac{1}{4} (\mu_{11} + \mu_{12} + \mu_{21} + \mu_{22}) \\ \alpha = \frac{1}{4} (\mu_{11} - \mu_{12} - \mu_{21} + \mu_{22}) \\ \beta = \frac{1}{4} (\mu_{11} - \mu_{12} + \mu_{21} - \mu_{22}) \\ \gamma = \frac{1}{4} (\mu_{11} + \mu_{12} - \mu_{21} - \mu_{22}) \end{cases}$$

We may think of  $\mu$  being interpreted as a general mean,  $\alpha$  the treatment effect,  $\beta$  the period effect and  $\gamma$  as an interaction between treatment and period. However, if one inserts equation (8) into the equations of Table 1 and the reparametrisation (6) and (7), we arrive at

$$(9) \quad \begin{cases} \mu = m + \frac{1}{2} (a_A + a_B) + \frac{1}{4} (d_A + d_B) + \frac{1}{4} (e_{AB} + e_{BA}) \\ \alpha = \frac{1}{2} (a_A - a_B) - \frac{1}{4} (d_A - d_B) - \frac{1}{4} (e_{AB} - e_{BA}) \\ \beta = b + \frac{1}{2} (c_A + c_B) - \frac{1}{4} (d_A + d_B) - \frac{1}{4} (e_{AB} + e_{BA}) \\ \gamma = \frac{1}{2} (c_A - c_B) + \frac{1}{4} (d_A - d_B) + \frac{1}{4} (e_{AB} - e_{BA}) \end{cases}$$

Thus, all parameters are confounded with the carryover effects  $d_A, d_B$  and the treatment by treatment interactions  $e_{AB}, e_{BA}$ . Moreover,  $\beta$  is additionally confounded with the treatment by period interactions ( $c_A, c_B$ ), and  $\gamma$  is a combination of all these effects.

In this context, a treatment effect is well defined only by comparing both treatments with each other (rather than comparing them with the no-treatment situation). For the first period, this results in defining the »direct effect«:

$$(10) \quad \begin{aligned} \alpha^* &= \frac{1}{2} (\mu_{11} - \mu_{21}) \\ &= \alpha + \gamma \end{aligned}$$

It may be attained from the second period as well, if

$$(11) \quad \gamma = 0$$

Then we have

$$\begin{aligned} \frac{1}{2} (-\mu_{12} + \mu_{22}) &= \alpha - \gamma \\ &= \alpha + \gamma \quad (\text{since } \gamma = 0) \\ &= \alpha^* \end{aligned}$$

so  $\alpha^*$  may be estimated by

$$(12) \quad \hat{\alpha}^* = w_1 \frac{1}{2} (\hat{\mu}_{11} - \hat{\mu}_{21}) + w_2 \frac{1}{2} (\hat{\mu}_{22} - \hat{\mu}_{12})$$

where the  $\hat{\mu}_{ij}$  are estimates for the location parameters  $\mu_{ij}$ , and the weights  $w_1$  and  $w_2$  sum up to 1. As is indicated by (9), in

practical applications the condition  $\gamma = 0$  affords that all the following effects are zero: Interaction treatment by period ( $c_A, c_B$ ), carryover effects ( $d_A, d_B$ ) and treatment by treatment interaction ( $e_{AB}, e_{BA}$ ).

One may hope that after adding new treatment sequences to the simple two-period design, some of these effects can be identified and therefore, even if present, do not make the second period unusable for analyzing the direct effect. To this aim, we regard the »extended« two-period crossover design, together with a parameter transformation, in Table 3:

Table 3. Extended two-period crossover

Treatment	Period 1	Period 2
1: AB	$\mu_{11} = \mu + \alpha + \beta + \gamma$	$\mu_{12} = \mu - \alpha - \beta + \gamma + \delta - \epsilon$
2: BA	$\mu_{21} = \mu - \alpha + \beta - \gamma$	$\mu_{22} = \mu + \alpha - \beta - \gamma - \delta - \epsilon$
3: AA	$\mu_{31} = \mu + \alpha + \beta + \gamma$	$\mu_{32} = \mu + \alpha - \beta - \gamma + \delta + \epsilon$
4: BB	$\mu_{41} = \mu - \alpha + \beta - \gamma$	$\mu_{42} = \mu - \alpha - \beta + \gamma - \delta + \epsilon$

Note that there are now 6 different location parameters (since  $\mu_{11} = \mu_{31}$  and  $\mu_{21} = \mu_{41}$ ), corresponding to the 6 transformed parameters. These are expressed in terms of the  $\mu_{ij}$  by (13):

$$\begin{aligned}
 \mu &= \frac{1}{8} (\mu_{11} + \mu_{12} + \mu_{21} + \mu_{22} + \mu_{31} + \mu_{32} + \mu_{41} + \mu_{42}) \\
 \alpha &= \frac{1}{8} (\mu_{11} - \mu_{12} - \mu_{21} + \mu_{22} + \mu_{31} + \mu_{32} - \mu_{41} - \mu_{42}) \\
 \beta &= \frac{1}{8} (\mu_{11} - \mu_{12} + \mu_{21} - \mu_{22} + \mu_{31} - \mu_{32} + \mu_{41} - \mu_{42}) \\
 \gamma &= \frac{1}{8} (\mu_{11} + \mu_{12} - \mu_{21} - \mu_{22} + \mu_{31} - \mu_{32} - \mu_{41} + \mu_{42}) \\
 \delta &= \frac{1}{4} (\mu_{12} - \mu_{22} + \mu_{32} - \mu_{42}) \\
 \epsilon &= \frac{1}{4} (-\mu_{12} - \mu_{22} + \mu_{32} + \mu_{42})
 \end{aligned}
 \tag{13}$$

where  $\mu_{11} + \mu_{31}$  and  $\mu_{21} + \mu_{41}$  may be replaced by any corresponding weighted sum with weights adding up to 1.

Again, one may first think of the transformed parameters being interpreted as follows:

- $\mu$  = general mean
- $\alpha$  = treatment effect
- $\beta$  = period effect
- $\gamma$  = interaction treatment by period
- $\delta$  = carryover effect
- $\epsilon$  = interaction (treatment AB or BA) by (treatment AA or BB)

But if we insert equation (13) into the equation of Table 1 and the reparametrisation (6) and (7) we get the following full interpretation of the parameters:

$$\begin{aligned}
 \mu &= m + \frac{1}{2} (a_A + a_B) + \frac{1}{4} (d_A + d_B) \\
 &\quad + \frac{1}{8} (e_{AB} + e_{BA} + e_{AA} + e_{BB}) \\
 \alpha &= \frac{1}{2} (a_A - a_B) \\
 &\quad + \frac{1}{8} (-e_{AB} + e_{BA} + e_{AA} - e_{BB}) \\
 \beta &= b + \frac{1}{2} (c_A + c_B) - \frac{1}{4} (d_A + d_B) \\
 &\quad - \frac{1}{8} (e_{AB} + e_{BA} + e_{AA} + e_{BB})
 \end{aligned}
 \tag{14}$$

$$\begin{aligned}
 \gamma &= \frac{1}{2} (c_A - c_B) \\
 &\quad + \frac{1}{8} (e_{AB} - e_{BA} - e_{AA} + e_{BB}) \\
 \delta &= \frac{1}{2} (d_A - d_B) \\
 &\quad + \frac{1}{4} (e_{AB} - e_{BA} + e_{AA} - e_{BB}) \\
 \epsilon &= \frac{1}{4} (-e_{AB} - e_{BA} + e_{AA} + e_{BB})
 \end{aligned}$$

It follows that the parameters  $\alpha$  and  $\gamma$  are no longer confounded with the carryover effects  $d_A$  and  $d_B$ , but all the parameters  $\mu, \alpha, \beta, \gamma$  and  $\delta$  are confounded with the treatment by treatment interaction terms  $e_{AB}, e_{BA}, e_{AA}$  and  $e_{BB}$ , and  $\beta$  is still confounded with the treatment by period interaction  $c_A$  and  $c_B$ , and the carryover effects  $d_A$  and  $d_B$ .

The »direct effect« of the treatment is again defined by comparing the location parameters for different treatments in the first period:

$$\begin{aligned}
 (15) \quad \alpha^* &= \frac{1}{2} (\mu_{11} - \mu_{21}) = \frac{1}{2} (\mu_{31} - \mu_{41}) \\
 &= \alpha + \gamma
 \end{aligned}$$

It may be attained from the second period as well, if

$$(16) \quad \gamma = 0,$$

for in this case we see from Table 3:

$$\begin{aligned}
 \frac{1}{4} (-\mu_{12} + \mu_{22} + \mu_{32} - \mu_{42}) &= \alpha - \gamma \\
 &= \alpha + \gamma \text{ (since } \gamma = 0) \\
 &= \alpha^*.
 \end{aligned}$$

So  $\alpha^*$  may be estimated by

$$\begin{aligned}
 (17) \quad \hat{\alpha}^* &= w_1 (w_{11}\hat{\mu}_{11} - w_{21}\hat{\mu}_{21} + w_{31}\hat{\mu}_{31} - w_{41}\hat{\mu}_{41}) \\
 &\quad + w_2 \frac{1}{4} (-\hat{\mu}_{12} + \hat{\mu}_{22} + \hat{\mu}_{32} - \hat{\mu}_{42})
 \end{aligned}$$

where the weights  $w_i, w_{ij}$  are positive with

$$(18) \quad \begin{cases} w_{11} + w_{31} = w_{21} + w_{41} = \frac{1}{2} \\ w_1 + w_2 = 1, \end{cases}$$

and the  $\hat{\mu}_{ij}$  are estimates for the location parameters  $\mu_{ij}$ .

From (14) we see that the condition  $\gamma = 0$ , under which this estimate of the direct effect  $\alpha^*$  is attainable implies (in practice) the absence of treatment by period and treatment by treatment interaction ( $c_A, c_B, e_{AB}, e_{BA}, e_{AA}, e_{BB}$ ). However, the carryover effects  $d_A, d_B$  may be present in this design without making the second period unusable for the analysis of the direct effect.

We now return to the complete two-period two-treatment crossover design to see under which conditions the second period is usable for the analysis of the direct effect in this case. To this aim we rewrite the complete design of Table 1 using the transformations (6) and (7) (Tab. 4, next page).

Note that the treatment effects  $a_A, a_B$  in the second period are always combined with the negative treatment by period interaction terms  $c_A, c_B$ , that is  $a_A - c_A$  and  $a_B - c_B$ , compared with the terms  $a_A + c_A$  and  $a_A + c_B$  in the first period, and  $c_A$  and  $c_B$  never appear alone or in connection with other terms. Therefore there is no linear combination of the location parameters  $\mu_{i2}$  of the second period that is equal to  $c_A$  or  $c_B$  or to the direct effects  $a_A + c_A$  or  $a_B + c_B$  or their difference. Thus, even in the complete two-period crossover design, the

Table 4. The complete two-period crossover design

Treatment	Period 1	Period 2
1: AB	$\mu_{11} = m + b + a_A + c_A$	$\mu_{12} = m - b + a_B - c_B + d_A + e_{AB}$
2: BA	$\mu_{21} = m + b + a_B + c_B$	$\mu_{22} = m - b + a_A - c_A + d_B + e_{BA}$
3: AA	$\mu_{31} = m + b + a_A + c_A$	$\mu_{32} = m - b + a_A - c_A + d_A + e_{AA}$
4: BB	$\mu_{41} = m + b + a_B + c_B$	$\mu_{42} = m - b + a_B - c_B + d_B + e_{BB}$
5: A-	$\mu_{51} = m + b + a_A + c_A$	$\mu_{52} = m - b + a_A - c_A + d_A$
6: -A	$\mu_{61} = m + b$	$\mu_{62} = m - b + a_A - c_A$
7: B-	$\mu_{71} = m + b + a_B + c_B$	$\mu_{72} = m - b + a_B - c_B + d_B$
8: -B	$\mu_{81} = m + b$	$\mu_{82} = m - b + a_B - c_B$
9: --	$\mu_{91} = m + b$	$\mu_{92} = m - b$

second period is not usable for the estimation of the direct effect, if treatment by period interactions are present. This is of course a quite evident result, since if the treatment effects are period-dependent in an unspecified way, no informations of the second period results can be transformed to the treatment effects of the first period.

However, the treatment by treatment interactions  $e_{AB}$ ,  $e_{BA}$ ,  $e_{AA}$  and  $e_{BB}$  and the carryover effects  $d_A$  and  $d_B$  can be identified in this design, as was shown in (5). But the presence of treatment by treatment interaction is still crucial in this complete design: In this case only the 6th and 8th treatment sequence (-A and -B) can be used to estimate the direct effect  $a_A - a_B$  (if  $c_A = c_B = 0$ ) from the observations of the second period, for in the treatment sequences 1 to 4 (AB, BA, AA, BB), the treatment effects are confounded with the treatment by treatment interaction, and in the remaining sequences 5, 7 and 9 they do not appear at all. So with only the two complete sequences (6: -A and 8: -B) and the data of the first period for analyzing the direct effect, it does not make any sense to use a crossover design. Also, if the treatment by treatment (and treatment by period) interactions are zero, the carryover effects  $d_A$  and  $d_B$  are already eliminated by the treatment sequences 1 to 4, so it is not necessary to add further sequences to the design for reason of estimability of the direct effect.

On the other hand, if one is interested in the treatment effect as a comparison with the no treatment situation in the first period (rather than the difference of the two treatments in the first period), that is in  $a_A$  and  $a_B$  rather than  $a_A - a_B$ , further sequences are to be added to the »extended« two-period design and one may use the complete two-period design.

Although the complete two-period crossover design has been introduced primarily for better understanding and defining all possible effects arising in treatment sequences, it may also be regarded as a three-treatment two-period crossover design. However, if none of the three treatments is a placebo or »no« treatment, one should use different parameter transformations, since the interpretation of some of the parameters is misleading in this case.

With these remarks concerning the complete two-period crossover design in mind, we may summarize the conditions for the applicability of the crossover design with both periods for the analysis of the direct effect in Table 5. In this table we neglect the fact that the conditions can be formally weakened. For example, in the simple two-period design one has only to demand:  $c_A = c_B$ ,  $d_A = d_B$  and  $e_{AB} = e_{BA}$  or, even less:

$$\frac{1}{2} (c_A - c_B) + \frac{1}{4} (d_A - d_B) + \frac{1}{4} (e_{AB} - e_{BA}) = 0 \text{ (see (9))}$$

But in practice these conditions are usually fulfilled only if the

Table 5. Two-period crossover designs and analysis of the direct effect

Design	Conditions for applicability of both periods of the design for the analysis of the direct effect		
	Treatment by period ( $c_A, c_B$ )	Treatment by treatment ( $e_{AB}, e_{BA}, e_{AA}, e_{BB}$ )	Carryover ( $d_A, d_B$ )
Simple: AB, BA	zero	zero ( $e_{AB}, e_{BA}$ )	zero
Extended: AB, BA, AA, BB	zero	zero	no restriction
Complete: AB, BA, AA, BB, A-, -A, B-, -B, --	zero	zero	no restriction

stronger versions of Table 5 can be assumed to hold in the experimental situation.

#### 4. Hypothesis testing

Although there is an extensive literatur on the analysis of the two-period crossover design, some remarks should be added on this topic. Thereby we restrict our attention to the case of normally distributed observations with equal, but unknown and arbitrary covariance matrices in each treatment sequence.

The «simple» design, defined in Table 2 may be analyzed following the lines of ZIMMERMANN and RAHLFS [9] or SCHNEIDER [7]. However, the first method is not applicable for the extended design defined in Table 3, and the second method mentioned needs some further explications for this case. To show this we begin with the MANOVA-approach of ZIMMERMANN and RAHLFS. This method assumes that the test of interest can be written as

$$H_0 : A (\mu_{ij})_{i,j} B = 0$$

with proper matrices A and B, and  $(\mu_{ij})_{i,j}$  the matrix of the expected values for period j in the treatment sequence i. But it can be easily shown that neither the test

$$H_{01} : \gamma = \frac{1}{8} (\mu_{11} + \mu_{12} - \mu_{21} - \mu_{22} + \mu_{31} - \mu_{32} - \mu_{41} + \mu_{42}) = 0$$

nor the test

$$H_{02} : \alpha = \frac{1}{8} (\mu_{11} - \mu_{12} - \mu_{21} + \mu_{22} + \mu_{31} + \mu_{32} - \mu_{41} - \mu_{42}) = 0$$

can be put in this form. So the MANOVA-model needs to be generalized to cover the extended two-period crossover design (an extension of the MANOVA-model following this line is in preparation).

For the second method described by SCHNEIDER one has to find estimates  $s_{\hat{\alpha}}^2, s_{\hat{\gamma}}^2$  etc. of the variance of the parameter estimates  $\hat{\alpha}, \hat{\gamma}$  etc. which are  $\chi^2$ -distributed. However, exact solutions are not attainable in general if the variance components of each treatment sequence of the parameter estimates are different (for instance  $\text{var}(\hat{\mu}_{11} + \hat{\mu}_{12}) \neq \text{var}(\hat{\mu}_{31} - \hat{\mu}_{32})$  if the correlation of the observations in the two periods is not zero). So we use the approximation of WELCH [8] to derive hypothesis tests for the parameters in this case.

Let  $Y_{ijk}$  ( $i = 1, \dots, 4; j = 1, 2; k = 1, \dots, n_i$ ) denote the random variable corresponding to the k-th experimental unit of treatment sequence i and period j, and

$$(19) \quad \bar{Y}_{ij} = \frac{1}{n_i} \sum_{k=1}^{n_i} Y_{ijk} \quad (i = 1, \dots, 4; j = 1, 2)$$

Furthermore, let

$$(20) \quad \sigma_{ij_2} = \text{cov}(Y_{ij_1k}, Y_{ij_2k}) \quad (j_1, j_2 = 1, 2)$$

denote the elements of the covariance matrix of the paired observations (independent of  $i$  and  $k$ ).

Following (13), we define as parameter estimators

$$(21) \quad \begin{aligned} \hat{\mu} &= \frac{1}{8} \sum_{i=1}^4 (\bar{Y}_{i_1} + \bar{Y}_{i_2}) \\ \hat{\alpha} &= \frac{1}{8} [(\bar{Y}_{11} - \bar{Y}_{12}) - (\bar{Y}_{21} - \bar{Y}_{22}) + (\bar{Y}_{31} + \bar{Y}_{32}) - (\bar{Y}_{41} + \bar{Y}_{42})] \\ \hat{\beta} &= \frac{1}{8} \sum_{i=1}^4 (\bar{Y}_{i_1} - \bar{Y}_{i_2}) \\ \hat{\gamma} &= \frac{1}{8} [(\bar{Y}_{11} + \bar{Y}_{12}) - (\bar{Y}_{21} + \bar{Y}_{22}) + (\bar{Y}_{31} - \bar{Y}_{32}) - (\bar{Y}_{41} - \bar{Y}_{42})] \\ \hat{\delta} &= \frac{1}{4} (\bar{Y}_{12} - \bar{Y}_{22} + \bar{Y}_{32} - \bar{Y}_{42}) \\ \hat{\varepsilon} &= \frac{1}{4} (-\bar{Y}_{12} - \bar{Y}_{22} + \bar{Y}_{32} + \bar{Y}_{42}) \end{aligned}$$

If we denote the variance of the sum and the difference of two paired observations by  $\sigma_+^2$  and  $\sigma_-^2$  respectively, we find

$$(22) \quad \begin{cases} \sigma_+^2 = \sigma_{11} + \sigma_{22} + 2 \sigma_{12} \\ \sigma_-^2 = \sigma_{11} + \sigma_{22} - 2 \sigma_{12} \end{cases}$$

With these notations we find for the variance of the parameter estimators:

$$(23) \quad \begin{aligned} \sigma^2(\hat{\mu}) &= \left( \sum_{i=1}^4 \frac{1}{n_i} \right) \frac{1}{64} \sigma_+^2 \\ \sigma^2(\hat{\alpha}) &= \left( \frac{1}{n_1} + \frac{1}{n_2} \right) \frac{1}{64} \sigma_-^2 + \left( \frac{1}{n_3} + \frac{1}{n_4} \right) \frac{1}{64} \sigma_+^2 \\ \sigma^2(\hat{\beta}) &= \left( \sum_{i=1}^4 \frac{1}{n_i} \right) \frac{1}{64} \sigma_-^2 \\ \sigma^2(\hat{\gamma}) &= \left( \frac{1}{n_1} + \frac{1}{n_2} \right) \frac{1}{64} \sigma_+^2 + \left( \frac{1}{n_3} + \frac{1}{n_4} \right) \frac{1}{64} \sigma_-^2 \\ \sigma^2(\hat{\delta}) &= \left( \sum_{i=1}^4 \frac{1}{n_i} \right) \frac{1}{16} \sigma_{22} \\ \sigma^2(\hat{\varepsilon}) &= \left( \sum_{i=1}^4 \frac{1}{n_i} \right) \frac{1}{16} \sigma_{22} \end{aligned}$$

To derive estimators of these variances we define for  $i = 1, \dots, 4$

$$(24) \quad \begin{cases} s_{i+}^2 = \frac{1}{n_i-1} \sum_{k=1}^{n_i} [(Y_{i1k} + Y_{i2k}) - (\bar{Y}_{i1} + \bar{Y}_{i2})]^2 \\ s_{i-}^2 = \frac{1}{n_i-1} \sum_{k=1}^{n_i} [(Y_{i1k} - Y_{i2k}) - (\bar{Y}_{i1} - \bar{Y}_{i2})]^2 \\ s_{i2}^2 = \frac{1}{n_i-1} \sum_{k=1}^{n_i} (Y_{i2k} - \bar{Y}_{i2})^2 \end{cases}$$

Then the following variance estimators are unbiased and independent of the corresponding parameter estimators:

$$s^2(\hat{\mu}) = \left( \sum_{i=1}^4 \frac{1}{n_i} \right) \frac{1}{64} \frac{1}{N-4} \sum_{i=1}^4 (n_i-1) s_{i+}^2$$

$$\begin{aligned} s^2(\hat{\alpha}) &= \left( \frac{1}{n_1} + \frac{1}{n_2} \right) \frac{1}{64} \frac{1}{n_1+n_2-2} \\ &\quad [(n_1-1) s_{1-}^2 + (n_2-1) s_{2-}^2] \\ &\quad + \left( \frac{1}{n_3} + \frac{1}{n_4} \right) \frac{1}{64} \frac{1}{n_3+n_4-2} \\ &\quad [(n_3-1) s_{3-}^2 + (n_4-1) s_{4+}^2] \\ (25) \quad s^2(\hat{\beta}) &= \left( \sum_{i=1}^4 \frac{1}{n_i} \right) \frac{1}{64} \frac{1}{N-4} \sum_{i=1}^4 (n_i-1) s_{i-}^2 \\ s^2(\hat{\gamma}) &= \left( \frac{1}{n_1} + \frac{1}{n_2} \right) \frac{1}{64} \frac{1}{n_1+n_2-2} \\ &\quad [(n_1-1) s_{1+}^2 + (n_2-1) s_{2+}^2] \\ &\quad + \left( \frac{1}{n_3} + \frac{1}{n_4} \right) \frac{1}{64} \frac{1}{n_3+n_4-2} \\ &\quad [(n_3-1) s_{3-}^2 + (n_4-1) s_{4-}^2] \\ s^2(\hat{\delta}) &= s^2(\hat{\varepsilon}) = \left( \sum_{i=1}^4 \frac{1}{n_i} \right) \frac{1}{16} \frac{1}{N-4} \sum_{i=1}^4 (n_i-1) s_{i2}^2 \end{aligned}$$

where  $N = \sum_{i=1}^4 n_i$  denotes the total sample size.

Since  $s^2(\hat{\mu})$ ,  $s^2(\hat{\beta})$ ,  $s^2(\hat{\delta})$  and  $s^2(\hat{\varepsilon})$  are  $\chi^2$ -distributed with  $N-4$  degrees of freedom and all parameter estimators are normally distributed, the following test statistics are t-distributed with  $N-4$  degrees of freedom if the corresponding parameters are zero, and may be used to test the hypotheses  $\mu = 0$ ,  $\beta = 0$ ,  $\delta = 0$  and  $\varepsilon = 0$ .

$$(26) \quad \begin{cases} t_\mu = \hat{\mu}/s(\hat{\mu}) \\ t_\beta = \hat{\beta}/s(\hat{\beta}) \\ t_\delta = \hat{\delta}/s(\hat{\delta}) \\ t_\varepsilon = \hat{\varepsilon}/s(\hat{\varepsilon}) \end{cases}$$

For the more interesting hypotheses  $\alpha = 0$  and  $\gamma = 0$  we define test statistics of the same type, but they are not t-distributed in general.

$$(27) \quad \begin{cases} t_\alpha = \hat{\alpha}/s(\hat{\alpha}) \\ t_\gamma = \hat{\gamma}/s(\hat{\gamma}) \end{cases}$$

Note, however, that the variances of  $\hat{\alpha}$  and  $\hat{\gamma}$  are both of the form  $\lambda_1 \sigma_1^2 + \lambda_2 \sigma_2^2$  with  $\lambda_1 = 1/n_1 + 1/n_2$  and  $\lambda_2 = 1/n_3 + 1/n_4$ , and the corresponding estimators may be written as  $\lambda_1 s_1^2 + \lambda_2 s_2^2$  with  $s_1^2$  and  $s_2^2$  independent,  $\chi^2$ -distributed with  $f_1 = n_1 + n_2 - 2$  and  $f_2 = n_3 + n_4 - 2$  degrees of freedom, and unbiased for  $\sigma_1^2$  and  $\sigma_2^2$  respectively.

Thus, the formulas of WELCH [8] or the tables of ASPIN [1] may be used to find the critical values for the distribution of  $t_\alpha$  and  $t_\gamma$ . For practical purposes it is sufficient to use the t-distribution with the approximate degrees of freedom  $f$ , where

$$(28) \quad f = \frac{1}{c^2/f_1 + (1-c^2)/f_2}$$

and

$$(29) \quad c = \lambda_1 s_1^2 / (\lambda_1 s_1^2 + \lambda_2 s_2^2)$$

(see the appendix by WELCH in ASPIN [1]).

In particular,  $t_\alpha$  and  $t_\gamma$  may approximately be regarded as t-distributed with  $f_\alpha$  and  $f_\gamma$  degrees of freedom, where

$$(30) \quad \begin{cases} 1/f_\alpha = c_\alpha^2 / (n_1+n_2-2) + (1-c_\alpha)^2 / (n_3+n_4-2) \\ 1/f_\gamma = c_\gamma^2 / (n_1+n_2-2) + (1-c_\gamma)^2 / (n_3+n_4-2) \end{cases}$$

where

$$(31) \quad \begin{cases} c_{\alpha} = \left( \frac{1}{n_1} + \frac{1}{n_2} \right) \frac{1}{64} \frac{1}{n_1+n_2-2} [(n_1-1) s_{1-}^2 \\ + (n_2-1) s_{2-}^2] / s^2 (\hat{\alpha}) \\ c_{\gamma} = \left( \frac{1}{n_1} + \frac{1}{n_2} \right) \frac{1}{64} \frac{1}{n_1+n_2-2} [(n_1-1) s_{1+}^2 \\ + (n_2-1) s_{2+}^2] / s^2 (\hat{\gamma}) \end{cases}$$

## 5. Conclusion

It could be shown that adding a second period to the first treatment period in order to compare two treatments with each other may result in four different types of new effects: the period effect, the carryover effect, the treatment by treatment interaction and the treatment by period interaction. Only the first of them (the period effect) can be identified and eliminated by the simple crossover design AB, BA. So this design can be used to analyze the direct treatment effect only, if the carryover effect and both interaction effects (treatment by treatment and treatment by period) are zero. On the other hand, in this case the simple crossover design is optimal (in the sense of minimal variance of the parameter estimator) as was shown by LASKA and MEISNER [6].

However, if there is no clear evidence in the experimental situation that carryover effects will not occur, one should use the extended design AB, BA, AA, BB in order to identify and eliminate these effects, for otherwise the second period cannot be used to analyze the direct effect. Again, according to the results of LASKA and MEISNER the stated design (AB, BA, AA, BB) is optimal in this case. Note, however, that both interaction effects (treatment by period and treatment by treatment) must be zero in order to use the observations of the second period for the analysis of the direct effect.

Finally, if one cannot be sure that both the treatment by treatment and the treatment by period interaction are zero, one should not apply the carryover design at all to analyze the direct treatment effect, because in this case the second period is of no use. Unfortunately the experimental conditions under which these effects could be eliminated, are often contrary to each other. For instance, if a possible treatment by treatment

interaction can be eliminated by a long wash-out period, the time delay of the second period may generate a treatment by period effect if the situation of the experimental unit or all other essential conditions are not stationary, and vice versa.

So it may be concluded that in a concrete situation a good knowledge about the existence or non-existence of all effects mentioned is crucial for the decision of applying or not applying a crossover design, and, if applicable for choosing the proper version of it.

## References

- [1] ASPIN, A. (1949): Tables for use in comparisons whose accuracy involves two variances, separately estimated. *Biometrika* **36**, 290-296.
- [2] BROWN, B.W. (1980): The crossover experiment for clinical trials. *Biometrics* **36**, 69-79.
- [3] GRIZZLE, J. E. (1965): The two-period change-over design and its use in clinical trials. *Biometrics* **21**, 467-480.
- [4] HILLS, M. and P. ARMITAGE (1979): The two-period cross-over clinical trial. *Br. J. clin. Pharmac.* **8**, 7-20.
- [5] KERSHNER, R. P. and W. T. FEDERER (1981): Two-treatment crossover design for estimating a variety of effects. *Journal of the American Statistical Association* **76**, 612-619.
- [6] LASKA, E. M. and M. MEISNER (1985): A variational approach to optimal two-treatment crossover designs: Application to carry-over-effect models. *Journal of the American Statistical Association* **80**, 704-710.
- [7] SCHNEIDER, B. (1983): Crossover designs and repeated measurements. *Neuropsychobiology* **10**, 49-55.
- [8] WELCH, B. L. (1947): The generalization of Student's problem when several different population variances are involved. *Biometrika* **34**, 28-35.
- [9] ZIMMERMANN, H. and W. RAHLFS (1980): Model Building and Testing for the Change-over Design. *Biometrical Journal* **22**, 197-210.

Eingegangen am 21. April 1986.

Anschrift des Verfassers: Dr. Hartmut Hecker, Medizinische Hochschule Hannover, Institut für Biometrie, Konstanty-Gutschow-Straße 8, 3000 Hannover 61

EDV in Medizin und Biologie 17 (3), 67–70, ISSN 0300-8282  
© Verlag Eugen Ulmer GmbH & Co., Stuttgart; Gustav Fischer Verlag KG, Stuttgart

## Explorative Variablenselektion und Anpassungstests bei Regressionsmodellen zur Analyse der stationären Aufenthaltsdauer nach Unfallverletzungen im Schulsport

P. Kemény, F. Rothmeier und A. Hamerle

### Zusammenfassung

*Untersucht wird die Dauer des stationären Krankenhausaufenthaltes von  $N = 554$  im Jahre 1981 bei Ballspielen im Schulsport verletzten Schülern in Abhängigkeit von 64 demographischen, die Unfallgenese beschreibenden und biometrischen Kovariablen. Mit Hilfe der Maximum-Partial-Likelihood-Quotientenmethode wurde eine Variablenselektion durchgeführt und der Kovariablenvektor auf 4 signifikante Komponenten reduziert. Die Parameterschätzungen des Cox-Modells wurden mit den Daten des Jahres 1981 gewonnen. Zur Überprüfung von Survival-Modellspezifikationen wurde ein verallgemeinerter  $\chi^2$ -Test konstruiert. Anhand eines neuen Datensatzes aus dem Jahre 1983 wurde dieser Test durchgeführt mit dem Ergebnis, daß das Cox-Modell eine gute Anpassung an die beobachteten Daten darstellt. Ein parametrischer Weibull-Regressionsansatz mit einer monotonen Grundhazardrate mußte zugunsten des Cox-Modells, das keine explizite Spezifizierung der Grundhazardrate voraussetzt, abgelehnt werden.*

### Summary

*For a data set of  $N = 554$  pupils and students injured in the year 1981 by ball-game accidents occurring in the course of sports lessons, the duration of hospital sojourn time is investigated depending on a vector of 64 covariates. A variable selection procedure using Maximum-Partial-Likelihood Method results in a reduction of the covariate vector to four significant components. The Cox-Model is fitted to the reduced 1981 data set. A generalized  $\chi^2$ -test developed to test survival model hypotheses is introduced. This test is applied to a 1983 data set. The parametric Weibull regression model assuming a monotonous baseline hazard rate is rejected, whereas it can be shown that the Cox-Model characterized by an unspecified baseline hazard rate provides an adequate fit to the observed data.*

### 1. Problemstellung

Untersucht wird die Dauer  $T$  des stationären Aufenthalts von  $N = 554$  im Jahre 1981 bei Ballspielen im Schulsport verletzten

Schülern in Abhängigkeit von demographischen (Alter, Geschlecht, Schulart), die Unfallgenese beschreibenden (verletzungsauslösende Spielphase) sowie biometrischen Kovariablen (Verletzungsdiagnosen).

Dieser Datensatz ist Teil einer Studie über Krankenhausverweildauern, die im Rahmen einer von der gesetzlichen Unfallversicherung in den Jahren 1982 und 1984 erstellten medizinischen Dokumentation zur Rehabilitation Unfallverletzter und arbeitsbedingt Erkrankter durchgeführt wird. Dabei besitzt aus epidemiologischer Sicht die Untersuchung der medizinischen Rehabilitation bei schweren Unfällen im Schulsport ein gesondertes Interesse. Im einzelnen wurden folgende Kovariablen erhoben:

- (1) Alter
- (2) Geschlecht
  - männlich
  - weiblich
- (3) Schulart
  - Gymnasium
  - Realschule
  - Hauptschule
- (4) Sportart
  - Handball
  - Basketball
  - Volleyball
  - Fußball
- (5) Verletzungsauslösende Spielphase
  - Ballannahme, Balltreffer
  - Umknicken, Sturz
  - Zusammenstoß mit Mitspieler, Körperberührung
  - sonstige verletzungsauslösende Spielphase
- (6) Verletzter Körperteil
  - Unterarm, Hand, Finger
  - Kniescheibe, Oberschenkel, Hüfte
  - Kniegelenk, Schienbein, Unterschenkel, Knöchelbereich
  - Sprunggelenk, Fuß
  - sonstiger verletzter Körperteil
- (7) Art der Verletzung
  - Contusio
  - Distorsion
  - Luxation
  - Zerreißung
  - Fraktur
  - sonstige Verletzungsart

**2. Regressionsansatz und Variablenselektion**

Ziel der Studie ist es, die Entlassungsrate  $\lambda(t|z)$  eines Individuums in Abhängigkeit von den in Abschnitt 1 aufgelisteten Kovariablen geeignet zu modellieren.

Da über die Zusammenhangsstrukturen sowie die Grundhazardfunktion  $\lambda_0(t)$  keine a-priori-Informationen vorliegen, wird als Startansatz für die Hazardfunktion eines verletzten Schülers mit Kovariablenvektor  $z$  das Cox-Modell (COX, 1972)

$$\lambda(t|z) = \lambda_0(t) \exp(\beta'z) \tag{1}$$

gewählt, wobei auch Zwei- und Drei-Faktor-Interaktionen in den Ansatz mitaufgenommen werden. Damit steht zunächst das Problem einer geeigneten explorativen Modellsuchstrategie bei hoher Kovariablenanzahl (61 Regressoren) im Vordergrund.

Dazu wurde der Ausgangsregressionsvektor  $z = (z_1, \dots, z_{61})'$  nach inhaltlichen Gesichtspunkten in drei Teilvektoren partitioniert:

$z = (y_1, y_2, y_3)'$ ,  $y_1 = (z_1, \dots, z_{19})$ ,  $y_2 = (z_{20}, \dots, z_{34})$  und  $y_3 = (z_{35}, \dots, z_{61})$ , wobei  $y_1$  die Haupteffekte (Basisvariablen),  $y_2$  und  $y_3$  entsprechende Interaktionen 1. und 2. Ordnung beinhalten. Die schrittweise Regression zur Variablenselektion erfolgte dann mit Hilfe der Prozedur BMDP2L (DIXON [ed.], 1981) in drei Schritten. Im ersten Schritt wurde lediglich der Teilvektor  $y_1$  in den Regressionsansatz (1) aufgenommen und mit Hilfe der Maximum-Partial-Likelihood-Quotienten (MPLQ)-Methode sowohl eine Vorwärtsselektion wie auch eine Rückwärtselimination ausgeführt. Man erhielt einen reduzierten Variablensatz  $\tilde{y}_1$  mit vier Variablen. Im zweiten Schritt wurde der Vektor  $(\tilde{y}_1, y_2)$  in das Modell (1) aufgenommen. Als Ergebnis der zwei schrittweisen Prozeduren erhielt man wiederum den Vektor  $\tilde{y}_1$ . Im dritten Schritt verfuhr man mit  $(\tilde{y}_1, y_3)$  analog und erhielt wieder den Variablensatz  $\tilde{y}_1$ .

Damit war keiner der untersuchten Interaktionsterme signifikant, übrig blieb ein Haupteffektmodell in (1) mit vier dichotomen Kovariablen. Den der Person  $i$  zugeordneten Vektor dieser Kovariablen bezeichnen wir mit  $x_i = (x_{i1}, x_{i2}, x_{i3}, x_{i4})'$ .

Die Komponenten des nach der MPL-Methode (COX, 1975; KALBFLEISCH/PRENTICE, 1980; MILLER, 1981; LAWLESS, 1982) geschätzten Parametervektors  $\hat{\beta} = (\hat{\beta}_1, \hat{\beta}_2, \hat{\beta}_3, \hat{\beta}_4)'$ , deren Standardabweichungen, die standardisierten Parameterschätzungen sowie die Überschreitungswahrscheinlichkeiten sind in Tabelle 1 angegeben.

Zur Variablenselektion bei Regressionsmodellen insbesondere auch im Hinblick auf numerisch effiziente Verfahren vgl. man FAHRMEIR/HAMERLE (1984).

**3. Anpassungstests**

*3.1 Graphische Methoden*

Graphische Möglichkeiten für Modell-Checks wurden von SCHOENFELD (1982), LAGAKOS (1980) und KAY (1977) angegeben.

Um die Güte der Anpassung des Cox-Modells an die beobachteten Daten beurteilen zu können, hat KAY (1977) vorgeschlagen, verallgemeinerte Residuen

$$\hat{\varepsilon}_i = \int_0^{t_i} \hat{\lambda}_0(u) \exp(\hat{\beta}'x_i) du \quad (i = 1, \dots, N) \tag{2}$$

(vgl. COX/SNELL, 1968) zu verwenden, und gegen  $-\ln \hat{S}(t_i|x_i)$

zu plotten. Liegt eine Stichprobe mit nichtzensierten Daten vor, so sind die kumulativen Hazardraten

$$\hat{\Lambda}(t_i|x_i) = -\ln \hat{S}(t_i|x_i)$$

unabhängige Realisierungen einer standard-exponential verteilten Zufallsvariablen. Wenn das spezifizierte Modell die Daten angemessen beschreibt, sollten sich die Residuen, die die Parameterschätzungen enthalten, approximativ ebenfalls wie exponentialverteilte Zufallsvariablen mit Parameter  $\lambda = 1$  verhalten, d. h., der Plot sollte eine Gerade mit Steigung 1 ergeben.

Generell ist dabei festzustellen, daß objektive Kriterien zur Beurteilung eines Plots nicht vorhanden sind, d. h., anhand welchen Maßes soll man feststellen, ob z. B. die durch eine graphische Darstellung angedeuteten Abweichungen vom Modell signifikant sind. Deshalb ist als Ergänzung ein formaler statistischer Test wünschenswert.

*3.2 Verallgemeinerter  $\chi^2$ -Test*

Da im vorliegenden Fall die Krankenhausverweildauern nicht zensiert sind, läßt sich zur numerischen Beurteilung der Modellanpassung ein verallgemeinerter  $\chi^2$ -Test konstruieren. Dazu wird die Zeitachse in  $r$  paarweise disjunkte Intervalle  $I_v = [a_{v-1}, a_v)$ ,  $v = 1, \dots, r$  mit  $a_0 = 0$  und  $a_r = \infty$  zerlegt.

Sei  $d_v$  die Anzahl der beobachteten Aufenthaltsdauern, welche in das Intervall  $I_v$  fallen, und  $p_v$  die Wahrscheinlichkeit, daß die Zufallsvariable  $T$  einen Wert aus  $I_v$  annimmt. Die Hypothese, daß das Cox-Modell (1) eine adäquate Anpassung an die beobachteten Daten darstellt, kann dahingehend spezifiziert werden, daß sich die Wahrscheinlichkeiten  $p_v$  als spezielle Funktionen  $\pi_v(\beta)$  des Parametervektors des Cox-Modells darstellen lassen:

$$H_0: \text{Es gibt ein } \beta \in \mathbb{R}^p, \text{ so daß} \tag{3}$$

$$p_v = \pi_v(\beta) \text{ für } v = 1, \dots, r$$

$$H_1: H_0 \text{ trifft nicht zu.}$$

Die Überprüfung der Hypothese kann mit der Teststatistik

$$Q = \sum_{v=1}^r \frac{[d_v - N\pi_v(\hat{\beta})]^2}{N\pi_v(\hat{\beta})} \tag{4}$$

ausgeführt werden. Der vorgeschlagene Test basiert also auf einem Vergleich zwischen beobachteten und entsprechend der Modellspezifikation erwarteten Verweildauerhäufigkeiten.

Zur Berechnung der Teststatistik  $Q$  in (4), insbesondere der  $\pi_v(\hat{\beta})$ , geht man folgendermaßen vor:

Es gibt maximal  $2^p$  verschiedene Kovariablenvektoren  $\tilde{x}_s$  mit  $p$  binären Komponenten.

Ist  $x_i$  der zur Person  $i$  gehörende Kovariatenvektor, dann sei

$$G_s = \{i | x_i = \tilde{x}_s\} \text{ für } s = 1, \dots, 2^p.$$

$$\text{Es ist } G_{s_1} \cap G_{s_k} = \emptyset \text{ für } s_1 \neq s_k \text{ und } \sum_{s=1}^{2^p} |G_s| = N.$$

Die Wahrscheinlichkeiten  $p_v$  berechnen sich wie folgt:

$$\begin{aligned} p_v &= P(T \in I_v) = P \left\{ \bigcup_{s=1}^{2^p} [(T \in I_v) \cap G_s] \right\} = \sum_{s=1}^{2^p} P[(T \in I_v) \cap G_s] \\ &= \sum_{s=1}^{2^p} P(G_s) \cdot P[(T \in I_v) | G_s] = \sum_{s=1}^{2^p} P(G_s) [S(a_{v-1} | \tilde{x}_s) - S(a_v | \tilde{x}_s)] \\ \hat{p}_v &= \sum_{s=1}^{2^p} \frac{|G_s|}{N} \{ [\hat{S}_0(a_{v-1})]^{\exp(\hat{\beta}'\tilde{x}_s)} - [\hat{S}_0(a_v)]^{\exp(\hat{\beta}'\tilde{x}_s)} \} =: \pi_v(\hat{\beta}). \end{aligned}$$



Mit den Daten des Jahres 1981 wurde die Selektion der signifikanten Variablen durchgeführt. Nun ist  $\hat{\beta}$  der aus den ursprünglichen (nicht gruppierten) Beobachtungswerten gewonnene MPL-Schätzer und nicht der nach Gruppierung der Daten erhaltene multinominale ML-Schätzer. Damit ist die asymptotische Verteilung von Q nicht mehr  $\chi^2(r-1-p)$ . Sie liegt zwischen  $\chi^2(r-1-p)$  und  $\chi^2(r-1)$ , so daß lediglich eine untere bzw. obere Signifikanzschranke angebar ist (LAWLESS, 1982; DE GROOT, 1975, CHERNOFF/LEHMANN, 1954). In diesem Sinn ist der von GREENBERG/BAYARD/BYAR (1974) ausgeführte Test nicht korrekt.

Anhand eines neuen Datensatzes aus dem Jahre 1983 wurde der eben beschriebene Test durchgeführt. Der Wert der Teststatistik  $Q = 11,05 < \chi^2_{0,95}(6) = 12,59 < \chi^2_{0,95}(10)$  führte zu einem nichtsignifikanten Ergebnis, so daß das Cox-Modell (1) eine gute Anpassung an die beobachteten Daten darstellt.

### 3.3 Anpassung eines parametrischen Regressionsmodells

Schließlich wurde untersucht, ob sich die Grundhazardfunktion  $\lambda_0(t)$  in parametrischer Form darstellen läßt. Ein Plot der geschätzten logarithmierten kumulativen Hazardfunktion  $\hat{\Lambda}_0(t)$  gegen  $\ln(t)$  legt eine Darstellung der Form

$$\ln \hat{\Lambda}_0(t) = \alpha \ln(t) + \gamma \text{ nahe.}$$

Dies würde eine Weibull-Spezifikation von  $\lambda_0(t)$  in (1) implizieren:  $\lambda_0(t) = \lambda \alpha t^{\alpha-1}$  mit  $\lambda = e^\gamma$ . Jedoch mußte der entsprechende Weibull-Regressionsansatz wegen  $Q = 48,37 > \chi^2_{0,95}(10) = 18,31$  verworfen werden.

Auffallend ist in diesem Zusammenhang die gute Übereinstimmung der Regressionskoeffizienten  $\hat{\beta}_i$  zwischen Cox- und Weibull-Modell (Tab. 2).

Daß der Weibull-Regressionsansatz die vorliegenden Daten dennoch nicht adäquat beschreibt, liegt im wesentlichen daran, daß das Weibull-Modell nur monotone Grundhazardfunktionen zuläßt. Hier wird die Flexibilität des Cox-Modells, das keine Spezifikation der Grundhazardfunktion voraussetzt, deutlich.

### 4. Abschließende Bemerkung

Die stationäre Aufenthaltsdauer nach Unfallverletzungen bei Ballspielen im Schulsport wird also durch ein Cox-Modell mit den in der Tabelle angegebenen Regressionsparametern adäquat beschrieben. Die inhaltliche Interpretation besagt, daß unabhängig von Alter, Geschlecht, Art des Ballspiels sowie verletzungsauslösender Spielphase Knie- und Knöchelverletzungen generell sowie Frakturen die stationäre Verweildauer im Krankenhaus signifikant verlängern. Der Hauptschulbereich weist im Gegensatz zu den anderen Schularten eine analoge Tendenz auf.

Tab. 1.

j	Kovariable $x_{ij}$	$\hat{\beta}_j$	$\hat{\sigma}_j$	$\frac{\hat{\beta}_j}{\hat{\sigma}_j}$	P-Wert
1	Kniescheibe, Oberschenkel, Hüfte	-0,8387	0,1811	-4,6319	0,000
2	Kniegelenk, Schienbein, Unterschenkel, Knöchelbereich	-0,4318	0,1020	-4,2340	0,000
3	Fraktur	-0,4626	0,1042	-4,4384	0,000
4	Hauptschule	-0,2019	0,0987	-2,0449	0,038

Tab. 2.

j	Kovariable $x_{ij}$	$\hat{\beta}_j$	
		COX	WEIBULL
1	Kniescheibe, Oberschenkel, Hüfte	-0,8387	-0,9507
2	Kniegelenk, Schienbein, Unterschenkel, Knöchelbereich	-0,4318	-0,4742
3	Fraktur	-0,4626	-0,5081
4	Hauptschule	-0,2019	-0,2562
$\hat{\alpha}$			0,8180
$\hat{\lambda}$			0,3518

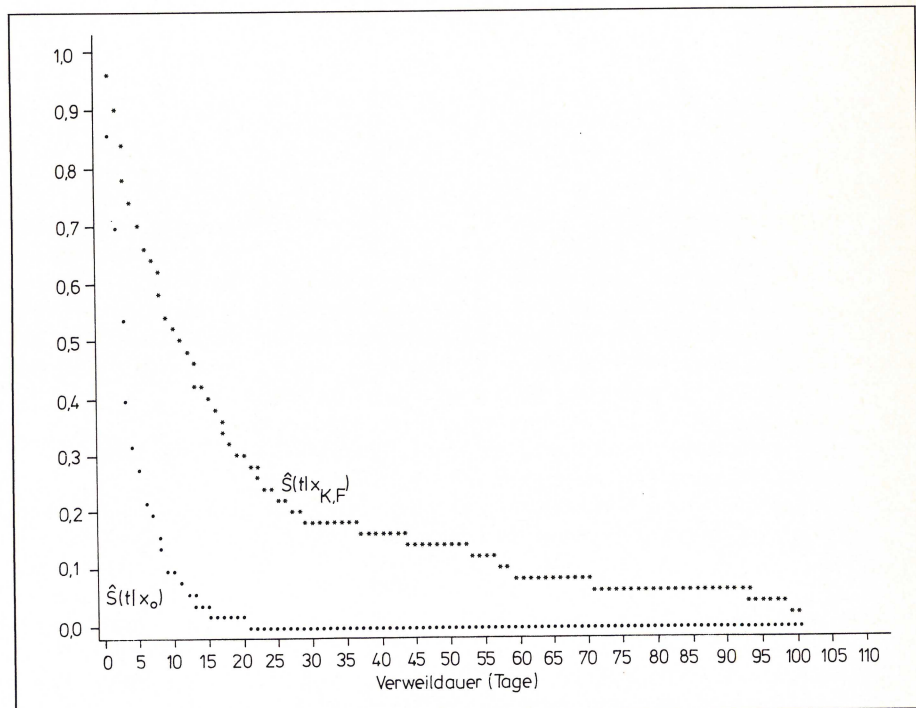


Abb. 1. Plot der geschätzten Verweildauerfunktionen für die Kovariablenvektoren  $x_0 = (0,0,0,0)'$  und  $x_{K,F} = (0,1,0,1)'$  (= Teilkollektiv mit Knie-scheibenverletzungen oder Frakturen).

Die numerische Berechnung der Parameterschätzungen und die Variablenselektion für das Cox-Modell wurden mit dem Programm BMDP2L durchgeführt. Die ML-Schätzung für das Weibull-Regressionsmodell erfolgte mit Hilfe des Programmsystems NAG. Für den Anpassungstest wurde eine eigene Prozedur erstellt, die auf Wunsch von den Autoren bezogen werden kann. Eine Überblicksdarstellung zum derzeit verfügbaren Software-Angebot zur regressionsanalytischen Behandlung von Verweildauern findet sich in KEMÉNY (1986).

## Literatur

- CHERNOFF, H., E. L. LEHMANN, 1954: The use of the maximum likelihood estimates in  $\chi^2$  test for goodness of fit. *Annals of Math. Statistics*, **25**, 579.
- COX, D. R., 1972: Regression models and life tables (with discussion). *J. R. Statist. Soc. B*, **34**, 187–220.
- COX, D. R., 1975: Partial likelihood. *Biometrika*, **62**, 269–276.
- COX, D. R., E. J. SNELL, 1968: A general definition of residuals. *J. R. Statist. Soc. B*, **30**, 248–275.
- DE GROOT, M. H., 1975: *Probability and Statistics*. Addison Wesley, Menlo Park.
- DIXON, W. J. (ed.), 1981: *BMDP Statistical Software*. University of California Press, Berkeley.
- FAHRMEIR, L., A. HAMERLE, 1984: *Multivariate statistische Verfahren*. Berlin.
- GREENBERG, R. A., S. BAYARD, D. BYAR, 1974: Selecting concomitant variables using a likelihood ratio step-down procedure and a method of testing goodness of fit in an exponential survival model. *Biometrics*, **30**, 601–608.
- KALBFLEISCH, J. D., R. L. PRENTICE, 1980: *The Statistical Analysis of Failure Time Data*. Wiley, New York.
- KAY, R., 1977: Proportional hazard regression models and the analysis of censored survival data. *Appl. Statist.*, **26**, 227–237.
- KEMÉNY, P., 1986: Regressionsmodelle zur Analyse von Verweildauern – ein Software-Vergleich. In: W. LEHMACHER, A. HÖRMANN (Hg.): *Statistik-Software. 3. Konferenz über wissenschaftliche Anwendung von Statistik-Software, 1985*. Stuttgart, New York 1986.
- LAGAKOS, S. W., 1980: The graphical evaluation of explanatory variables in proportional hazard regression models. *Biometrika*, **68**, 93–98.
- LAWLESS, J. F., 1982: *Statistical Models and Methods for Lifetime Data*. Wiley, New York.
- MILLER, R. G., 1981: *Survival Analysis*. Wiley, New York.
- SCHOENFELD, D., 1982: Partial residuals for the proportional hazards regression model. *Biometrika*, **69**, 239–241.

Eingegangen am 16. Mai 1986.

Anschrift der Verfasser:

Dr. P. Kemény, Dipl.-Math. F. Rothmeier, Bundesverband der Unfallversicherungsträger der öffentlichen Hand (BAGUV), Bereich Statistik und EDV, Marsstr. 46, 8000 München 2.  
Prof. Dr. A. Hamerle, Universität Konstanz, Abteilung Statistik, Postfach 55 60, 7750 Konstanz 1.

EDV in Medizin und Biologie **17** (3), 70–74, ISSN 0300-8282

© Verlag Eugen Ulmer GmbH & Co., Stuttgart; Gustav Fischer Verlag KG, Stuttgart

# Der Vorzeichentest bei wiederholter Beobachtung – die exakte Verteilung für kleine Stichproben

E. Brunner

## Zusammenfassung

Wenn Meßergebnisse mit wiederholten Messungen vorliegen, bei denen nur Größenbeziehungen ( $<$ ,  $>$ ,  $=$ ) verglichen werden können, kann dieser Versuchsplan mit einem von BRUNNER/NEUMANN (1984) vorgeschlagenen Testverfahren ausgewertet werden. Die Verteilung der dort angegebenen Statistik konnte jedoch nur für »große« Stichprobenumfänge hergeleitet werden. Im vorliegenden Artikel wird für kleine Stichprobenumfänge ein bedingter Test angegeben. Die Verteilung der Statistik kann mit einem Shift-Algorithmus rasch berechnet werden. Das Verfahren wird anhand eines Beispiels erläutert.

## Summary

If data with repeated measurements can only be compared as  $<$ ,  $>$ ,  $=$  then this design can be evaluated by means of a

procedure given by BRUNNER/NEUMANN (1984). The drawback of this test is that the distribution of the statistic is only known for large sample sizes. A conditional test for small sample sizes is derived in the present article. The distribution of this statistic can be calculated quickly by means of a shift algorithm. The procedure is illustrated by means of a practical example.

## 1. Einleitung

Es soll angenommen werden, daß eine Behandlung B gegen eine Kontrolle K an  $n$  unabhängigen Versuchseinheiten  $E_i$ ,  $i = 1, \dots, n$  untersucht wird. Das Ergebnis  $X_{iB}$  bzw.  $X_{iK}$  soll an derselben Versuchseinheit nur »ganz grob« als  $X_{iB} >$ ,  $<$  oder  $= X_{iK}$  beurteilt werden können. Die verbale Hypothese »kein Behandlungseffekt« kann dann beschrieben werden durch  $H_0^D$ :  $P(X_{iB} < X_{iK}) = P(X_{iK} < X_{iB})$ . Diese Hypothese kann getestet werden durch den bekannten Vorzeichen-Test (mit Bindun-

gen). Wird nun, um eine bessere Meßgenauigkeit zu erreichen, die Beurteilung  $X_{iB} <, >$  oder  $= X_{iK}$  mehrfach an derselben Versuchseinheit durchgeführt (mehrere Beurteiler, mehrfache Messungen), so liegen abhängige Beobachtungen  $X_{ijB}$  und  $X_{ij' B}$  bzw.  $X_{ijK}$  und  $X_{ij' K}$  für  $j, j' = 1, \dots, m$  vor, da die  $m$  Beurteiler zwar unabhängig voneinander aber jedesmal dasselbe Versuchsergebnis beurteilt haben. Es wird vorausgesetzt, daß die Anzahl der  $m$  Wiederholungen (Beurteiler, Meßgeräte) je Versuchseinheit konstant ist. Die Hypothese »kein Behandlungseffekt« kann dann so formuliert werden, daß für jeden Beurteiler  $j$  die Wahrscheinlichkeit, das Ergebnis der Behandlung »besser« einzustufen als das Ergebnis der Kontrolle gleich der Wahrscheinlichkeit ist, es »schlechter« einzustufen, d. h.:

$$H_0^D: \Delta_{ij}^* = P(X_{ijB} < X_{ijK}) - P(X_{ijB} > X_{ijK}) = 0 \quad \forall j = 1, \dots, m \text{ und } i = 1, \dots, n.$$

Sinnvollerweise nimmt man an, daß diese Wahrscheinlichkeiten nicht von der Versuchseinheit  $i$  abhängen, so daß  $\Delta_{ij}^* \equiv \Delta_j^*$  gilt.

Für die etwas allgemeinere Hypothese  $H_0^D$ :

$$\Delta^* = \frac{1}{m} \sum_{j=1}^m \Delta_j^* = 0$$

wurde von BRUNNER/NEUMANN (1984) im Modell mit additiven Effekten ein asymptotisch verteilungsfreier Test angegeben. Die Hypothese  $H_0^D$  ist etwas allgemeiner als  $H_0^D$ , ist jedoch für die Untersuchung der finiten Verteilung der hier vorgeschlagenen Teststatistik zu groß (d. h. liefert nicht genug Information), wenn man Bindungen zuläßt. Dieser Sachverhalt soll für unterschiedliche Modellannahmen im Anhang diskutiert werden. Im folgenden soll für  $H_0^D: \Delta_j^* = 0$  ein bedingter Test für kleine Stichprobenumfänge hergeleitet werden. Dabei sind »Bindungen« zugelassen.

## 2. Teststatistik und Verteilung

Es bezeichne

$$c^*(a, b) = \begin{cases} 0 & \text{falls } a < b \\ 1/2 & \text{falls } a = b \\ 1 & \text{falls } a > b \end{cases}$$

die Zählfunktion für Paarvergleiche mit Bindungen. Die Nullhypothese wird beschrieben durch  $H_0^D: \Delta_j^* = 0 \quad \forall j = 1, \dots, m$ . Aus  $H_0^D$  folgt  $H_0^D$ :

$$\Delta^* = \frac{1}{m} \sum_{j=1}^m \Delta_j^* = 0.$$

Man schätzt nun  $\Delta^*$  erwartungstreu und konsistent durch

$$\hat{\Delta}_n^* = \frac{1}{n} \sum_{i=1}^n \hat{\Delta}_{i,m}^*$$

mit

$$\hat{\Delta}_{i,m}^* = \frac{1}{m} \sum_{j=1}^m [c^*(X_{ijK} < X_{ijB}) - c^*(X_{ijB} < X_{ijK})].$$

Die beiden Statistiken  $C_{ij}^- = c^*(X_{ijK}, X_{ijB})$  und  $C_{ij}^+ = c^*(X_{ijB}, X_{ijK})$  sind unter  $H_0^D$  verteilungsgleich, da  $c^*(a, b) = 1 - c^*(b, a)$  ist. Weiter folgt für die Summen

$$\begin{aligned} \frac{1}{m} \sum_{j=1}^m C_{ij}^- - \frac{1}{2} &= \frac{1}{m} \sum_{j=1}^m [1 - C_{ij}^+] - \frac{1}{2} \\ &= \frac{1}{2} - \frac{1}{m} \sum_{j=1}^m C_{ij}^+. \end{aligned}$$

Somit sind die Summen

$$\frac{1}{m} \sum_{j=1}^m C_{ij}^- \text{ und } \frac{1}{m} \sum_{j=1}^m C_{ij}^+ \text{ unter } H_0^D \text{ symmetrisch zu } 1/2 \text{ verteilt}$$

(RANDLES/WOLFE, Satz 1. 3. 16). Daraus folgt weiter, daß

$$\hat{\Delta}_{i,m}^* = \frac{1}{m} \sum_{j=1}^m C_{ij}^- - \frac{1}{m} \sum_{j=1}^m C_{ij}^+$$

symmetrisch zu 0 verteilt ist.

Dies bedeutet, daß in der Summe

$$n \hat{\Delta}_n^* = \sum_{i=1}^n \hat{\Delta}_{i,m}^*$$

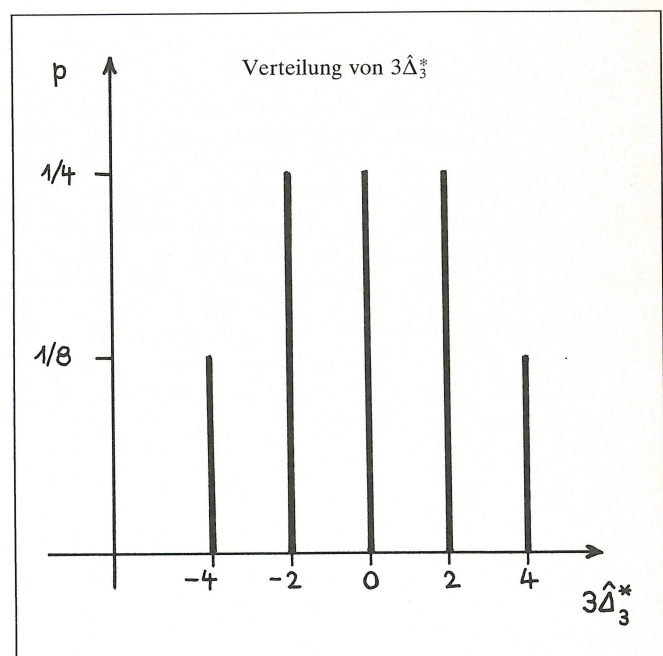
alle Vorzeichenpermutationen gleichwahrscheinlich sind, so daß die auf  $\hat{\Delta}_{1,m}^*, \dots, \hat{\Delta}_{n,m}^*$  bedingte Verteilung von  $n \hat{\Delta}_n^*$  unter  $H_0^D$  ausgezählt werden kann. Dabei ist es technisch einfacher, nicht mit den Mittelwerten  $\hat{\Delta}_{i,m}^*$ , sondern mit den Summen  $m \hat{\Delta}_{i,m}^*$  zu arbeiten. Zur Erläuterung soll das folgende einfache Zahlenbeispiel dienen:

$$n = 3, m = 2$$

i	j	$c^*(X_{ijK}, X_{ijB})$	$\hat{\Delta}_{i,2}^*$	$2\hat{\Delta}_{i,2}^*$	
1	1	1	1	2	$P(2) = P(-2)$
	2	1			
2	1	0	$-1/2$	$-1$	$P(-1) = P(1)$
	2	$1/2$			
3	1	1	$1/2$	1	$P(1) = P(-1)$
	2	$1/2$			

Die folgenden 8 Summen sind gleichwahrscheinlich:

i	Permutation							
	1	2	3	4	5	6	7	8
1	2	2	2	2	-2	-2	-2	-2
2	-1	-1	1	1	-1	-1	1	1
3	1	-1	1	-1	1	-1	1	-1
Summe $3\hat{\Delta}_3^*$	2	0	4	2	-2	-4	0	-2



Zur schnelleren Berechnung verwendet man den von ZIMMERMANN (1985) angegebenen Shift-Algorithmus: Die Summe  $3\hat{\Delta}_3^*$  kann maximal den Wert

$$S = \sum_{i=1}^3 |m\hat{\Delta}_{i,m}^*| = 4$$

und minimal den Wert  $-S = -4$  annehmen. Man beginnt z. B. mit dem maximalen Wert 4, notiert die Häufigkeit, »shiftet« um 2  $|m\hat{\Delta}_{i,m}^*|$ , addiert die Häufigkeiten, usw.:

Summe $3\hat{\Delta}_3^*$	Häufigkeiten			Ergebnis	kumul. Häufigk.
	Start	1. shift	2. shift		
4	1	1	1	1	1
3	0	0	0	0	1
2	0	0	1	1	2
1	0	0	0	0	3
0	0	1	1	0	5
-1	0	0	0	0	5
-2	0	0	1	1	7
-3	0	0	0	0	7
-4	0	0	0	0	8

Da die Verteilung zu 0 symmetrisch ist, braucht eigentlich nur der obere Teil der Tabelle berechnet zu werden.

### 3. Beispiel und Programm

Die Anwendung des Verfahrens soll anhand des von BRUNNER/NEUMANN (1984) angegebenen Beispiels gezeigt werden. Dazu wird ein SAS-Programm<sup>1)</sup> angegeben, das die als SAS-Datensatz vorhandenen Daten verarbeitet. Man erhält hier eine sehr gute Übereinstimmung mit der asymptotisch gefundenen Überschreitungswahrscheinlichkeit.

### Anhang

Die in der Einleitung erwähnte Beziehung zwischen  $H_0^D$  und  $\bar{H}_0^D$  bedarf einer kurzen Diskussion.

Zunächst soll das Modell *ohne Bindungen* betrachtet werden: Das Versuchsergebnis unter Behandlung bzw. Kontrolle werde beschrieben durch die Zufallsvariablen  $X_B$  bzw.  $X_K$ . Dieses Ergebnis wird bei der Versuchseinheit Nr.  $i$  vom Beurteiler Nr.  $j$  als Zufallsvariable  $X_{ijB}$  bzw.  $X_{ijK}$  beobachtet. Ein Vergleich zwischen  $X_{ijB}$  und  $X_{ijK}$  soll nur als  $X_{ijB} < X_{ijK}$  oder  $X_{ijB} > X_{ijK}$  angegeben werden können. Es bezeichne  $p_{ij} = P(X_{ijB} < X_{ijK})$  und  $q_{ij} = P(X_{ijB} > X_{ijK}) = 1 - p_{ij}$ , da Bindungen ausgeschlossen sind. Wenn die Beobachtung neuer Versuchseinheiten als Wiederholung angesehen wird, dann ist  $p_{ij} \equiv p_j$  und  $q_{ij} \equiv q_j \forall i = 1, \dots, n$ . Eine Gleichheit der  $p_j$  für alle Beurteiler ist hier nicht nötig.

Die Hypothese  $\bar{H}_0^D$ :

$$p = \frac{1}{m} \sum_{j=1}^m p_j = \frac{1}{2} \left( \Leftrightarrow \frac{1}{m} \sum_{j=1}^m (p_j - q_j) = 0 \right)$$

bedeutet, daß »im Mittel« die Beurteiler die Behandlung nicht besser und nicht schlechter einstufen als die Kontrolle.

Mit Hilfe der Zählfunktion

$$c(a, b) = \begin{cases} 0, & a < b \\ 1, & a > b \end{cases}$$

erhält man für  $p$  einen konsistenten und erwartungstreuen Schätzer

$$\hat{p}_n = \frac{1}{n} \sum_{i=1}^n \frac{1}{m} \sum_{j=1}^m c(X_{ijK}, X_{ijB}).$$

Schränkt man nun das hier angegebene allgemeine Modell auf ein lineares Modell ein, so erhält man:

$$X_{ijv} = \mu_v + \delta_j + Z_i + \varepsilon_{ijv} \\ i = 1, \dots, n \\ j = 1, \dots, m \\ v = B, K$$

```
DATA DATALIB.BEISPIEL;
  INPUT PATNR APLUS AMINUS ANULL;
  CARDS;
```

```
01 6 0 1
02 5 0 2
03 5 0 2
04 2 0 5
05 4 0 3
06 4 0 3
07 0 4 3
08 6 1 0
09 0 1 6
10 7 0 0
11 4 0 3
12 0 0 7
13 5 0 2
14 1 0 6
15 2 0 5
16 0 1 6
17 1 0 6
18 2 0 5
19 3 3 1
20 1 2 4
21 0 0 7
22 7 0 0
23 1 2 4
24 2 2 3
25 0 4 3
26 2 0 5
27 7 0 0
28 0 1 6
29 1 0 6
30 3 1 3
31 1 2 4
32 1 4 2
33 0 2 5
34 3 1 3
35 2 4 1
36 4 0 3
37 7 0 0
38 1 3 3
39 2 2 3
40 2 3 2
41 3 2 2
42 0 0 7
43 3 1 3
```

<sup>1)</sup> Herrn cand. math. D. Compagnone danke ich für die Erstellung des SAS-Programms.

$\mu_v$ : Behandlungseffekt  
 $\delta_j$ : Gutachtereffekt  
 $Z_i$ : Blockeffekt  
 $\varepsilon_{ijv}$ : Beobachtungsfehler, unabhängig, identisch verteilt nach  $F(x)$ , stetig.

BRUNNER/NEUMANN (1984) haben gezeigt, daß  $\mu_B = \mu_K$  genau dann gilt, wenn  $\hat{H}_0^D$  gilt. Damit folgt unter

$$\hat{H}_0^D : P(X_{ijB} < X_{ijK}) = P(\varepsilon_{ijB} < \varepsilon_{ijK}) = \frac{1}{2}$$

und daraus folgt, daß  $\sum_{j=1}^m c(X_{ijK}, X_{ijB})$  symmetrisch verteilt ist zu  $m/2$ .

Im allgemeinen Modell folgt dieser Zusammenhang nur für  $m = 2$ ; für  $m > 2$  kann man leicht Gegenbeispiele angeben.

Läßt man nun *Bindungen* zu, so benutzt man auch für  $m = 2$  die im Abschnitt 2 definierte Zählfunktion  $c^*(a, b)$ . Der Parameter

$$\Delta^* = \frac{1}{m} \sum_{j=1}^m \Delta_j^*$$

beschreibt einen Beurteilungsunterschied. Nun liefert die Hypothese  $\hat{H}_0^D : \Delta^* = 0$  nicht genügend Information, um unter Gültigkeit von  $\hat{H}_0^D$  die Verteilungsgleichheit von

$$\sum_{j=1}^m C_{ij}^- \quad \text{und} \quad \sum_{j=1}^m C_{ij}^+$$

folgern zu können. Es lassen sich sofort im allgemeinen Modell für  $m = 2$  Gegenbeispiele angeben, die belegen, daß zwar  $\hat{H}_0^D$  gilt, jedoch die Summen

```

PROC MATRIX ;

  FETCH APLUS DATA=DATALIB.BEISPIEL(KEEP=APLUS) ;
  FETCH AMINUS DATA=DATALIB.BEISPIEL(KEEP=AMINUS) ;

  ADIFF = APLUS - AMINUS ;
  ADIFF = ADIFF(LOC(ADIFF),) ; * DIE NULLEN WERDEN BESEITIGT ;
  S = SUM(ADIFF) ;
  T = SUM(ADIFF)#/SQRT( SUM(ADIFF#ADIFF) ) ;
  PT = 2#{1-PROBNORM(ABS(T))} ;
* S IST EINE STATISTIK FUER DEN PERMUTATIONS-TEST ;
* T IST EINE ASYMTOTISCH N(0,1)-VERTEILTE STATISTIK ;
* PT IST DIE ZWEISEITIGE UEBERSCHREITUNGS-WAHRSCHEINLICHKEIT FUER T ;

  AABS = ABS(ADIFF) ;
  SMAX = SUM(AABS) ; N = 2#SMAX+1 ;
  P = J(N,1,0) ; P(1,) = 1 ;
* DIE FOLGENDEN BEIDEN DO-SCHLEIFEN BERECHNEN DURCH EINEN SHIFT- ;
* ALGORITHMUS DIE VERTEILUNG VON S BEDINGT AUF AABS ;
  DO I=1 TO NROW(AABS) ;
    DO K=N TO 2#AABS(I,)+1 BY -2 ;
      P(K,) = P(K,) + P(K-2#AABS(I,)) ;
    END ;
    P = P#/2 ;
  END ;
* P(K,) = P( S=SMAX+1-K / AABS ) UNTER HO ;
  PS = SUM( P(1:SMAX+1-ABS(S),) )#2 ;
* PS IST DIE ZWEISEITIGE UEBERSCHREITUNGS-WAHRSCHEINLICHKEIT FUER S ;

  ERGEBNIS = T!!PT!!S!!PS ; ERGEBNIS=ERGEBNIS//ERGEBNIS ;
  ERGEBNIS(1,2 4) = ERGEBNIS(1,2 4)#/2 ;
  C = 'T' 'P>=!T!' 'S' 'P>=!S!' ;
  R='1-SEITIG' '2-SEITIG' ;
  PRINT ERGEBNIS COLNAME=C ROWNAME=R;

RUN;

ERGEBNIS  T    P>=!T!          S    P>=!S!

1-SEITIG  2.89714 .001883    64    .001604
2-SEITIG  2.89714 .003766    64    .003208

```

$$\sum_{j=1}^m C_{ij}^- \text{ und } \sum_{j=1}^m C_{ij}^+$$

nicht verteilungsgleich sind.

Für die enger gefaßte Hypothese  $H_0^D: \Delta_j^* = 0 \forall j = 1, \dots, m$  wurde in Abschnitt 2 die Verteilungsgleichheit der beiden Summen gezeigt. Trivialerweise folgt aus  $H_0^D \Rightarrow \tilde{H}_0^D$ , aber nicht umgekehrt.

Die Hypothese  $H_0^D$  bringt keine große Einschränkung gegenüber  $\tilde{H}_0^D$ ; sie kann gut anschaulich interpretiert werden und bedeutet, daß kein Gutachter eine Präferenz für B oder K hat, falls kein Behandlungseffekt vorliegt.

Im linearen Modell folgt aus der Unabhängigkeit und identischen Verteilung der  $\varepsilon_{ijv}$  unter  $\tilde{H}_0^D$ :

$$\sum_{j=1}^m [P(\varepsilon_{ijK} < \varepsilon_{ijB}) - P(\varepsilon_{ijB} < \varepsilon_{ijK})] = 0$$

$$\Rightarrow m [P(\varepsilon_{11K} < \varepsilon_{11B}) - P(\varepsilon_{11B} < \varepsilon_{11K})] = 0$$

$$\Rightarrow P(\varepsilon_{11K} < \varepsilon_{11B}) = P(\varepsilon_{11B} < \varepsilon_{11K}).$$

$$\Rightarrow \Delta^* \equiv \Delta_j^* = 0 \forall j = 1, \dots, m, \text{ d. h. } H_0^D.$$

Das bedeutet, daß im linearen Modell auch bei Bindungen  $\tilde{H}_0^D$  mit dem angegebenen Permutationstest getestet werden kann.

Unberührt von diesen Überlegungen, die nur für die finite Verteilung der Teststatistik angestellt wurden, bleiben die asymptotischen Resultate. Hier genügt es, daß  $\hat{\Delta}_n^*$  ein konsistenter, erwartungstreuer Schätzer für  $\Delta^*$  ist, um die asymptotische Normalität folgern zu können.

#### Literatur

- BRUNNER, E., und N. NEUMANN (1984): Vorzeichen-Tests in mehrfaktoriellen Versuchsanlagen, EDV in Med. Biol. **15**, 44–49.  
 RANLDES, R. H. and D. A. WOLFE (1979): Introduction to the Theory of Nonparametric Statistics. Wiley, New York.  
 ZIMMERMANN, H. (1985): Exact Calculation of Permutational Distributions for Two Dependent Samples I. Biom. J. **24**, 349–352.

Eingegangen am 28. Juli 1986.

Anschrift des Verfassers: Prof. Dr. E. Brunner, Abt. Medizinische Statistik, Fachbereich Medizin der Universität Göttingen, Windaus Weg 2, D-3400 Göttingen.

## NACHRICHTEN UND BERICHTE

### Johann-Peter-Süßmilch-Medaille der GMDS

Die Deutsche Gesellschaft für Medizinische Dokumentation, Informatik und Statistik e.V. schreibt hiermit die *Johann-Peter-Süßmilch-Medaille 1987* für hervorragende Arbeiten im Bereich der medizinischen Dokumentation, Informatik und Statistik aus.

Die zur Bewerbung einzureichenden Arbeiten in deutscher oder englischer Sprache müssen nach dem 1. 4. 1985 publiziert oder zur Publikation angenommen worden sein. Jeder Bewerber kann nur eine Arbeit einreichen. Wird eine Arbeit mehrerer Autoren vorgelegt, so muß der Beitrag des einreichenden Bewerbers deutlich erkennbar sein.

Die Medaille wird in der Regel jährlich an Wissenschaftler/-innen unter 40 Jahren verliehen. Über die Vergabe entscheidet ein Gutachter-Kollegium.

Einreichungsschluß ist der **31. März 1987**.

Bewerbungen sind zu richten an die Geschäftsstelle der GMDS, Herbert-Lewin-Str. 1, 5000 Köln 41.

### 32. Jahrestagung der GMDS

5.–7. Oktober 1987, Tübingen

Thema: Medizinische Informationsverarbeitung und Epidemiologie im Dienste der Gesundheit.  
Anmeldung von Vorträgen oder Poster bis zum **7. März 1987** an: Tagungssekretariat, Abteilung für Med. Dokumentation und Datenverarbeitung, Westbahnhofstraße 55, 7400 Tübingen, Telefon (0 70 71) 29-52 18

## BUCHBESPRECHUNGEN

PAASS, G. und WAUSCHKUHN, K.

### Datenzugang, Datenschutz und Anonymisierung

Analysepotential und Identifizierbarkeit von anonymisierten Industriedaten

GMD Bericht Nr. 148

1985, 401 S., DM 70,-

R. Oldenbourg Verlag, München – Wien

Im Zusammenhang mit der Datenschutzdiskussion ergeben sich zunehmend Schwierigkeiten bei der Weitergabe von Daten aus der amtlichen Statistik und dem Verwaltungsvollzug für Forschungsvorhaben. Das Projekt AIMIPH hat sich die Aufgabe gestellt, Kriterien für die Anonymität von Einzeldaten zu erarbeiten sowie geeignete Anonymisierungstechniken zu entwickeln und zu erproben.

Dazu wurde auch ein Verfahren zur Re-Identifikation von Einzelangaben aus Stichproben entwickelt, das insbesondere die bei allen Erhebungen nicht zu vermeidenden Datenfehler berücksichtigt.

Die umfangreiche Darstellung dieses Forschungsvorhabens ist ein wertvolles Hilfsmittel für weitere Arbeiten und sollte auch zu einer Versachlichung der Diskussion beitragen. Ge.

LEHMACHER, W. und HÖRMANN, A. (Hrsg.)

### Statistik-Software

3. Konferenz über die wissenschaftliche Anwendung von Statistiksoftware, 1985

1986, 399 S., DM 58,-

G. Fischer Verlag, Stuttgart – New York

In diesem Band ist ein Großteil der Konferenz-Beiträge zusammengestellt. Ziel der Tagung war es, Themen über die Anwendung von Statistik-Software in den verschiedenen wissenschaftlichen Bereichen vorzustellen. Im Vordergrund standen dabei Programmvergleiche und -bewertungen, Übersichtsvorträge zu Software für spezielle statistische Fragestellungen sowie Vorträge über Neuerungen bei den eingeführten Softwaresystemen.

Damit liegt insgesamt eine instruktive Hilfe beim Einsatz von Statistik-Software vor.

SCHAFFLAND, H.-J. und WILTFANG, N.

### Bundesdatenschutzgesetz (BDSG)

Ergänzbarer Kommentar nebst einschlägigen Rechtsvorschriften

Ergänzbare Ausgabe, einschl. 19. Lieferung,

1250 Seiten, DIN A5, DM 86,-, zuzügl. Ordner DM 11,80

Ergänzungen folgen von Fall zu Fall

Erich Schmidt Verlag, Berlin – Bielefeld – München

Die Bedeutung des Bundesdatenschutzgesetzes (BDSG) gewinnt in der Praxis und in der Rechtsprechung immer mehr an Bedeutung.

Die 19. Lieferung enthält den gegenwärtig im Gesetzgebungsverfahren befindlichen Koalitionsentwurf zum BDSG nebst Begründung sowie einer ersten Kommentierung. Der Praktiker hat damit die Möglichkeit, sich auf die demnächst geltenden Neuregelungen rechtzeitig einzustellen.

BURHENNE, W. E. und PERBAND, K. (Hrsg.)

### EDV-Recht

Systematische Sammlung der Rechtsvorschriften, organisatorischen Grundlagen und Entscheidungen zur elektronischen Datenverarbeitung

Herausgegeben im Auftrage der Interparlamentarischen Arbeitsgemeinschaft

Ergänzbare Ausgabe, einschl. 46. Lieferung,

4214 Seiten und 3 Ausschlagtafeln, DIN A5, DM 148,-

zuzügl. 3 Spezialordner je DM 11,80

Ergänzungen folgen von Fall zu Fall

Erich Schmidt Verlag, Berlin – Bielefeld – München

Der Textteil dieses stets aktuellen und umfassenden Grundlagenwerkes enthält die einschlägigen Rechtsvorschriften, Dokumente, insbesondere die Ausführungsbestimmungen, und Gerichtsentscheidungen sowie die Materialien zur Organisation und Zuständigkeit für die EDV in der Verwaltung jeweils für den Bund und die einzelnen Länder.

Der Erläuterungsteil enthält Kommentierungen zum Bundesdatenschutzgesetz und zu den Vertragsbedingungen für EDV-Anlagen und -Geräte. Durch diese Erläuterungen wird die komplizierte Materie des EDV-Rechts auch Nichtjuristen verständlich gemacht.

In der 46. Lieferung erscheint u. a. das Volkszählungsgesetz 1987 sowie zahlreiche einschlägige Antworten auf parlamentarische Anfragen.

NIEMEYER, G.

### Einführung in das Programmieren in ASSEMBLER

5. bearb. und erw. Aufl.

1984, 332 S., DM 44,-

W. de Gruyter, Berlin – New York

Die Frage nach der Notwendigkeit einer Programmierung in ASSEMBLER kann man sehr unterschiedlich beantworten. Unbestritten ist aber wohl, daß es für Studenten keine bessere Einführung in die Programmierung von Computern gibt als der Weg über die ASSEMBLER-Programmierung.

Das vorliegende Buch zeichnet sich vor allem durch eine geschickte didaktische Darstellung aus, die es wirklich zu einem hervorragenden Lehrbuch macht. Ge.

SCHULZ, A.

**Einführung in das Programmieren in PL/1**

2. bearb. u. erw. Aufl.

1984, 332 S., DM 49,50

W. de Gruyter, Berlin - New York

In den letzten Jahren hat sich die strukturierte Programmierung in der Praxis voll durchgesetzt. Obwohl neben PL/1 jetzt eine ganze Reihe neuer, das strukturierte Programmieren unterstützende Programmiersprachen zur Verfügung stehen, hat PL/1 wegen seiner Universalität, z. B. in der Datenverarbeitung, nicht an Bedeutung verloren. Allerdings muß diese Eigenschaft systematisch angewandt werden.

Die Neuauflage des Buches trägt dieser Entwicklung Rechnung, indem nicht nur die Syntax von PL/1 anhand von Beispielen dargestellt wird, sondern auch Regeln vermittelt werden, wie diese Syntax in einem gut strukturierten Programm anzuwenden ist. Darüber hinaus werden neue Sprachelemente zur strukturierten Programmierung eingeführt.

GITTINS, R.

**Canonical Analysis**

A Review with Applications in Ecology

Biomathematics, Vol. 12

1985, 320 S., DM 128,-

Springer-Verlag, Berlin - Heidelberg - New York - Tokyo

Das Interesse an Beziehungen zwischen Sätzen von Variablen besteht in den unterschiedlichsten Bereichen. In der vorliegenden Monographie über die Canonische Analyse wird im ersten Abschnitt die Theorie dargestellt. Im zweiten Abschnitt werden eine Anzahl Anwendungen aus der Ökologie ausführlich behandelt und interpretiert. In einem dritten Teil werden Hinweise für weitere Arbeiten zusammengestellt. - Das klar gegliederte Buch ist als Einführung und als Lehrbuch ausgezeichnet geeignet. Ge.

HANUS, M.

**Probleme lösen mit PROLOG**

1986, 208 S., DM 23,80

B. G. Teubner, Stuttgart

Die Entwicklung immer neuer Computeranwendungen bedingt auch eine Weiterentwicklung der Programmiersprachen und der Programmier-techniken. Hier spielt PROLOG zur Zeit eine wesentliche Rolle. Das vorliegende Buch stellt eine brauchbare Einführung in die Sprache PROLOG und in die Anwendung dieser Sprache für die Lösung von Problemen dar. Ge.

MÖBUS, C. und SCHNEIDER, W. (Hrsg.)

**Strukturmodelle für Längsschnittdaten und Zeitreihen**

LISREL, Pfad- und Varianzanalyse.

Methoden der Psychologie, Bd. 5

1986, 276 S., DM 44,-

Verlag Hans Huber, Bern - Stuttgart - Toronto

Obwohl die Analyse von Strukturgleichungsmodellen schon seit geraumer Zeit angewandt wird, gibt es nach wie vor wenig einführende Literatur. Die für diese Problematik entwickelten Computerprogramme, z. B. LISREL, haben dazu geführt, daß in immer mehr Arbeiten diese Methoden benutzt werden. Die vorliegende Zusammenstellung will den Zugang zu diesem Themenbereich erleichtern, wozu neben einführenden Aufsätzen insbesondere auch die dargestellten Anwendungen beitragen werden. Ge.

HEITFELD, U.

**Einführung in das relationale Datenbanksystem SIR/DBMS**

1986, 279 S., DM 38,-

Gustav Fischer Verlag, Stuttgart - New York

Mit diesem Handbuch liegt die erste umfassende deutschsprachige Einführung in das Datenbanksystem SIR vor. Sie soll für die gegenwärtigen Anwender eine Hilfe sein und neuen Anwendern den Einstieg erleichtern. Beschrieben wurden die Systemmodule: DBMS, SQL+, FORMS und HELP, die in dieser Form auch schon für PC's

zur Verfügung stehen. Die Darstellung orientiert sich stark an Beispielen, wobei die verwendete Datenbank mit der Software ausgeliefert wird, d. h. für jeden Anwender verfügbar sein sollte. Jedes Kapitel wurde mit einem Referenzteil abgeschlossen, der die meisten Kommandos mit ihren Optionen in Kurzform dokumentiert. - Insgesamt eine nützliche Darstellung, die zu einer intensiveren Nutzung von SIR beitragen wird. Ge.

CÔTÉ, R. A., PROTTI, D. J., and SCHERRER, J. R.

**Role of Informatics in Health Data Coding and Classification Systems**

1985, 394 S., \$ 59,25

North-Holland, Amsterdam - New York - Oxford

Diese Konferenz war durch die WG 6 der IMIA organisiert. Die vorliegenden Proceedings enthalten die Referate sowie Resolutionen, die z. T. recht progressiv klingen. Aus der Zusammensetzung der Teilnehmer wird deutlich, daß die behandelten Themen ein weltweites Interesse besitzen. Ge.

EISENFELD, J. and DELISI, C. (Ed.)

**Mathematics and Computers in Biomedical Applications**

1985, 390 S., \$ 40,-

North-Holland, Amsterdam - New York - Oxford

Der vorliegende Band enthält die Proceedings einer interdisziplinären Konferenz. Die zusammengestellten Referate umspannen ein breites Spektrum von Themen, wie z. B. Sequentialanalyse, Proteinstrukturen, Neurobiologie, Immunologie, Pharmakokinetik, Planung von Versuchen, Dynamische Systeme, Bildverarbeitung und Signalverarbeitung. Ge.

RENNER, E.

**Statistikprogramme in Basic**

Pareys Studentexte, Nr. 56

1986, 185 S., DM 29,-

mit Diskette DM 49,-

Verlag Paul Parey, Berlin - Hamburg

Das vom gleichen Autor im gleichen Verlag in zweiter Auflage erschienene Buch (Studentexte Nr. 31) „Mathematisch-statistische Methoden in der praktischen Anwendung“ wird nun durch eine Sammlung von BASIC-Programmen ergänzt.

Als Computer wurde der Commodore C 64 verwendet, für den die Programme auch auf Diskette erhältlich sind. - Man mag über den heute bei Programmen zu erwartenden Komfort geteilter Meinung sein, dennoch wäre es wünschenswert, wenn durchgängig bei der statistischen Bearbeitung die Dateneingabe, die Datenmanipulation (z. B. Transformationen) und die möglichen Dateneingaben von den eigentlichen statistischen Berechnungen getrennt würden, um so auch ohne große Mühe mit den Ergebnissen (z. B. Residuen) weiterarbeiten zu können. Die Druckqualität bei den Programm Listings ist insbesondere auch durch die kaum entzifferbaren Sonderzeichen nicht befriedigend. Eine Eingabe aufgrund dieser Listings dürfte daher viel Mühe bereiten. Für die Anpassung an andere, gleichfalls verbreitete Rechner und Betriebssysteme fehlen leider konkrete Hinweise. Ge.

ALLETSEE, R., SCHMIDT, K.-D. und ZELLER, M.

**PASCAL-Praktikum**

Teil 1: Lernprogramm

Teil 2: Katalog

1984, DM 162,-

Siemens AG, Berlin und München

Das vorliegende PASCAL-Praktikum ist als programmierter Selbstunterricht mit Terminalübungen (PU) konzipiert. Grundkenntnisse in der Datenverarbeitung werden vorausgesetzt.

Die Darstellung ist aber so klar und übersichtlich, daß nach einer gründlichen Durcharbeitung Programme mittleren Schwierigkeitsgrades erstellt werden können. - Im Katalog sind die PASCAL-Anwendungen alphabetisch aufgeführt. Dieser Teil des Praktikums ist auch als Nachschlagewerk geeignet. Ge.