

## Automatic area calculation by microcomputer-controlled video analysis

D.-P. Häder

### Zusammenfassung

*Es wird ein kostengünstiges, mikrocomputer-gesteuertes Bildverarbeitungssystem beschrieben, das automatisch die Fläche von makroskopischen oder mikroskopischen Objekten bestimmt. Das Bild des Objektes wird von einer Videocamera erfaßt, sequentiell digitalisiert und im Computermemory gespeichert. Die Software ist in Assemblersprache geschrieben und erlaubt neben der Flächenberechnung auch die Bestimmung des Durchmessers an einer beliebigen Stelle des Objektes und die Berechnung des Flächenschwerpunktes. Biologische Applikationen sind u. a. Blattflächenbestimmung, Berechnung von Zellgrößen sowie Querschnittanalysen.*

### Summary

*A low cost, microcomputer-controlled image analysis system is described which automatically detects the area of macroscopic or microscopic objects. The image of the object is recorded by a video camera, digitized sequentially and stored in the computer memory. The software is written in assembler language and allows to calculate the area of the object, diameters at any chosen position and the center of gravity. Biological applications are leaf area detection, cell size determination or cross section analysis.*

### 1. Introduction

Recent developments in both video and microcomputer technology have opened new possibilities for solving old problems (SHIPTON, 1979; SERRA, 1980; KEMNITZ and HOUGARDY, 1980; SKARNULIS, 1982). Combining both techniques leads to automatic analysis in many fields such as image analysis, cell counting, shape detection and tracking of organisms (DESAI and REIMER, 1985; JAFFE et al., 1985; BRYAN et al., 1985). Other examples include quantification of particle or organism numbers (ERHARDT et al., 1980), cell size (PRESTON and DEKKER, 1980), image processing of high resolution electron microscopy (BOYES et al., 1982); GRAVEKAMP et al., 1982) or calculation of stomatal apertures (OMASA and ONOE, 1984).

By following temporal changes in the processed images it is possible to track the movement of particles or organisms and detect e.g. the orientation with respect to a stimulus. Depending on the speed requirements and complexity of the task, image analysis in the time domain can be performed by

recording and subsequent analysis (LEWANDOWSKA et al., 1979; DUSENBERY, 1985; GUALTIERI et al., 1985). Real time analysis of movement patterns requires high speed digitization and parallel data transfer as well as fast algorithms written in assembly language since high level languages need too much overhead which drastically reduces execution speed (HÄDER and LEBERT, 1985).

Area calculation is a tool for many technological and scientific fields. Leaf area detection is a classical example and is used by e.g. physiologists, morphologists as well as by ecologists (KOTHARI and SHAH, 1975; KINCAID and SCHNEIDER 1983; TERAMURA, 1980; TERAMURA et al., 1984). In addition to manual determination, area detection both on the microscopic and the macroscopic scale can be performed by several techniques including photometric procedures (CHAUBAL et al., 1981), planimetric (NULTSCH and HÄDER, 1974) or gravimetric determination (HÄDER and NULTSCH, 1973), Fast Fourier Analysis (KINCAID and SCHNEIDER, 1983), or statistical techniques such as the Monte Carlo technique (FINNEY, 1982).

Most automatic area determinations are based on a video recording of the object(s). The video image is digitized by an electronic analog/digital (A/D) converter and stored in a dedicated video read and write memory (RAM). Alternatively, part of the computer memory can be used to temporarily store the image. The hardware costs depend on the resolution of the image (number of pixels in horizontal times vertical directions), the number of gray levels and the time requirements. Real time digitization performed by flash A/D conversion is more expensive than scanning the image in slower motion using successive approximation registers which may require several seconds to minutes per video frame.

Depending on the available hardware several techniques are used to access the stored information including bus-oriented transfer, serial or parallel input/output (I/O) or memory-mapped access. Whatever the specific procedure is, the image is processed sequentially: Individual pixels (picture elements) are examined in a horizontal or vertical sequence and the key features of the image are extracted depending on the individual requirements.

The aim of this paper is to describe a low cost system based on an inexpensive b/w video camera and a non real time video A/D converter connected to a microcomputer. The software developed for this system enables area calculation on a microscopic or macroscopic scale, diameter measurements at any point of the sample as well as determination of the center of gravity. Possible applications include leaf area calculation, cross section measurement and cell size determination.

## 2. Material and methods

### 2.1 Hardware

Figure 1 shows the essential components of the analysis system. The video camera used for this purpose is a general purpose inexpensive b/w camera (Sony AVC 3250E). Since some of the applications require infrared measurements it is equipped with a newvicon tube which has a sensitivity maximum at wavelengths  $> 700$  nm. An additional feature is the built-in capability of an electronic inversion which is very useful since e.g. the optical contrast between microorganisms and their background can be drastically enhanced by dark field irradiation of the microscope image. The electronic image inversion restores a normal picture with dark objects on a bright background. The video information can also be stored on a  $\frac{1}{2}$  inch open reel time lapse video recorder (National NV 8030E) and analyzed later.

The signal is analyzed by a Digisector DS-80 digitizer (Micro Works, Del Mar, California) which plugs into any S-100 bus microcomputer. The resolution of this device is  $256 \times 256$  pixels and allows for 64 gray levels. While it digitizes 1 pixel in about  $4 \mu\text{s}$ , the actual time required to digitize a full video frame exceeded several seconds and could be minutes depending on the strategy of scanning. The digitizer board accepts either interlaced NTSC or standard USA industrial norm (525 lines) or the European 625 line norm (CCIR).

The system is built around an S-100 bus microcomputer based on a Z80 CPU (S.D. Systems, Dallas, Texas) running under CP/M (Digital Research, Pacific Grove, California). The microcomputer has two 8" disk drives, 64 kByte RAM and a video controller for  $80 \times 24$  characters on a monochrome display. The digitizer uses one slot in the bus and can be

accessed by sending two 8 bit addresses (X and Y coordinates) via two adjacent 8 bit parallel output ports. The gray level found at the corresponding pixel is returned at an 8 bit parallel input port. The bytes representing the gray level of the tested pixels are stored in the main microcomputer memory in an order appropriate for subsequent data analysis. The currently tested pixel appears white during digitization which facilitates debugging.

Macroscopic objects to be studied are placed on a translucent screen illuminated from below. Thus the camera records a dark image on a bright background. The scale can be calibrated by placing a transparent graded ruler on the screen and adjusting the camera distance so that the size of the image corresponds with the desired scale on the monitor screen. For area and distance calculations it has to be taken into account that the pixels are not square so that scales are different in x and y directions.

Microscopic objects are studied using a non-inverted light microscope (Zeiss Standard, Oberkochen, FRG). Since most cells have too low an absorbance, cutoff or interference band-pass filters (Schott & Gen., Mainz, FRG) can be placed in the light beam which select wavelength bands at which the organisms absorb maximally. Alternatively dark field irradiation can be used and the image electronically inverted. In each case uniform background irradiation is an important prerequisite since the algorithms to detect absolute values over a uniform background are easier than those which calculate differences between an object and a variable background. As for macroscopic objects the magnification is calibrated by inserting a graded scale.

### 2.2 Software

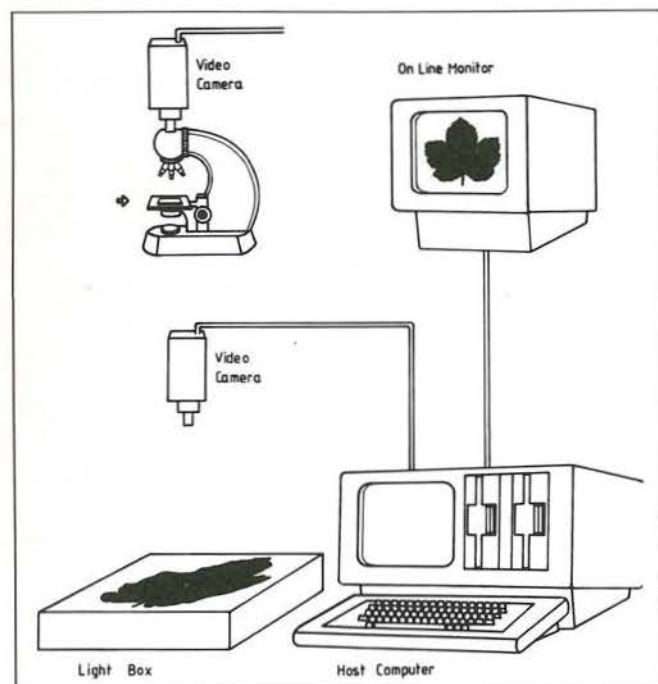
The basic software is written in Z80 or 8080 assembler language. Further programs for mathematical and statistical treatment of the images are developed in CBASIC (Compiler Systems Inc.) or PASCAL (Turbo Pascal, Borland Int., Scotts Valley, California).

The system requires that the image is digitized pixel by pixel and the data stored in the microcomputer memory. The scanning strategy has a large impact on the digitization time. Since each pixel can be digitized only when it is accessed by the video camera, under worst case conditions the delay time is 20 ms (US standard 17 ms) until the pixel is scanned by the camera. At a resolution of  $256 \times 256$  pixels the total scan time amounts to 22 min. Since digitization and data transfer require more time than the tv scanning speed allows between horizontal adjacent pixels digitization along horizontal lines is prohibitive; therefore pixels are digitized along vertical columns. Due to these timing considerations every second row is accessed in sequence and a second pass is made in order to cover the interlacing pixels.

As a complete frame would fill the total amount of memory (64 kByte) and leave no room for the operating systems and the analyzing program the resolution could be reduced to half by scanning only every second horizontal line. However, in order to obtain the maximal resolution, the area is slightly reduced by skipping the top and bottom rows as well as the lateral columns. For control purposes the memory image can be plotted as hard copy on a dot matrix printer (Epson, Nagano, Japan).

Area detection of either macroscopic or microscopic objects can easily be performed by integrating over the number of pixels differing from the background by a preselected threshold. Since the magnification can be controlled easily the

Fig. 1. Schematic representation of the system for area calculation. The video camera is either mounted over a light box or on top of a light microscope. The video image is digitized and stored in the microcomputer main memory. The video image currently digitized is displayed on a separate monitor



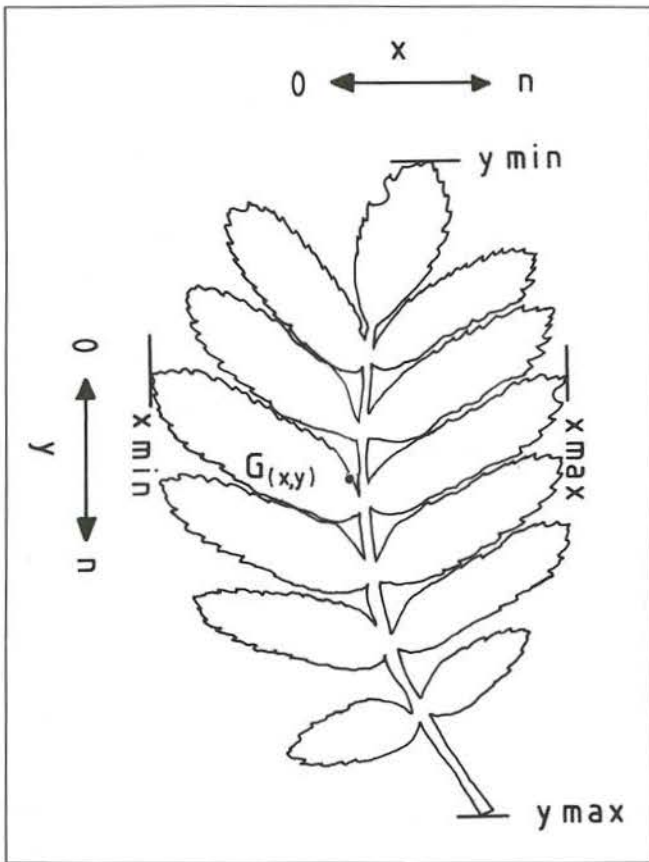


Fig. 2. Calculation of the geometrical center  $G_{(x,y)}$  in the image of a *Sorbus aucuparia* leaf from the minimal and maximal x and y coordinates  $x_{min}$ ,  $x_{max}$ ,  $y_{min}$  and  $y_{max}$

absolute area can be calculated from the number of pixels found in the object image. After placing the image in memory several parameters can be determined. From the minimal and maximal x and y coordinates ( $x_{max}$ ,  $x_{min}$ ,  $y_{max}$ ,  $y_{min}$ ) the geometrical center  $G_{(x,y)}$  is calculated (Fig. 2).

$$G_{(x,y)} = G_{(((x_{max}-x_{min})/2)+x_{min}, ((y_{max}-y_{min})/2)+y_{min})} \quad (1)$$

The algorithm to determine the center of gravity  $C_{(x,y)}$  is more complicated: Starting from the geometric center  $G_{(x,y)}$  the number of object pixels  $P_x$  and  $P_y$  on either side in x and y directions are determined and weighted according to their distance from the center (Fig. 3). When the weighted sum of the pixels in one half is out of balance with that in the other half (in both x and y directions) the center is moved accordingly to a new  $G_{(x,y)}$ .

$$\text{If } \sum_{x=1}^{G_x-1} (G_x-x) P_{(G_x-x)} + \sum_{x=G_x+1}^n (G_x-x) P_{(G_x-x)} > 0 \quad (2)$$

then  $C_{(x,y)} = C_{(x-1,y)}$  else  $C_{(x,y)} = C_{(x+1,y)}$  and

$$\text{if } \sum_{y=1}^{G_y-1} (G_y-y) P_{(G_y-y)} + \sum_{y=G_y+1}^n (G_y-y) P_{(G_y-y)} > 0 \quad (3)$$

then  $C_{(x,y)} = C_{(x,y-1)}$  else  $C_{(x,y)} = C_{(x,y+1)}$ .

By iterative recalculation the center of gravity  $C_{(x,y)}$  is approached. Iteration is stopped when the determined new

center oscillates around a fixed pixel in subsequent calculations.

The diameter of the object  $D_y$  at any x coordinate can be determined from the number of pixels by placing the object in such a way that a predefined scan line cuts the image at the desired position. The program sums the number of object pixels along the scan line and determines the real length according to the scaling factor.

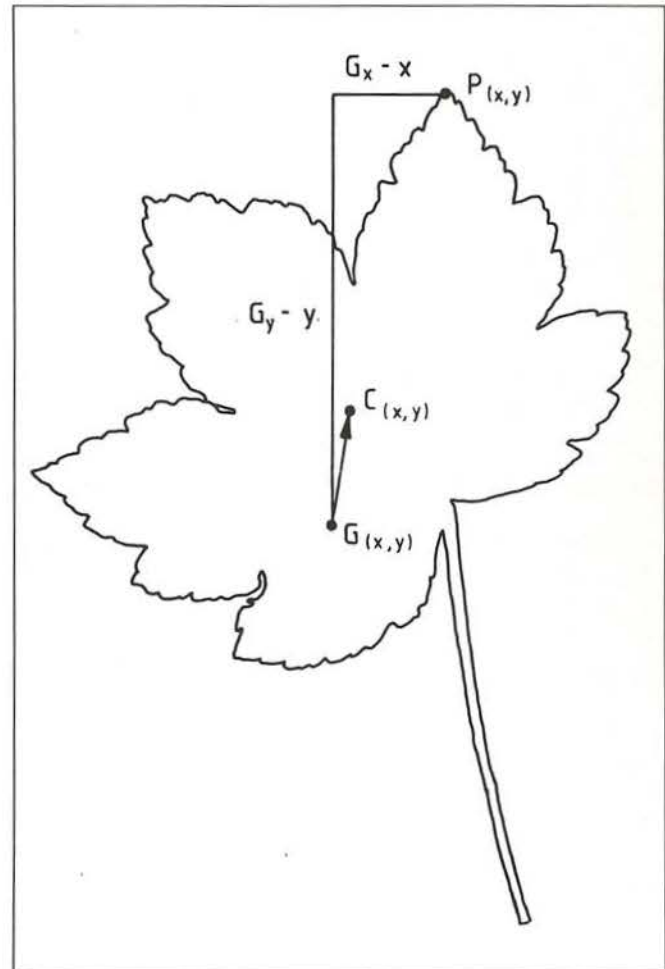
$$D_y = \sum_{x=1}^n P_{(x,i)} \text{ for all } i = \text{object pixels} \quad (4)$$

Some graphic terminals have a crosshair option by which the position of the scan line can be defined. Other methods to define the position of the scan line in addition to numerical input from the keyboard utilize an optical mouse or a joystick.

### 3. Results and discussion

The hardware and software system described in this article allows automatic area detection and calculation of important parameters. The system does not only calculate single areas

Fig. 3. Calculation of the center of gravity  $D_{(x,y)}$  in the image of an *Acer pseudo-platanus* leaf using an iterative process integrating over the weighted sums of the pixels either side of the currently calculated center in x and y direction



but also integrates over several subcomponents such as in compound leaves. While it is not a real time system it executes within a few seconds. Due to the interlaced scan mode the total area is covered within about 4 seconds. The area determination is almost instantaneous as well as the calculation of the geometrical center and the diameter at any cross section. The calculation of the center of gravity takes several seconds up to a couple of minutes depending on the size of the object due to the iterative process of calculation.

Calibration of both area and distance parameters is easy by proper selection of the magnification. The accuracy of this low cost system depends on the relative size of the object in the field of observation and thus on the pixel size. Best results are obtained by adjusting the object size so that it fills most of the screen. The second parameter which determines the accuracy is the linearity of the video camera. Tests with an object of only 10% the size of the screen have shown an accuracy of within  $\pm 1.2\%$  SEM when placing the object at different positions on screen.

The techniques described in this article can easily be implemented in other computer systems. For e.g. the IBM PC/XT/AT and compatibles several video digitizer boards are available. The programs to perform the described tasks can be converted into 80286/80386 assembly language. Recently real time frame grabbers have been introduced which speed up the process of digitization drastically since they digitize a video frame within 20 ms (European standard, 17 ms US standard) using a flash A/D converter.

All programs described in this article are available on 8" SD/SS or 5 1/2" IBM compatible disks from the author upon request at a nominal fee.

#### 4. Conclusions

The fully automatic system described in this article determines the area of macroscopic or microscopic samples of any shape. The hardware is suited to even a low budget and consists of a video camera which records the image of an object on a light box or placed on the stage of a microscope and a non real time analog/digital video digitizer. The image is stored in the main memory of a microcomputer and is analyzed by several programs written in assembly language which calculate the area, diameter at a chosen position, the geometrical center or the center of gravity.

#### Acknowledgements

This work was supported by the Deutsche Forschungsgemeinschaft (Ha 984/5-5). The author gratefully acknowledges the skillful technical assistance of U. Neis.

#### Literature

- BOYES, E. D., B. J. MUGGRIDGE, M. J. GORINGE (1982): On-line image processing in high resolution electron microscopy. *J. Microsc.* **127**, 321.  
 BRYAN, S. R., W. S. WOODWARD, D. P. GRIFFIS, R. W. LINTON

- (1985): A microcomputer based digital imaging system for ion microanalysis. *J. Microscopy* **138**, 15.  
 CHAUBAL, K. A., C. S. GODBOLE, V. S. AROSKAR (1981): A simple method to measure cell area photometrically. *Indian J. Exp. Biol.* **20**, 220.  
 DESAI, V., L. REIMER (1985): Digital image recording and processing using an Apple II microcomputer. *Scanning* **7**, 185.  
 DUSENBERY, D. B. (1985): Using a microcomputer and videocamera to simultaneously track 25 animals. *Comput. Biol. Med.* **15**, 169.  
 ERHARDT, R., E. R. REINHARDT, W. SCHLIPF, W. H. BLOSS (1980): FAZYTAN A system for fast automated cell segmentation, cell image analysis and feature extraction based on TV-image pickup and parallel processing. *Anal. Quant. Cytol. J.* **2**, 25.  
 FINNEY, J. L. (1982): Monte Carlo techniques. *Biochem. Soc. Trans.* **10**, 305.  
 GRAVEKAMP, C., H. K. KOERTEN, N. P. VERWOERD, W. C. DE BRUIJN, W. T. DEAMS (1982): Automated image analysis applied to electron micrographs. *Cell Biol. Int. Rep.* **6**, 656.  
 GUALTIERI, P., G. COLOMBETTI, F. LENCI (1985): Automatic analysis of the motion of microorganisms. *J. Microscopy* **139**, 57.  
 HÄDER, D.-P. (1974): Participation of two photosystems in the photophototaxis of *Phormidium uncinatum*. *Arch. Microbiol.* **96**, 255.  
 HÄDER, D.-P., M. LEBERT (1985): Real time computer-controlled tracking of motile microorganisms. *Photochem. Photobiol.* **42**, 509.  
 HÄDER, D.-P., W. NULTSCH (1973): Negative photo-phobotactic reactions in *Phormidium uncinatum*. *Photochem. Photobiol.* **18**, 311.  
 JAFFE, M. J., A. H. WAKEFIELD, F. TELEWSKI, E. GULLEY, R. BIRO (1985) Computer-assisted image analysis of plant growth, thigmomorphogenesis, and gravitropism. *Plant Physiol.* **77**, 722.  
 KEMNITZ, H.-D., H. P. HOUGARDY (1980): Design of an interface for digital image storage. *Mikroskopie* **7**, 415.  
 KINCAID, D. T., R. B. SCHNEIDER (1983): Quantification of leaf shape with a microcomputer and Fourier transform. *Can. J. Bot.* **61**, 2333.  
 KOTHARI, M. J., G. L. SHAH (1975): Epidermal structures and ontogeny of stomata in the Papilionaceae (Tribe heydysareae). *Bot. Gaz.* **136**, 372.  
 LEWANDOWSKA, K., J. DOROSZEWSKI, G. HAEMMERLI, P. STRÄULI (1979): An attempt to analyze locomotion of leukemia cells by computer image processing. *Comput. Biol. Med.* **9**, 331.  
 OMASA, K., M. ONOE (1984): Measurement of stomatal aperture by digital image processing. *Plant & Cell Physiol.* **25**, 1379.  
 PRESTON, K. jr., A. DEKKER (1980): Differentiation of cells in abnormal human liver tissue by computer image processing. *Anal. Quant. Cytol. J.* **2**, 14.  
 SERRA, J. (1980): Digitalization. *Mikroskopie* **37** Suppl., 109.  
 SHIPTON, H. W. (1979): The microprocessor, a new tool for the biosciences. *Ann. Rev. Biophys. Bioeng.* **8**, 269.  
 SKARNULIS, A. J. (1982): A computer system for on-line image capture and analysis. *J. Microsc.* **127**, 39.  
 TERAMURA, A. H. (1980): Effects of ultraviolet-B irradiances on soybean. I. Importance of photosynthetically active radiation in evaluating ultraviolet-B irradiance effects on soybean and wheat growth. *Physiol. Plant.* **48**, 333.  
 TERAMURA, A. H., M. C. PERRY, J. LYDON, M. S. MCINTOSH, E. G. SUMMERS (1984): Effects of ultraviolet-B radiation on plants during mild water stress. III. Effects on photosynthetic recovery and growth in soybean. *Physiol. Plant.* **60**, 484.

Date of receipt: April 23rd, 1987.

The authors' address: Dr. Donat-P. Häder, Fachbereich Biologie - Botanik, Philipps-Universität Marburg, Lahnberge, D-3550 Marburg, Federal Republic of Germany.

## Versatile digital image analysis by microcomputer to count microorganisms

D.-P. Häder and K. Griebenow

### Zusammenfassung

*Ein Bildanalyse-System aus modernen Komponenten wird beschrieben, das aus einem üblichen Microcomputer und einem in Echtzeit arbeitenden, hochauflösenden Einplatinen-Digitalisierer besteht. Bilder werden mit einer CCD Kamera im Makro- oder Mikromaßstab aufgenommen und in einem Speicher abgelegt, auf den der Computer Zugriff hat. Die Analysesoftware ist in 8088/8086 Assemblersprache geschrieben und erlaubt die Quantifizierung von Organismen mit jeder mikroskopischen Technik oder auf makroskopischer Ebene. Die meisten nutzbaren mathematischen Verfahren, wie Kontraststeigerung und Glättung, Laplace Filterung, Kanten-erkennung, Trennung von sich berührenden Objekten, sind implementiert. Das nachfolgende Statistikpaket ist in Turbo Pascal geschrieben und ermöglicht eine statistische Analyse sowie grafische Darstellung der Daten.*

### Summary

*A state-of-the-art image analysis system is described which is based on a common microcomputer in combination with a high resolution, real time single-board digitizer. Images are recorded by a CCD camera on a macro or micro scale and stored in a memory accessible by the computer. The analysis software is written in 8088/8086 assembly language and allows identification and quantification of organisms from any type of microscopy or macro image. Most useful mathematical treatments, such as contrast enhancement and smoothing, Laplace filtering, edge detection, separation of objects in contact, are implemented. A subsequent statistical package written in Turbo Pascal allows statistical analysis and graphical presentation of the data.*

### 1. Introduction

Modern microcomputer technology facilitates high performance image analysis to tackle a number of problems and allow hardware and software solutions of biologically relevant problems (SHIPTON, 1979; SERRA, 1980; KEMNITZ and HOUGARDY, 1980; SKARNULIS, 1982). The basic idea is to record a video image from any source, such as a video recorder, CCD or video camera, digitize the image and analyze it on a pixel by pixel basis with a microcomputer. The amount of calculations necessary depends on the task to be performed: Detecting objects of a given gray level on a uniform

background of a different gray level is rather easy compared to a detailed shape analysis (DESAI and REIMER, 1985; JAFFE et al., 1985; BRYAN et al., 1985).

Image analysis techniques can be employed to quantify particle or cell numbers or to process images such as electron microscopy photographs (ERHARDT et al., 1980; PRESTON and DEKKER, 1980; BOYES et al., 1982; GRAVEKAMP et al., 1982) or calculate sizes of areas such as leaves or stomatal apertures (OMASA and ONOE, 1984; HÄDER, 1987).

Analysis of motile microorganisms is more complicated since it requires to follow sequences of images in the time domain (LEWANDOWSKA et al., 1979; DUSENBERRY, 1985). Due to the enormous calculation requirements often a video sequence is recorded and analyzed afterwards frame by frame which may take far longer than the real time recording (GUALTIERI et al., 1985). Real time analysis of the movement of motile microorganisms requires first of all real time digitization, i.e., one video frame every 40 or 33 ms (depending on the video standard). The subsequent processing needs to be performed in the small time frame allowed by the stream of incoming video images. Thus, only assembly language can be used for this task since all higher languages are far too slow to be useful (HÄDER and LEBERT, 1985).

Cell counting is routinely performed in many laboratories mostly using manual techniques. This applies to blood cell counting as well as to the quantification of microorganisms such as individual bacteria or colonies. Most growth curves published in the literature are based on manual counting using microscopical techniques and a calibrated chamber such as a Thoma chamber. These tasks can be facilitated enormously by an automatic counting system. Some devices have been introduced such as the Coulter counter or high cost counting systems which incorporate a dedicated computer and a mechanical positioning gear. In contrast, the system described here uses a general purpose microcomputer and fast and versatile algorithms to solve a wide variety of problems. These techniques can as well be utilized to count and quantify hardware objects such as industrial products or grains of sand or rock. In this article we demonstrate the capabilities of the system by quantification of growth curves of the green flagellate, *Euglena gracilis*.

The system described here is based on a standard IBM PC/XT/AT (or compatible) microcomputer in combination with a plug-in real-time digitizer card. The basic software, written in Assembly language, is capable of dealing with all kinds of image situations since it includes a number of recently developed software techniques to smooth and enhance the images (ASMUS, 1987; DAWSON, 1987).

In addition, the software also computes the areas of individual objects, which is relevant for the analysis of, e.g., leaves, which in the past has been done manually, photometrically, planimetrically or gravimetrically (KOTHARI and SHAH, 1975; TERAMURA 1980; TERAMURA et al., 1984; CHAUBAL et al., 1981; NULTSCH and HÄDER, 1973, 1974). Fast Fourier Analysis (KINCAID and SCHNEIDER, 1983) or statistical methods such as the Monte Carlo technique have also been used for area calculation (FINNEY, 1982).

## 2. Materials and Methods

### 2.1 Organisms and Growth

Cell counting in dependence of culture age has been carried out in the green flagellate, *Euglena gracilis* (strain Z) (HÄDER, 1985, 1986). The cultures were grown in 100 ml Erlenmeyer flasks with 40 ml of an organic growth medium (STARR, 1964). Fresh flasks were inoculated with about 4 ml of a culture in logarithmic growth (MIKOLAJCZYK et al. 1986) and grown in static culture at 23°C continuously irradiated with constant white light at 600 lx from fluorescent lamps (warm tone and cool white mixed) to prevent synchronisation and establishment of a circadian rhythm (BRINKMANN, 1976).

For comparison with the automatically obtained cell densities, the number of flagellates per volume was determined using a light microscope (Zeiss Standard, Oberkochen, FRG) in combination with a conventional Thoma chamber. In addition, the cell densities were quantified by measuring the absorbance of aliquots taken from the culture at fixed time intervals using a Shimadzu dual beam spectrophotometer (MPS 5000).

### 2.2 Hardware

The image of the objects was recorded using a CCD camera (Philips No. LDH 0600/00) in combination with either an adjustable macro objective (Tarcus CC TV 12.5–75 mm,  $F=1.6$ ) or a conventional microscope (Olympus, BH2). Depending on the optical properties of the objects such as absorbance and refractive index, dark field or light microscopical techniques were employed (Fig. 1).

The recorded image was displayed on an on-line b/w monitor and simultaneously digitized in a Matrox single board digitizer (PIP 512) which plugs into the bus of a standard IBM PC/XT/AT or compatible. The microcomputer was a Picotron with 640 KB RAM (read and write memory) equipped with two 5 $\frac{1}{4}$ " floppy disk drives, two 5 $\frac{1}{4}$ " hard disk drives (10 and 20 MB, respectively) and a color monitor (CGA).

The digitizer is very versatile since it allows to switch between various video standards and to choose from a wide variety of built-in hardware features (Fig. 2). The incoming video signal can be controlled by a software gain and offset according to the requirements before it is digitized in a flash A/D converter with 256 gray levels. The spatial resolution is 512  $\times$  512 pixels. The gray level of the individual pixels can be mapped using one of eight programmable input look-up tables, e.g., all values below a certain predefined threshold can be mapped to represent black and all others white, or a specific range can be spread out etc. After this manipulation the data are stored in a dedicated on-board video memory (RAM).

The host computer has access to the video memory via input and output (I/O) ports and can both read and write individual pixels on a random basis and thus analyze or manipulate the image. A mask can be defined for overlaying the video image

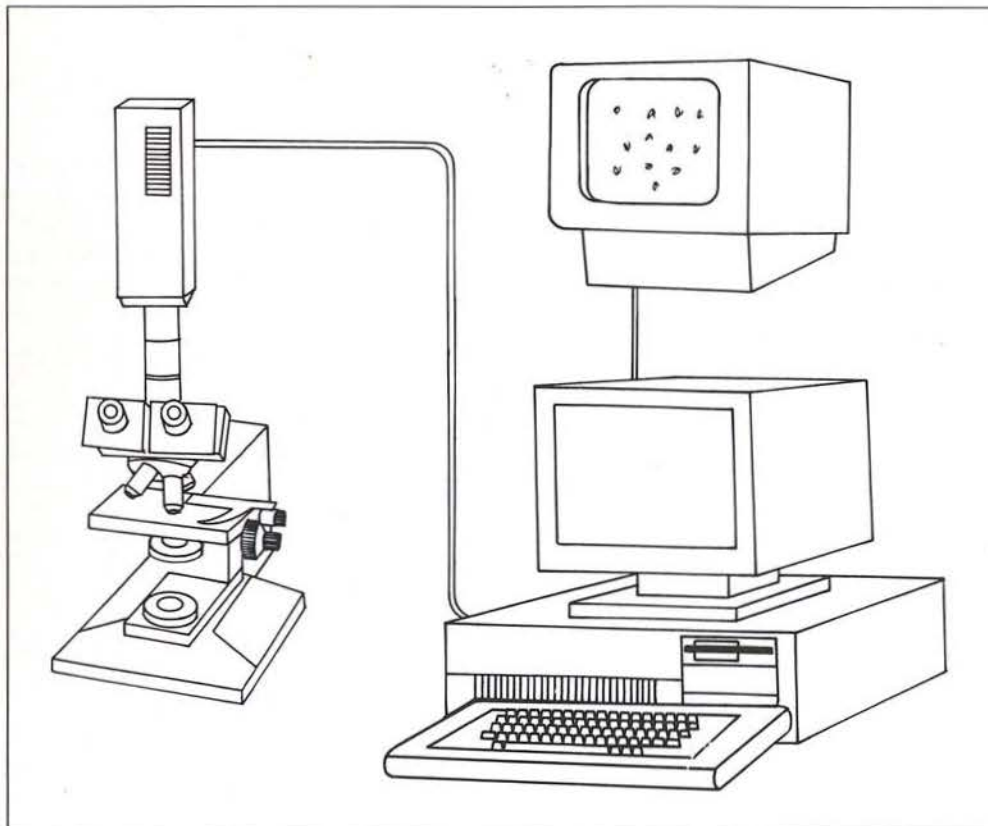


Fig. 1. Hardware system to analyze the number and areas of objects recorded by a CCD camera. The image is digitized in real time and stored in a video RAM. Both on-line and digitized images can be seen on separate video monitors

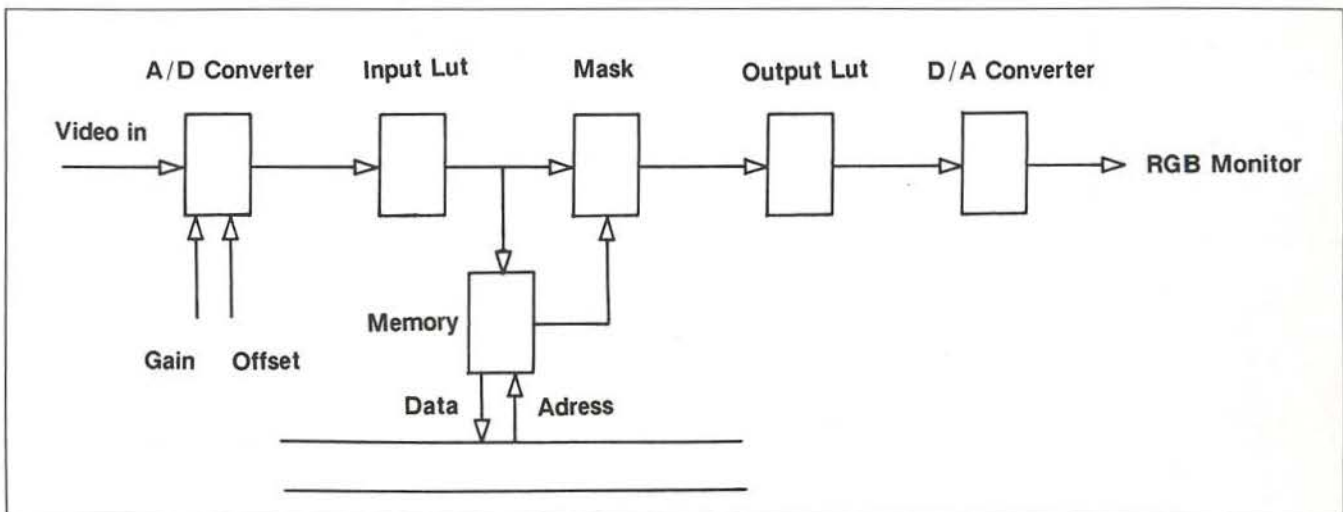


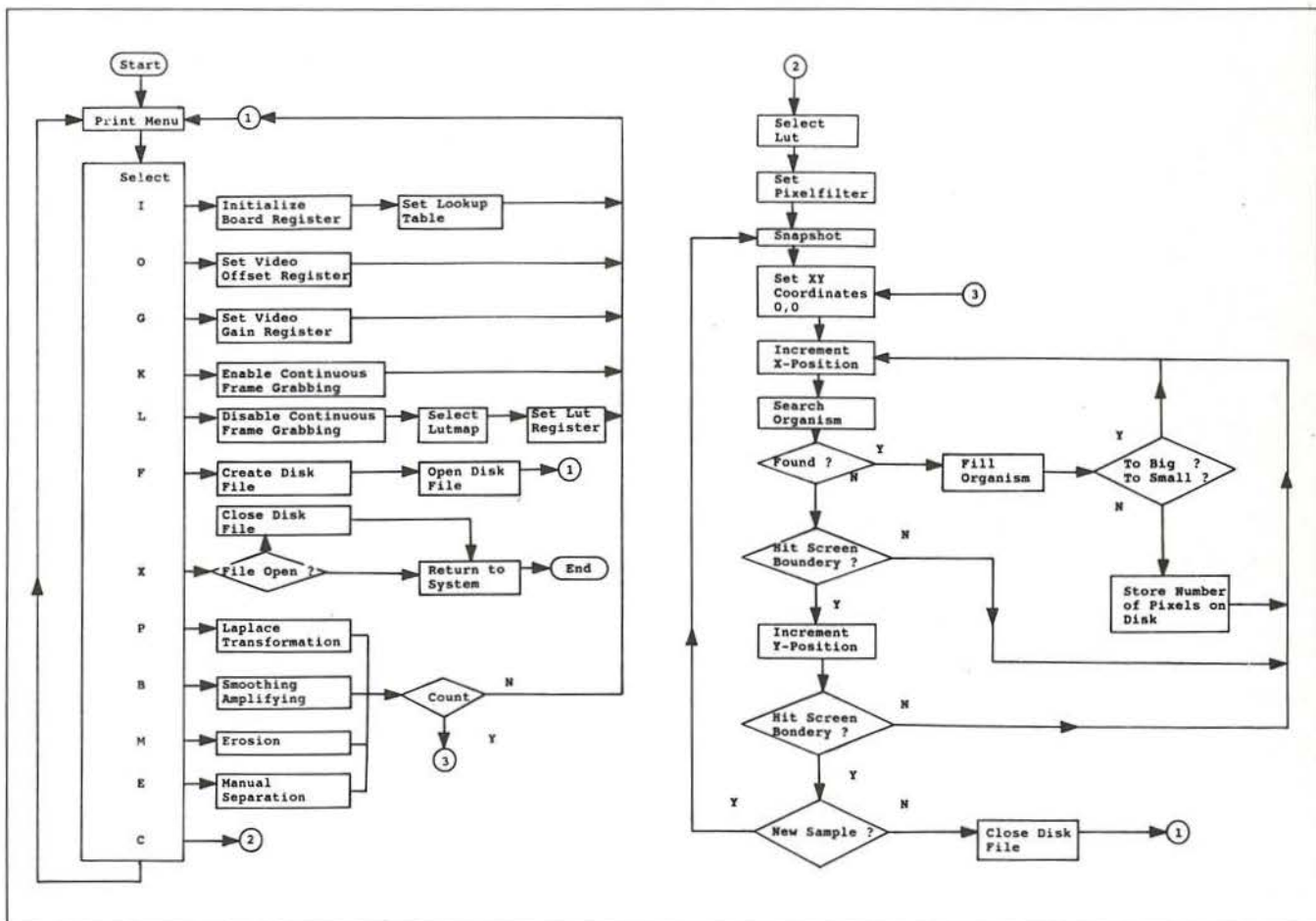
Fig. 2. Block diagram of the video digitizer showing the key features of the system. For further details see text

e.g. with text or geometrical patterns. The processed video image is then mapped in three parallel look-up tables which can be used for false color presentation (e.g. dark gray levels activate green, medium red and bright blue hues on the monitor). The signal is then converted by a digital/analog (D/A) converter and displayed on a RGB (color) analog monitor.

2.3 Software

The main program is written in 8088/8086 assembly language to facilitate fast computing and analysis. The program displays a main menu and upon selection of a feature an informative comment or question to be answered by the operator (Fig. 3).

Fig. 3. Flow chart of the counting algorithm with the additional options



Independent of the predefined values which are selected at the beginning of the program, a specific set of LUTs can be selected, the video gain and offset altered and either continuous or single frame grabbing chosen. All calculated data can be stored in a disk file for further statistical treatment. Before the count and area calculation the image can be manipulated by a number of useful mathematical techniques: Laplace filtering allows a specific edge detection by recalculating each individual pixel in the image using a kernel defined in a  $3 \times 3$  matrix.

$$\begin{array}{ccc} -1 & -1 & -1 \\ -1 & 8 & -1 \\ -1 & -1 & -1 \end{array} \quad (1)$$

The gray level of each pixel is multiplied by 8 and the values of the neighboring 8 pixels are subtracted; the result is written back into the pixel matrix (DAWSON, 1987). Changing the factor with which the central pixel is multiplied ( $<8$  or  $>8$ ) results in a substantial increase in contrast and in a visibility enhancement of continuous areas. Similarly, the noise level can be reduced drastically by averaging over a  $3 \times 3$  matrix of original pixel values  $P(t-1)$  and using this new pixel value  $P(t)$  at the row and column positions  $n$  and  $m$  for the reconstructed image.

$$P_{(n,m)}(t) = \frac{\sum_{i=n-1}^{n+1} \sum_{j=m-1}^{m+1} P_{(i,j)}(t-1)}{\sum_i * \sum_j} \quad (2)$$

The counting routine starts at the top left position and scans the image in horizontal rows until it detects an object which differs from the background by a predefined threshold value. It then tries to fill the area of the found organism by scanning all adjacent pixels of the found one and storing their  $x$  and  $y$  coordinates temporarily provided they belong to the same organism. These pixels are then taken as the new centers for a recursive search of their neighbors until the whole area has been analyzed and set to a different color or gray level. At this stage the number of organisms found is incremented and the number of pixels in the analyzed organism is stored in the disk file, provided the object is neither too big nor too small according to predefined limits in order to avoid disturbances from debris or optical artefacts. When the whole screen is scanned, the file can be closed or a new sample can be analyzed after initiating the digitization of a new image.

The raw data on disk are the basis for a statistical package written in Turbo Pascal. The individual values can be binned in a number of classes selected by the operator and presented in graphical form as column or line diagrams on the monitor from where they can be printed out as hard copy.

### 3. Results and Discussion

In order to determine the usefulness and accuracy of the hardware and software system, a number of tests were performed in which both motile organisms and non-moving objects were analyzed. The objects included hardware items of various shapes as well as biological objects such as electron microscopical images of proteins and metacystodes of *Taenia crassiceps*.

However, quantitative measurements were performed studying growing curves of the green flagellate, *Euglena gracilis*. Figure 4 shows the number of cells per ml in log units as counted with a Thoma chamber in comparison with the

automatic count in dependence of the age of the culture after inoculation. In addition to the convenient and fast determination of the cell density using the computer system, the cells do not need to be immobilized by glutaraldehyde fixation (2.5%) which is necessary for manual counting due to the fast swimming of the cells in culture; the digitization is performed on a snapshot taken in a 40 ms period. Another disadvantage of the manual count is the larger error at low and high densities due to minimal counts in the microscope image or a too crowded image. Using the repetition option of the automatic system a high degree of statistical significance can be achieved.

The comparison of the cell densities measured by the automatic system with the absorbance of the sample shows a close resemblance during growth of the culture (data not shown). The disadvantage of the spectrophotometric method is, however, that it is limited to low absorptions up to an extinction of about 0.5. Above this value scattering and multiple absorption causes distortions. Another problem is obvious in figure 5: The spectrophotometer cannot distinguish between the counted organisms and any debris or contaminations because it detects both absorbance and scattering as a signal. The culture in this comparison was contaminated by bacteria which took over in growth at about the fifth day after inoculation while the cell number of the flagellates remained constant and eventually even declined. The values measured by the spectrophotometer increased due to the scatter component, while the automatic system faithfully recorded the cell densities of the *Euglena* culture, because the lower threshold for organism size discriminated them from the smaller bacteria.

Fig. 4. Cell numbers of a culture as determined with a Thoma chamber (closed symbols) in comparison with the computer system (open symbols)

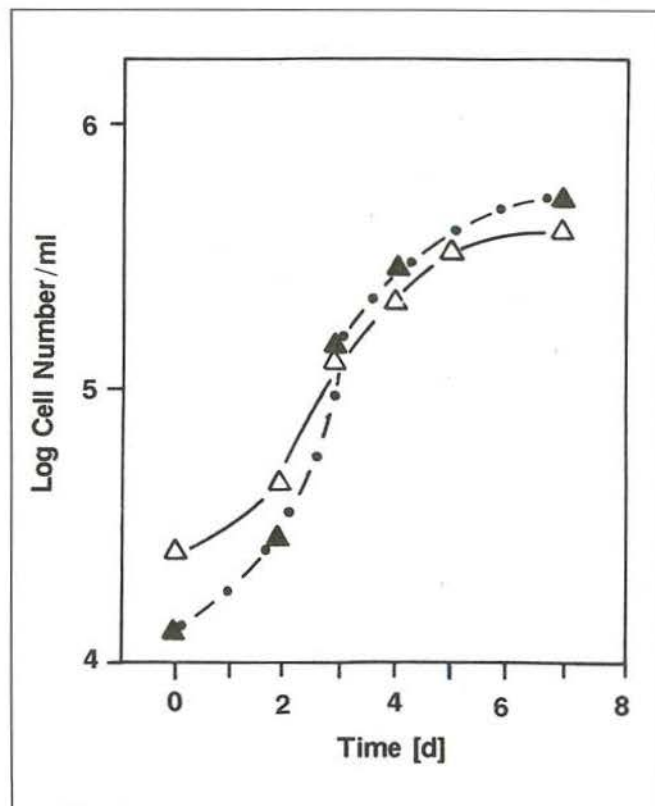
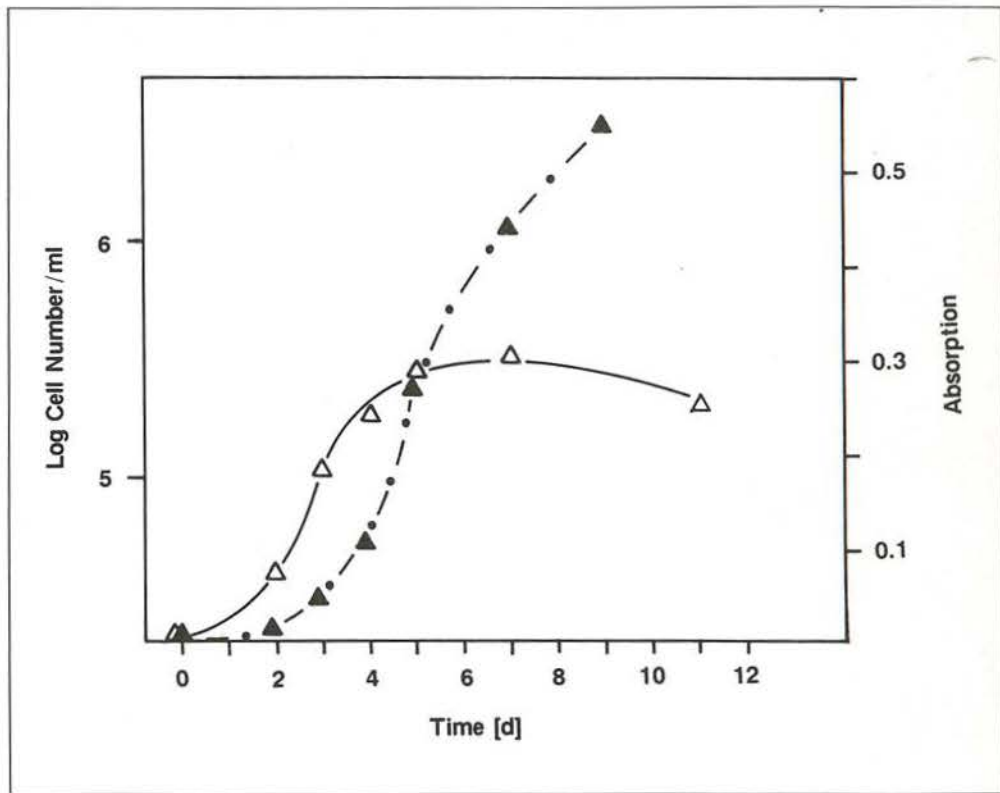




Fig. 5. Cell numbers of a culture as measured with the computer system (open symbols) in comparison with the absorption of the culture. The deviation between the two curves starting at about five days after inoculation is due to a bacterial contamination which disturbs the absorption measurements due to the scatter component



#### 4. Conclusions

The hardware and software system described in this article has been shown to be an accurate, fast and reliable tool to count organisms or any other objects on a macroscopic or microscopic scale. Due to the built-in mathematical techniques it allows to process even images of low quality which can be enhanced in contrast and/or sharpness before analysis. Proper size filtering allows to discriminate between objects of various sizes and can thus be used to distinguish between classes of organisms according to their size.

#### Acknowledgements

This work was supported by the Bundesminister für Forschung und Technologie (KBF 57) and by the Deutsche Forschungsgemeinschaft (Ha 985/5-6). The authors gratefully acknowledge the skilful technical assistance of A. HOHMANN and M. RUDYK.

#### Literature

- ASMUS, J. F., 1987: Digital image processing in art conservation. *Byte* **12**, No. 3, 151.
- BOYES, E. D., B. J. MUGGRIDGE, M. J. GORINGE. 1982: On-line image processing in high resolution electron microscopy. *J. Microsc.* **127**, 321.
- BRINKMANN, K., 1976: The influence of alcohols on the circadian rhythm and metabolism of *Euglena gracilis*. *J. Interdiscipl. Cycle Res.* **7**, 149.
- BRYAN, S. R., W. S. WOODWARD, D. P. GRIFFIS, R. W. LINTON, 1985: A microcomputer based digital imaging system for ion microanalysis. *J. Microscopy* **138**, 15.
- CHAUBAL, K. A., C. S. GODBOLE, V. S. AROSKAR, 1981: A simple method to measure cell area photometrically. *Indian J. Exp. Biol.* **20**, 220.
- DAWSON, B. M., 1987: Introduction to image processing algorithms. *Byte* **12**, No. 3, 169.
- DESAI, V., L. REIMER, 1985: Digital image recording and processing using an Apple II microcomputer. *Scanning* **7**, 185.
- DUSENBERY, D. B., 1985: Using a microcomputer and videocamera to simultaneously track 25 animals. *Comput. Biol. Med.* **15**, 169.
- ERHARDT, R., E. R. REINHARDT, W. SCHLIPF, W. H. BLOSS, 1980: FAZYTAN A system for fast automated cell segmentation, cell image analysis and feature extraction based on TV-image pickup and parallel processing. *Anal. Quant. Cytol. J.* **2**, 25.
- FINNEY, J. L., 1982: Monte Carlo techniques. *Biochem. Soc. Trans.* **10**, 305.
- GRAVEKAMP, C., H. K. KOERTEN, N. P. VERWOERD, W. C. DE BRUIJN, W. T. DAEMS, 1982: Automated image analysis applied to electron micrographs. *Cell Biol. Int. Rep.* **6**, 656.
- GUALTIERI, P., G. COLOMBETTI, F. LENCI, 1985: Automatic analysis of the motion of microorganisms. *J. Microscopy* **139**, 57.
- HÄDER, D. P., 1985: Effects of UV-B on motility and photobehavior in the green flagellate, *Euglena gracilis*. *Arch. Microbiol.* **141**, 159.
- HÄDER, D. P., 1986: Effects of solar and artificial UV irradiation on motility and phototaxis in the flagellate, *Euglena gracilis*. *Photochem. Photobiol.* **44**, 651.
- HÄDER, D. P., 1987: Polarotaxis, gravitaxis and vertical phototaxis in the green flagellate, *Euglena gracilis*. *Arch. Microbiol.* **147**, 179.
- HÄDER, D.-P., M. LEBERT, 1985: Real time computer-controlled tracking of motile microorganisms. *Photochem. Photobiol.* **42**, 509.
- JAFFE, M. J., A. H. WAKEFIELD, F. TELEWSKI, E. GULLEY, R. BIRO, 1985: Computer-assisted image analysis of plant growth, thigmomorphogenesis, and gravitropism. *Plant Physiol.* **77**, 722.
- KEMNITZ, H.-D., H. P. HOUGARDY, 1980: Design of an interface for digital image storage. *Mikroskopie* **7**, 415.
- KINCAID, D. T., R. B. SCHNEIDER, 1983: Quantification of leaf shape with a microcomputer and Fourier transform. *Can. J. Bot.* **61**, 2333.
- KOTHARI, M. J., G. L. SHAH, 1975: Epidermal structures and ontogeny of stomata in the *Papilionaceae* (Tribe *Heydysareae*). *Bot. Gaz.* **136**, 372.
- LEWANDOWSKA, K., J. DOROSZEWSKI, G. HAEMMERLI, P. STRÄULI, 1979: An attempt to analyze locomotion of leukemia cells by computer image processing. *Comput. Biol. Med.* **9**, 331.

- MIKOLAJCZYK, M., D. P. HÄDER, W. NULTSCH, 1985: Photodynamically induced chemoresponses of the colorless flagellate, *Astasia longa*, in the presence of riboflavin. Arch. Microbiol. **142**, 397.
- NULTSCH, W., D.-P. HÄDER, 1973: Negative photo-phobotactic reactions in *Phormidium uncatum*. Photochem. Photobiol. **18**, 311.
- NULTSCH, W., D.-P. HÄDER, 1974: Über die Rolle der beiden Photosysteme in der Photo-phobotaxis von *Phormidium uncatum*. Ber. Deutsch. Bot. Ges. **87**, 83.
- OMASA, K., M. ONOE, 1984: Measurement of stomatal aperture by digital image processing. Plant & Cell Physiol. **25**, 1379.
- PRESTON, K. jr., A. DEKKER, 1980: Differentiation of cells in abnormal human liver tissue by computer image processing. Anal. Quant. Cytol. J. **2**, 14.
- SERRA, J., 1980: Digitalization. Mikroskopie **37** Suppl., 109.
- SHIPTON, H. W., 1979: The microprocessor, a new tool for the biosciences. Ann. Rev. Biophys. Bioeng. **8**, 269.
- SKARNULIS, A. J., 1982: A computer system for on-line image capture and analysis. J. Microsc. **127**, 39.
- STARR, R. C., 1964: The culture collection of algae at Indiana University. Am. J. Bot. **51**, 1013.
- TERAMURA, A. H., 1980: Effects of ultraviolet-B irradiances on soybean. I. Importance of photosynthetically active radiation in evaluating ultraviolet-B irradiance effects on soybean and wheat growth. Physiol. Plant. **48**, 333.
- TERAMURA, A. H., M. C. PERRY, J. LYDON, M. S. MCINTOSH, E. G. SUMMERS, 1984: Effects of ultraviolet-B radiation on plants during mild water stress. III. Effects on photosynthetic recovery and growth in soybean. Physiol. Plant. **60**, 484.

Date of receipt: June 26th, 1987.

The authors' address: Dr. Donat-P. Häder and Kai Griebenow, Fachbereich Biologie – Botanik, Philipps-Universität Marburg, Karl-von-Frisch-Straße, D-3550 Marburg, Federal Republic of Germany.

# Ein Weg zum exakten Testen von größeren Kontingenztafeln

H. Hochadel

## Zusammenfassung

Eine Formel wird angegeben zur Berechnung der Summe der Wahrscheinlichkeiten von Kontingenztafeln, die sich aus der vollständigen Variation einer Untermatrix ergeben. Diese Formel spart eine wesentliche Anzahl von Rechenschritten in einem Algorithmus für den exakten Test von Kontingenztafeln.

## Summary

A formula is given to compute the sum of probabilities of contingency-tables, that are built by the complete variation of a subtable. The formula will spare a lot of steps in an algorithm for the exact test of contingency-tables.

Bei größeren Kontingenztafeln scheitert ein exakter Test noch immer an der Zahl der Rechenschritte. Die nachstehenden Hilfen können die Zahl der Rechenschritte wesentlich verringern.

## 1. Teilsommen

Sei  $M_0 = (n_{ij})$  eine gegebene Kontingenztafel, so ist ihre Wahrscheinlichkeit mit

$$(1) W(M_0) = \frac{\prod n_{i.}! \cdot \prod n_{.j}!}{n! \cdot \prod n_{ij}!}$$

gegeben.

Bezeichnen wir die Randsummen  $n_{.j}$  und  $n_{i.}$  kurz mit  $R_0$  und schreiben

$$(1) W(M_0) = \frac{R_0!}{n! \cdot M_0!}$$

Zur Durchführung des exakten Tests müssen alle Tafeln  $M$  gebildet werden, welche den gegebenen Rand  $R$  besitzen. Eine solche Tafel sei mit  $M|R$  bezeichnet. Der exakte Test besteht damit aus der Bildung von zwei Summen  $S_1$  und  $S_2$  mit

$$S_1 = \sum W(M|R) \text{ mit } W(M) \leq W(M_0)$$

$$S_2 = \sum W(M|R) \text{ mit } W(M) > W(M_0)$$

Insgesamt gilt:

$$S_1 + S_2 = \sum_{M|R} W(M|R) = 1$$

$$(2) \sum_{M|R} W(M|R) = \sum \frac{R!}{n!M!} = \frac{R!}{n!} \cdot \sum_{M|R} \frac{1}{M!} = 1$$

Daraus folgt:

$$(3) \sum_{M|R} \frac{1}{M!} = \frac{n!}{R!}$$

Diese Formel (3) läßt sich auch anwenden, wenn  $M$  eine Teilmatrix einer größeren Kontingenztafel ist.

Sei  $M = (n_{ij})$  eine Kontingenztafel und  $M_0$  eine Teilmatrix in  $M$  mit dem Rand  $R_0$ , dann gilt:

$$(4) \sum_{M_0|R_0} W(M) = \frac{R!}{n!} \cdot \frac{1}{\prod_{ij \in M_0} n_{ij}} \cdot \frac{n_0!}{R_0!}$$

## 2. Ein Schritt zum Algorithmus

Die Formel (4) läßt sich dann für einen Algorithmus verwenden, wenn alle in der Summe erfaßten Tafeln entweder zur Summe  $S_1$  oder alle zur Summe  $S_2$  gehören.

Diese Entscheidung kann getroffen werden.

Es gilt:

$$(5) \max W(M|R) \Leftrightarrow n_{ij} = \frac{n_{i.} \cdot n_{.j}}{n}$$

(6)  $\min W(M|R) \Leftrightarrow n_{ij}$  haben möglichst maximale und minimale Werte.

Die Formel (5) ergibt sich z. B. aus der ML-Schätzung, Formel (6) aus  $n_1! \cdot n_2! \leq (n_1 + n_2)!$ . Beide Formeln sind Eigenschaften der Fakultätsfunktion und gelten auch für eine Teilmatrix  $M_0$  innerhalb einer Matrix  $M$ .

[Eine Tafel für das Minimum läßt sich leicht konstruieren: Man suche die Maxima der Zeilen- und Spaltensummen und setze am Kreuzungspunkt der kleineren der beiden Werte ein und fülle die restliche Zeile bzw. Spalte mit Nullen auf. Jetzt setzt man das Verfahren mit der verkleinerten Tafel fort.

Falls man mit Fakultäten für rationale Zahlen arbeiten will, stellt die Formel 5 eine obere Schranke für die Wahrscheinlichkeiten dar. Anderenfalls muß man die Formel 5 auf ganze Zahlen runden. Bei einem verbleibenden Überschuß in einer Reihe oder Spalte muß man den größten Wert reduzieren; wird eine Summe zu klein, muß der kleinste Wert erhöht werden.]

Die Formeln 4, 5, 6 lassen sich unschwer in einen Algorithmus einbinden, der aus einer schrittweisen Veränderung der Tafeln  $M|R$  die Summen  $S_1$  und  $S_2$  errechnet.

$M_0$  besteht dabei immer aus den Spalten (bzw. Zeilen), die von der schrittweisen Veränderung nicht erfaßt werden.

Zu klären ist noch die Frage der Pivotalisierung. Es erscheint effektiver, wenn  $M_0$  möglichst große Randwerte besitzt. Jedoch läßt sich die Grenze zwischen den Summen  $S_1$  und  $S_2$  in Spalten mit großen Randsummen leichter ansteuern. Ebenso erscheint es günstig, wenn der Algorithmus mit der Tafel der kleinsten Wahrscheinlichkeit startet.

### 3. Beispiel

Für eine  $4 \times 5$  Tafel  $M$  mit vorgegebenen Randsummen und festgelegter erste Spalte

3		6
3		6
2	$M_0$	6
2		6
10	8 2 2 2	24

		3
		3
	$M_0$	4
		4
8	2 2 2 2	14

gilt nach Formel (4):

$$\sum_{M_0|R_0} W(M) = \frac{10!8!2!2!6!6!6!}{24!} \cdot \frac{1}{3!3!2!2!} \cdot \frac{14!}{8!2!2!3!3!4!4!}$$

In dieser Summe sind über 1000 Tafeln enthalten. Das Maximum und Minimum unter diesen Summanden wird bei den Tafeln

2	1	0	0	3
2	0	1	0	3
2	1	0	1	4
2	0	1	1	4
8	2	2	2	14

0	2	0	1	3
0	0	2	1	3
4	0	0	0	4
4	0	0	0	4
8	2	2	2	14

angenommen. Man erkennt leicht, daß die Extremwerte auch mit anderen Anordnungen erreicht werden, jedoch nicht mit anderen Werten.

### Literatur

GAIL, MANTEL: »Counting the Number of  $r \times c$  Contingency Tables with fixed Margins.« Journal of the American Statistical Association, Vol. 72/360.

Eingegangen am 28. Januar 1987.  
 Anschrift des Verfassers: H. Hochadel, c/o BASF AG, Abtl. ZLJ/N, Bau Z 35, D-6700 Ludwigshafen/Rh.

## Eine verschärfte Holm-Prozedur zum Vergleich aller Mittelwertpaare

Th. Royen

### Zusammenfassung

Es wird eine multiple sequentiell ablehnende Testprozedur zum Vergleich aller Paare aus  $k$  Mittelwerten in einer Varianzanalyse vorgeschlagen. Diese Prozedur ist gleichmäßig schärfer (im Sinne von POPPER-SHAFFER, 1986) als die Anwendung der von POPPER-SHAFFER gegebenen allgemeinen modifizierten Bonferroni-Holm-Prozedur auf dieses Testproblem.

Schranken zum multiplen Niveau  $\alpha = 0,05$  werden angegeben für  $k = 3$  (1) 10 und 42 ausgewählte Freiheitsgrade zwischen 4 und  $\infty$ .

### Summary

A multiple sequentially rejective test procedure for the comparison of all pairs of  $k$  means in the Anova is proposed. This procedure is uniformly more powerful (cf. POPPER-SHAFFER, 1986) than the application of the general modified Bonferroni-Holm procedure of POPPER-SHAFFER to this test problem.

For the multiple level  $\alpha = 0,05$ , bounds are given for  $k = 3$  (1) 10 and 42 selected degrees of freedom between 4 and  $\infty$ .

### 1. Einleitung

Die klassischen nichtsimultanen Testprozeduren von NEWMAN-KEULS, DUNCAN und anderen Autoren zum Vergleich aller Mittelwerte untereinander bei einer Varianzanalyse halten das multiple Niveau nicht ein (vgl. BERTCHIER, 1981; EINOT und GABRIEL, 1975; SONNEMANN, 1982).

Die verbesserte abgeschlossene Newman-Keuls-Prozedur nach BEGUN und GABRIEL (1981) zur Prüfung der Homogenität aller Teilmengen von  $k$  Mittelwerten ist für die praktische Durchführung verhältnismäßig aufwendig und nur mit entsprechenden Programmen möglich. Die Autoren meinen selbst: "One objection to this procedure, however, may be the difficulty of explaining its logic to experimenters."

Allgemein sequentiell ablehnende Bonferroni-Holm-Prozeduren verwenden konservative Bonferroni-Schranken bzw. entsprechende Überschreitungswahrscheinlichkeiten (vgl. HOLM, 1979). Durch POPPER-SHAFFER (1986) wurde kürzlich eine verschärfte Bonferroni-Holm-Prozedur vorgeschlagen. Speziell für die Anwendung auf den Vergleich von  $k$  Mittelwerten untereinander läßt sich diese Methode noch verbessern bis hin zur Ausschöpfung des vereinbarten multiplen Niveaus  $\alpha$ , indem die Bonferroni-Schranken durch dem speziellen

Problem angepaßte Schranken ersetzt werden, die auf den Maxima mehrerer Spannweiten beruhen.

Da diese Schranken alle kleiner oder gleich den entsprechenden Bonferroni-Schranken der Methode nach POPPER-SHAFFER sind, ist die hier vorgeschlagene Prozedur von gleichmäßig höherer Power.

### 2. Methode

Es seien  $k$  vollständig unabhängige Stichproben vom gleichen Umfang  $n$  (zur Anwendung auf Stichproben ungleichen Umfangs siehe weiter unten in diesem Abschnitt) aus  $k$  Normalverteilungen mit Erwartungswerten  $\mu_i$  und gleicher Varianz  $\sigma^2$  gegeben.

$s^2$  bezeichnet eine auf  $\nu$  Freiheitsgraden basierende von den Mittelwerten  $\bar{y}_1, \dots, \bar{y}_k$  unabhängige Schätzung der Varianz  $\sigma^2$ ,  $H_{ij}^0$  die Einzelnullhypothese  $\mu_i = \mu_j$ ,  $H_{ij}^1$  die entsprechende zweiseitige Alternative  $\mu_i \neq \mu_j$  und  $T_{ij} = \sqrt{n} |\bar{y}_i - \bar{y}_j|/s$  die zugehörige Teststatistik ( $1 \leq i < j \leq k$ ).

Eine multiple sequentiell ablehnende Holm-Prozedur für den Test aller Hypothesen  $H_{ij}^0$  gegen  $H_{ij}^1$  geht von den geordneten Teststatistiken  $T_{(K)} \dots T_{(1)}$  aus ( $K = \binom{k}{2}$ ). Die entsprechenden Hypothesen seien  $H_{(K)}^0, \dots, H_{(1)}^0$ . Beginnend mit dem höchsten Rang  $r = K$  wird  $H_{(r)}^0$  abgelehnt, wenn  $T_{(r)} \geq c(\alpha, r)$  mit geeigneten Schranken  $c(\alpha, K) \dots c(\alpha, 1)$ . Wenn zum ersten Mal  $T_{(r)} < c(\alpha, r)$  gilt, endet die Prozedur dort, und die verbleibenden Hypothesen  $H_{(r)}^0, \dots, H_{(1)}^0$  werden nicht abgelehnt.

Die Prozedur hält das multiple Niveau  $\alpha$  (vgl. SONNEMANN, 1982) ein, wenn die Wahrscheinlichkeit  $p$  für mindestens einen Fehler 1. Art durch  $\alpha$  beschränkt bleibt bei beliebigen Werten der Parameter  $\mu_1, \dots, \mu_k$ .

Entsprechend ihren verschiedenen Werten können  $\mu_1, \dots, \mu_k$  in disjunkte Blöcke untereinander gleicher Werte zusammengefaßt werden.  $B_1, \dots, B_b$  seien die Mengen der zugehörigen Stichprobenindizes, die den nicht trivialen Blöcken – d. h. vom Umfang  $> 1$  – entsprechen. Ein Fehler 1. Art kann nur auftreten, wenn mindestens ein solcher Block existiert. Hat der  $\ell$ -te Block den Umfang  $k_\ell$ , so ist die Anzahl der wahren Einzelnullhypothesen durch  $\sum_{\ell=1}^b \binom{k_\ell}{2}$  gegeben.  $W_\ell = \max_{i,j \in B_\ell} T_{ij}$  sei die studentisierte Spannweite zum Block Nr.  $\ell$  und  $q(\alpha; k_1, \dots, k_b; \nu)$  die obere  $\alpha$ -Schranke für  $\max_{1 \leq \ell \leq b} W_\ell$ .

Ein Fehler 1. Art tritt nur auf, wenn die wahre Hypothese  $H_{(R)}^0$  mit dem höchsten Wert  $T_{(R)}$  abgelehnt wird.

Bezeichnet  $D(k,r)$  für irgendeinen Rang  $r \leq \binom{k}{2}$  die Menge aller möglichen »Zerlegungen«  $(k_1, \dots, k_b)$ , für welche  $\sum_{i=1}^b \binom{k_i}{2} \leq r$  und zwangsläufig  $\sum_{i=1}^b k_i \leq k$  gilt, so erkennt man nun, daß die sequentiell ablehnende Testprozedur, welche die Schranken  $c(\alpha; k, r, v) := \max_{(k_1, \dots, k_b) \in D(k,r)} q(\alpha; k_1, \dots, k_b; v)$  verwendet, das multiple Niveau  $\alpha$  einhält. Ist nämlich  $R$  der Rang der größten Teststatistik, die zu einer wahren Nullhypothese gehört, so gilt wegen  $R \geq \sum_{i=1}^b \binom{k_i}{2}$  für die Wahrscheinlichkeit  $p$  des Auftretens von mindestens einem Fehler 1. Art:

$$\begin{aligned} p &\leq P\{T_{(R)} \geq c(\alpha; k, R, v)\} \\ &= P\{\max_{1 \leq i \leq b} W_i \geq c(\alpha; k, R, v)\} \\ &\leq P\{\max_{1 \leq i \leq b} W_i \geq q(\alpha; k_1, \dots, k_b; v)\} = \alpha, \end{aligned}$$

wobei  $k_1, \dots, k_b$  hier die Umfänge der tatsächlichen Blöcke bezeichnen ( $p = 0$  für  $b = 0$ ).

Nach HAYTER (1984) sind die Tukey-Kramer-Schranken für studentisierte »Spannweiten« von Mittelwerten  $\bar{y}_i$  aus Stichproben ungleichen Umfangs zu beliebigem Niveau konservativ. Genauer gilt:

$$P\left\{ \max_{1 \leq i < j \leq k} \frac{|\bar{y}_i - \bar{y}_j|}{\sigma \sqrt{\frac{1}{2} \left( \frac{1}{n_i} + \frac{1}{n_j} \right)}} \leq c \right\}$$

wird für jedes  $c > 0$  strikt minimal, falls alle  $n_i$  identisch sind. Nach Studentisierung folgt für obige Testprozedur, daß sie konservativ bzgl. des multiplen Niveaus  $\alpha$  bleibt, wenn sie mit den gleichen Schranken auf die Teststatistiken

$$T_{ij} := \frac{|\bar{y}_i - \bar{y}_j|}{s \sqrt{\frac{1}{2} \left( \frac{1}{n_i} + \frac{1}{n_j} \right)}} \text{ angewandt wird.}$$

Eine asymptotische Anwendung auf nichtparametrische Vergleiche ist möglich. Liegen  $k$  unabhängige Stichproben vom Umfang  $n$  aus  $k$  stetigen Verteilungen vor, die sich nur durch Translationen voneinander unterscheiden, so gilt mit den Wilcoxon-Rangsummen  $S_{ij}$  des Wilcoxon-Zweistichprobentests für  $H_{ij}^0$ , daß obige Testprozedur mit den Teststatistiken

$$T_{ij} := \frac{|S_{ij} - n(n + \frac{1}{2})| - \frac{1}{2}}{\sqrt{\frac{n^2(2n+1)}{24}}} \quad (-\frac{1}{2}: \text{Stetigkeitskorrektur})$$

und den Schranken  $c(\alpha; k, r) := \max_{(k_1, \dots, k_b) \in D(k,r)} q(\alpha; k_1, \dots, k_b; \infty)$

für  $T_{(r)}$  asymptotisch für  $n \rightarrow \infty$  das Niveau  $\alpha$  einhält. Dies folgt aus der asymptotischen multivariaten Normalverteilung der zu einem Block identischer Verteilungen gehörenden standardisierten Wilcoxon-Rangsummen (vgl. MILLER, 1981, Chapter 4,4). Für  $n \geq 15$  ist die Normalapproximation sicher anwendbar.

Das Prinzip obiger Testprozedur kann auf andere paarweise Vergleiche übertragen werden, z. B. auf die Vergleiche von  $k$  Polynomialverteilungen (vgl. ROYEN, 1984). Hierzu würden jedoch Schranken für die Maxima mehrerer multivariater Spannweiten benötigt.

### 3. Tafeln

Zum Vergleich von  $k$  Mittelwerten werden die Teststatistiken nach ihrem Rang geordnet.  $T_{(2)} \geq \dots \geq T_{(r)} \geq \dots \geq T_{(1)}$ . In der mit  $k$  überschriebenen Spalte von Tafel 1 findet man die zum Rang  $r$  gehörende Zerlegung  $k_1, \dots, k_b$ . Falls  $r$  nicht angegeben ist, entnimmt man die nächst höher stehende Zerlegung. So ist z. B. für  $k = 6$  und  $r = 7$  oder  $r = 8$  die zugehörige Zerlegung  $k_1 = 4, k_2 = 2$ .

Zur Zerlegung  $k_1, \dots, k_b$  findet man aus der Tafel 2 in der Spalte mit dem richtigen Freiheitsgrad  $v$  den entsprechenden Schrankenwert  $q(\alpha; k_1, \dots, k_b; v)$  für  $T_{(r)}$ . Die Reihenfolge der Zeilen (Zerlegungen) in Tafel 2 ist entsprechend der Rangfolge der Schranken für  $v = \infty$  gewählt. Für kleinere Freiheitsgrade ändert sich diese Rangfolge der Schranken an einigen Stellen. In solchen Fällen müssen gewisse Zerlegungen übergangen werden, die in Tafel 1 jeweils angegeben sind. Damit nicht wegen fehlender Freiheitsgrade zu oft interpoliert werden muß, wurden alle Freiheitsgrade  $v \leq 60$  mitaufgenommen, die keine Primfaktoren  $\geq 11$  enthalten, da bei vielen Versuchsplänen  $v$  ein Produkt kleiner Zahlen ist. Ferner sind Tafelwerte für  $v = 60, 65, 70, 75, 80, 90, 100, 120, 150, 200, 300$  und  $\infty$  angegeben.

Ein einfaches Zahlenbeispiel zeige den Gebrauch der Tafeln: aus  $k = 5$  Stichproben zu je  $n = 4$  Werten seien die Mittelwerte  $\bar{y}_1 = 5,8, \bar{y}_2 = 6,4, \bar{y}_3 = 6,8, \bar{y}_4 = 10,8, \bar{y}_5 = 11,5$  und  $s = 2,0$  mit  $v = 5 \cdot 3 = 15$  Freiheitsgraden berechnet worden. Mit den Teststatistiken  $T_{ij} = \sqrt{n}|\bar{y}_i - \bar{y}_j|/s = |\bar{y}_i - \bar{y}_j|$  erhält man zum multiplen Niveau  $\alpha = 0,05$ :

Rang r	Teststatistik $T_{(r)}$	Zerlegung aus Tafel 1	Schranke aus Tafel 2	Signifikanz
10	$T_{1,5} = 5.7$	5	4.367	*
9	$T_{2,5} = 5.1$	4	4.076	*
8	$T_{1,4} = 5.0$	4	4.076	*
7	$T_{3,5} = 4.7$	4	4.076	*
6	$T_{2,4} = 4.4$	4	4.076	*
5	$T_{3,4} = 4.0$	3,2	3.899	*
4	$T_{1,3} = 1.0$	3,2	3.899	← Ende der Prozedur
3	$T_{4,5} = 0.7$	3	3.674	—
2	$T_{1,2} = 0.6$	2,2	3.499	—
1	$T_{2,3} = 0.4$	2	3.014	—

Der Wert  $T_{3,4} = 4$  wäre auch nach der durch POPPER-SHAFER verschärfte Bonferroni-Holm-Prozedur nicht mehr signifikant. Mit wachsender Anzahl zu vergleichender Mittelwerte werden die Unterschiede zwischen den vertafelten und den Bonferroni-Schranken größer.

Ein-Prozent-Schranken sind auf Wunsch vom Autor erhältlich.

### 4. Numerische Berechnung der Schranken

Zur Berechnung der Wahrscheinlichkeiten  $P\{\max_{1 \leq i \leq b} W_i \leq c\} = \frac{2 p^p}{\Gamma(p)} \int_0^\infty \left( \prod_{i=1}^b \left( k_i \int_{-\infty}^\infty (\Phi(x+cs) - \Phi(x))^{k_i-1} \varphi(x) dx \right) \right) s^{2p-1} e^{-ps^2} ds$  ( $p := v/2, \Phi :=$  Standardnormalverteilungsfunktion,  $\varphi :=$  zugehörige Dichte).

wurde der Integrationsbereich durch einen beschränkten Bereich mit einem Abschneidefehler  $\epsilon < 10^{-8}$  ersetzt. Dieser Bereich wurde in eine von den Parametern abhängige Anzahl (mindestens 16 bis höchstens 56) Teilbereiche zerlegt, deren

Beiträge durch Gaussquadraturen mit 10-Punkte-Formeln berechnet wurden. Vergleiche mit Quadraturen noch höherer Genauigkeit zeigen, daß die berechneten Wahrscheinlichkeiten auf 7 Dezimalstellen genau sind (mindestens 6 Dezimalstellen für  $v > 150$ ). Die Schranken wurden durch iterative quadratische Interpolation ermittelt. Vertafelt wurde der Wert, der im Vergleich zu den um  $\pm 0,001$  veränderten Werten die am dichtesten an  $1 - \alpha$  liegende Wahrscheinlichkeit ergab. Einige der so berechneten Schranken für die studentisierte Spannweite lägen um 0,001 niedriger als die von HARTER (1960) vertafelten Werte. Da somit letztere stets auf der sicheren Seite liegen und der Unterschied praktisch bedeutungslos ist, wurden HARTERS Werte übernommen.

**Literatur**

BEGUN, J. M. und K. R. GABRIEL (1981): Closure of the Newman-Keuls multiple comparisons procedures. *JASA* **76**, 241-245.  
 BERCHIER, P. (1981): Mittelwertvergleiche in Normalverteilungsmo-  
 dellern – univariate Verfahren. ROES-Seminar, Bad Ischl.

EINOT, I. und K. R. GABRIEL (1975): A study of the powers of several methods of multiple comparisons. *JASA* **70**, 574-583.  
 HARTER, H. L. (1960): Tables of range and studentized range. *AMS* **31**, 1122-1147.  
 HAYTER, A. J. (1984): A proof of the conjecture that the Tukey-Kramer multiple comparisons procedure is conservative. *Ann. of Stat.* **12**, 61-75.  
 HOLM, ST. (1979): A simple sequentially rejective multiple test procedure. *Scand. Stat.* **6**, 65-70.  
 MILLER, R. G., jr. (1981): Simultaneous statistical inference, 2<sup>nd</sup> ed., Springer, New York.  
 POPPER-SHAFFER, J. (1986): Modified sequentially rejective multiple test procedures. *JASA* **81**, 826-831.  
 ROYEN, TH. (1984): Multiple comparisons of polynomial distributions. *Biometrical Journal* **26**, 319-332.  
 SONNEMANN, E. (1982): Allgemeine Lösungen multipler Testprobleme. *EDV in Medizin und Biologie* **13**, 120-128.

Eingegangen am: 23. April 1987.  
 Anschrift des Verfassers: Prof. Dr. Th. Royen, Fachhochschule des Landes Rheinland-Pfalz, Abteilung Bingen, Rochus-Allee 4, D-6530 Bingen.

Tafel 1<sup>1)</sup>

k = 3		4		5		6	
Rang	Zerlegung	Rang	Zerlegung	Rang	Zerlegung	Rang	Zerlegung
3	3	6	4	10	5	15	6
2	2	5	3	9	4	14	5
		2	2,2	5	3,2	9	4,2
		1	2	3	3	6	3,3
				2	2,2	5	3,2
				1	2	3	2,2,2
						2	2,2
						1	2

10		9		8		7	
Rang	Zerlegung	Rang	Zerlegung	Rang	Zerlegung	Rang	Zerlegung
45	10	36	9	28	8	21	7
44	9	35	8	27	7,2	20	6
35	8,2	27	7,2	20	6,3	14	5,2
28	8	21	7	17	5,4	10	5 v > 34
27	7,3	15	6 v > 6	15	6 v > 6	9	4,3
23	6,4	14	5,3	14	5,3	8	4,2,2
20	5,5	12	4,4	12	4,4	7	3,3,2
19	6,3	11	5,2 v > 29	11	5,2	6	3,3
17	6,2,2 v > 5	10	4,3,2	10	5 v > 34	5	3,2,2
16	5,4	9	3,3,3	9	4,3	4	3,2
15	5,3,2	8	4,2,2	8	4,2,2	3	2,2,2
13	4,4,2	7	3,3,2	7	3,3,2	2	2,2
12	4,3,3	6	3,2,2,2	6	3,3	1	2
11	5,2 v > 29	5	3,2,2	5	3,2,2		
10	4,3,2	4	2,2,2,2	4	2,2,2,2		
9	3,3,3	3	2,2,2	3	2,2,2		
8	3,3,2,2	2	2,2	2	2,2		
7	3,3,2	1	2	1	2		
6	3,2,2,2						
5	2,2,2,2,2						
4	2,2,2,2						
3	2,2,2						
2	2,2						
1	2						

Falls der Freiheitsgrad  $v$  nicht größer als der angegebene Wert ist, so ist die dort angegebene Zerlegung durch die darunter stehende zu ersetzen, da dann zu dieser »tieferen« Zerlegung die höhere Schranke in Tafel 2 gehört.

<sup>1)</sup> Erklärung in Abschnitt 3.

Tafel 2.1. (Niveau Alpha = 0.05)

Freiheitsgrad:	4	5	6	7	8	9	10
Zerlegung :							
10 :	7.826	6.995	6.493	6.158	5.918	5.739	5.599
9 :	7.602	6.802	6.319	5.998	5.767	5.595	5.461
8 2 :	7.433	6.653	6.184	5.871	5.647	5.479	5.349
8 :	7.347	6.582	6.122	5.815	5.597	5.432	5.305
7 3 :	7.313	6.548	6.086	5.779	5.559	5.394	5.266
6 4 :	7.242	6.484	6.027	5.723	5.505	5.342	5.215
7 2 :	7.161	6.420	5.973	5.676	5.464	5.305	5.181
5 5 :	7.218	6.463	6.007	5.704	5.487	5.324	5.198
7 :	7.053	6.330	5.895	5.606	5.399	5.244	5.124
6 3 :	7.038	6.310	5.872	5.581	5.372	5.216	5.095
6 2 2 :	6.972	6.253	5.821	5.533	5.327	5.173	5.054
5 4 :	6.977	6.255	5.821	5.532	5.326	5.171	5.051
6 2 :	6.848	6.150	5.730	5.451	5.252	5.103	4.987
5 3 2 :	6.860	6.153	5.728	5.445	5.243	5.092	4.974
6 :	6.707	6.033	5.628	5.359	5.167	5.024	4.912
4 4 2 :	6.822	6.119	5.697	5.415	5.214	5.064	4.947
5 3 :	6.725	6.040	5.629	5.355	5.160	5.013	4.900
4 3 3 :	6.771	6.074	5.655	5.375	5.176	5.027	4.911
4 4 :	6.684	6.004	5.595	5.322	5.128	4.983	4.870
5 2 2 :	6.643	5.969	5.564	5.295	5.103	4.959	4.847
5 2 :	6.479	5.832	5.444	5.186	5.002	4.865	4.758
4 3 2 :	6.541	5.878	5.479	5.214	5.025	4.884	4.774
3 3 3 :	6.479	5.823	5.429	5.166	4.979	4.839	4.731
4 2 2 2 :	6.449	5.797	5.406	5.146	4.960	4.821	4.714
5 :	6.287	5.673	5.305	5.060	4.886	4.756	4.654
4 3 :	6.365	5.730	5.349	5.096	4.916	4.781	4.676
3 3 2 2 :	6.384	5.739	5.352	5.094	4.911	4.774	4.667
4 2 2 :	6.260	5.639	5.266	5.018	4.841	4.709	4.607
3 3 2 :	6.187	5.573	5.204	4.960	4.785	4.655	4.554
4 2 :	6.035	5.450	5.100	4.867	4.701	4.577	4.481
3 2 2 2 :	6.070	5.470	5.110	4.871	4.700	4.573	4.475
3 3 :	5.950	5.374	5.028	4.799	4.636	4.514	4.419
2 2 2 2 2 :	5.944	5.358	5.007	4.774	4.608	4.484	4.388
4 :	5.757	5.218	4.896	4.681	4.529	4.415	4.327
3 2 2 :	5.812	5.252	4.916	4.693	4.535	4.416	4.324
2 2 2 2 :	5.661	5.118	4.793	4.577	4.423	4.308	4.219
3 2 :	5.485	4.976	4.671	4.469	4.325	4.218	4.135
2 2 2 :	5.296	4.807	4.515	4.321	4.183	4.080	4.001
3 :	5.040	4.602	4.339	4.165	4.041	3.949	3.877
2 2 :	4.783	4.371	4.124	3.960	3.844	3.757	3.690
2 :	3.927	3.635	3.461	3.344	3.261	3.199	3.151

Tafel 2.2. (Niveau Alpha = 0.05)

Freiheitsgrad:	12	14	15	16	18	20	21
Zerlegung :							
10 :	5.395	5.254	5.198	5.150	5.071	5.008	4.981
9 :	5.265	5.131	5.077	5.031	4.956	4.896	4.870
8 2 :	5.159	5.028	4.977	4.932	4.858	4.800	4.776
8 :	5.119	4.990	4.940	4.897	4.824	4.768	4.743
7 3 :	5.080	4.951	4.901	4.857	4.785	4.728	4.704
6 4 :	5.031	4.904	4.853	4.810	4.739	4.682	4.658
7 2 :	5.001	4.877	4.828	4.786	4.717	4.662	4.639
5 5 :	5.014	4.887	4.837	4.794	4.723	4.667	4.643
7 :	4.950	4.829	4.782	4.741	4.673	4.620	4.597
6 3 :	4.919	4.797	4.749	4.708	4.640	4.586	4.564
6 2 2 :	4.880	4.760	4.713	4.672	4.605	4.552	4.530
5 4 :	4.876	4.756	4.708	4.667	4.600	4.547	4.524
6 2 :	4.819	4.703	4.657	4.618	4.553	4.502	4.480
5 3 2 :	4.804	4.686	4.639	4.599	4.534	4.482	4.460
6 :	4.751	4.639	4.595	4.557	4.495	4.445	4.424
4 4 2 :	4.777	4.660	4.614	4.574	4.509	4.457	4.435
5 3 :	4.735	4.621	4.577	4.538	4.475	4.425	4.403
4 3 3 :	4.743	4.627	4.578	4.542	4.477	4.426	4.404
4 4 :	4.707	4.594	4.549	4.511	4.448	4.398	4.377
5 2 2 :	4.686	4.574	4.530	4.492	4.430	4.381	4.360
5 2 :	4.603	4.496	4.454	4.418	4.358	4.311	4.291
4 3 2 :	4.615	4.505	4.462	4.425	4.364	4.315	4.295
3 3 3 :	4.573	4.464	4.422	4.385	4.324	4.276	4.256
4 2 2 2 :	4.557	4.450	4.408	4.371	4.311	4.263	4.243
5 :	4.508	4.407	4.367	4.333	4.277	4.232	4.213
4 3 :	4.524	4.419	4.378	4.343	4.284	4.238	4.219
3 3 2 2 :	4.513	4.406	4.364	4.328	4.269	4.222	4.202
4 2 2 :	4.458	4.356	4.316	4.281	4.224	4.179	4.160
3 3 2 :	4.407	4.306	4.267	4.232	4.176	4.132	4.113
4 2 :	4.341	4.246	4.208	4.175	4.122	4.080	4.062
3 2 2 2 :	4.331	4.233	4.194	4.161	4.106	4.063	4.044
3 3 :	4.282	4.187	4.150	4.118	4.066	4.024	4.007
2 2 2 2 2 :	4.248	4.152	4.115	4.082	4.028	3.986	3.968
4 :	4.199	4.111	4.076	4.046	3.997	3.958	3.942
3 2 2 :	4.191	4.100	4.064	4.033	3.982	3.942	3.925
2 2 2 2 :	4.090	4.002	3.967	3.937	3.888	3.849	3.832
3 2 :	4.014	3.931	3.899	3.871	3.825	3.788	3.773
2 2 2 :	3.885	3.806	3.775	3.748	3.703	3.669	3.654
3 :	3.773	3.702	3.674	3.649	3.609	3.578	3.565
2 2 :	3.592	3.525	3.499	3.476	3.439	3.410	3.397
2 :	3.082	3.033	3.014	2.998	2.971	2.950	2.941

Tafel 2.3. (Niveau Alpha = 0.05)

Freiheitsgrad:	24	25	27	28	30	32	35
Zerlegung :							
10 :	4.915	4.897	4.864	4.850	4.824	4.802	4.773
9 :	4.807	4.789	4.758	4.745	4.720	4.698	4.671
8 2 :	4.715	4.698	4.668	4.654	4.630	4.610	4.583
8 :	4.684	4.667	4.638	4.625	4.602	4.581	4.555
7 3 :	4.644	4.627	4.598	4.585	4.561	4.541	4.515
6 4 :	4.599	4.583	4.554	4.541	4.517	4.497	4.471
7 2 :	4.581	4.565	4.537	4.524	4.502	4.482	4.457
5 5 :	4.584	4.568	4.539	4.526	4.502	4.482	4.457
7 :	4.541	4.526	4.498	4.486	4.464	4.445	4.421
6 3 :	4.507	4.491	4.464	4.451	4.429	4.410	4.385
6 2 2 :	4.474	4.459	4.432	4.419	4.398	4.379	4.355
5 4 :	4.468	4.453	4.425	4.413	4.391	4.372	4.348
6 2 :	4.426	4.411	4.385	4.373	4.352	4.334	4.310
5 3 2 :	4.405	4.390	4.363	4.351	4.330	4.311	4.287
6 :	4.373	4.358	4.333	4.322	4.302	4.284	4.261
4 4 2 :	4.381	4.366	4.339	4.327	4.306	4.288	4.264
5 3 :	4.351	4.336	4.310	4.299	4.278	4.260	4.237
4 3 3 :	4.350	4.335	4.309	4.297	4.276	4.258	4.234
4 4 :	4.324	4.310	4.284	4.273	4.252	4.235	4.212
5 2 2 :	4.308	4.294	4.268	4.257	4.237	4.219	4.197
5 2 :	4.242	4.228	4.204	4.193	4.174	4.157	4.136
4 3 2 :	4.244	4.230	4.205	4.194	4.174	4.157	4.135
3 3 3 :	4.206	4.192	4.167	4.156	4.137	4.120	4.098
4 2 2 2 :	4.194	4.180	4.155	4.145	4.125	4.108	4.087
5 :	4.166	4.153	4.130	4.120	4.102	4.086	4.066
4 3 :	4.170	4.157	4.133	4.123	4.104	4.087	4.066
3 3 2 2 :	4.153	4.139	4.115	4.104	4.085	4.068	4.047
4 2 2 :	4.113	4.100	4.077	4.066	4.048	4.032	4.011
3 3 2 :	4.066	4.053	4.030	4.020	4.002	3.986	3.966
4 2 :	4.018	4.005	3.984	3.974	3.957	3.942	3.923
3 2 2 2 :	3.999	3.986	3.964	3.954	3.936	3.921	3.901
3 3 :	3.963	3.951	3.930	3.920	3.903	3.888	3.870
2 2 2 2 2 :	3.924	3.912	3.890	3.881	3.863	3.848	3.829
4 :	3.901	3.890	3.870	3.861	3.845	3.832	3.814
3 2 2 :	3.882	3.871	3.850	3.841	3.824	3.810	3.792
2 2 2 2 :	3.791	3.780	3.760	3.751	3.735	3.721	3.704
3 2 :	3.735	3.724	3.705	3.697	3.682	3.669	3.653
2 2 2 :	3.617	3.607	3.589	3.581	3.567	3.555	3.539
3 :	3.532	3.523	3.506	3.499	3.486	3.475	3.461
2 2 :	3.366	3.358	3.343	3.336	3.324	3.313	3.300
2 :	2.919	2.913	2.902	2.897	2.888	2.881	2.871



Tafel 2.4. (Niveau Alpha = 0.05)

Freiheitsgrad:	36	40	42	45	48	49	50
Zerlegung :							
10 :	4.764	4.735	4.722	4.705	4.690	4.686	4.681
9 :	4.663	4.635	4.622	4.606	4.592	4.588	4.584
8 2 :	4.575	4.548	4.536	4.520	4.507	4.503	4.499
8 :	4.547	4.521	4.509	4.494	4.481	4.477	4.473
7 3 :	4.507	4.480	4.469	4.453	4.440	4.436	4.432
6 4 :	4.464	4.437	4.426	4.411	4.398	4.394	4.390
7 2 :	4.450	4.424	4.413	4.398	4.385	4.381	4.378
5 5 :	4.449	4.422	4.411	4.396	4.383	4.379	4.375
7 :	4.414	4.389	4.378	4.364	4.351	4.347	4.344
6 3 :	4.378	4.353	4.342	4.328	4.315	4.311	4.308
6 2 2 :	4.347	4.323	4.312	4.298	4.285	4.282	4.278
5 4 :	4.340	4.315	4.305	4.291	4.278	4.274	4.271
6 2 :	4.303	4.279	4.269	4.255	4.244	4.240	4.236
5 3 2 :	4.280	4.256	4.246	4.232	4.220	4.216	4.212
6 :	4.255	4.232	4.222	4.209	4.197	4.194	4.190
4 4 2 :	4.257	4.233	4.222	4.209	4.197	4.193	4.189
5 3 :	4.231	4.207	4.197	4.184	4.172	4.168	4.165
4 3 3 :	4.228	4.203	4.193	4.180	4.168	4.164	4.161
4 4 :	4.205	4.182	4.172	4.158	4.147	4.143	4.140
5 2 2 :	4.190	4.167	4.157	4.144	4.133	4.129	4.126
5 2 :	4.130	4.108	4.098	4.086	4.075	4.071	4.068
4 3 2 :	4.128	4.106	4.096	4.083	4.072	4.068	4.065
3 3 3 :	4.091	4.069	4.059	4.047	4.036	4.032	4.029
4 2 2 2 :	4.080	4.058	4.049	4.036	4.025	4.022	4.018
5 :	4.060	4.039	4.030	4.018	4.008	4.005	4.002
4 3 :	4.060	4.039	4.029	4.017	4.006	4.003	4.000
3 3 2 2 :	4.041	4.019	4.009	3.997	3.986	3.983	3.979
4 2 2 :	4.005	3.984	3.975	3.963	3.953	3.950	3.947
3 3 2 :	3.960	3.939	3.930	3.919	3.908	3.905	3.902
4 2 :	3.917	3.897	3.889	3.878	3.868	3.865	3.862
3 2 2 2 :	3.895	3.875	3.867	3.855	3.845	3.842	3.839
3 3 :	3.864	3.845	3.836	3.825	3.816	3.813	3.810
2 2 2 2 2 :	3.824	3.804	3.795	3.784	3.775	3.772	3.769
4 :	3.809	3.791	3.783	3.773	3.764	3.761	3.758
3 2 2 :	3.786	3.768	3.760	3.749	3.740	3.737	3.734
2 2 2 2 2 :	3.699	3.680	3.673	3.662	3.653	3.651	3.648
3 2 :	3.648	3.631	3.623	3.614	3.605	3.603	3.600
2 2 2 :	3.534	3.518	3.511	3.502	3.494	3.491	3.489
3 :	3.457	3.442	3.436	3.428	3.420	3.418	3.416
2 2 :	3.296	3.282	3.276	3.269	3.262	3.260	3.258
2 :	2.868	2.858	2.854	2.848	2.843	2.842	2.841

Tafel 2.5. (Niveau Alpha = 0.05)

Freiheitsgrad:	54	56	60	65	70	75	80
Zerlegung :							
10 :	4.666	4.659	4.646	4.633	4.621	4.611	4.603
9 :	4.569	4.562	4.550	4.538	4.527	4.517	4.509
8 2 :	4.484	4.478	4.466	4.454	4.443	4.434	4.426
8 :	4.459	4.452	4.441	4.429	4.419	4.410	4.402
7 3 :	4.418	4.412	4.400	4.388	4.378	4.369	4.361
6 4 :	4.376	4.370	4.358	4.346	4.336	4.327	4.320
7 2 :	4.364	4.358	4.347	4.336	4.326	4.317	4.309
5 5 :	4.361	4.355	4.344	4.332	4.322	4.313	4.305
7 :	4.331	4.325	4.314	4.303	4.293	4.285	4.277
6 3 :	4.294	4.289	4.278	4.267	4.257	4.248	4.241
6 2 2 :	4.265	4.259	4.249	4.238	4.228	4.220	4.212
5 4 :	4.258	4.252	4.241	4.230	4.220	4.212	4.205
6 2 :	4.224	4.218	4.208	4.197	4.188	4.180	4.173
5 3 2 :	4.200	4.194	4.184	4.173	4.163	4.155	4.148
6 :	4.178	4.173	4.163	4.153	4.144	4.136	4.129
4 4 2 :	4.177	4.171	4.161	4.150	4.141	4.132	4.125
5 3 :	4.153	4.147	4.137	4.127	4.118	4.110	4.103
4 3 3 :	4.148	4.142	4.132	4.121	4.112	4.104	4.097
4 4 :	4.128	4.122	4.112	4.102	4.093	4.085	4.078
5 2 2 :	4.114	4.109	4.099	4.088	4.080	4.072	4.065
5 2 :	4.057	4.052	4.042	4.032	4.024	4.017	4.010
4 3 2 :	4.053	4.048	4.038	4.028	4.019	4.012	4.005
3 3 3 :	4.017	4.012	4.003	3.992	3.984	3.976	3.970
4 2 2 2 :	4.007	4.002	3.992	3.982	3.974	3.966	3.960
5 :	3.991	3.986	3.977	3.968	3.960	3.953	3.947
4 3 :	3.989	3.984	3.975	3.965	3.957	3.949	3.943
3 3 2 2 :	3.968	3.963	3.954	3.944	3.935	3.928	3.922
4 2 2 :	3.936	3.931	3.922	3.913	3.904	3.897	3.891
3 3 2 :	3.891	3.887	3.878	3.868	3.860	3.854	3.848
4 2 :	3.852	3.847	3.839	3.830	3.822	3.816	3.810
3 2 2 2 :	3.829	3.824	3.815	3.806	3.798	3.792	3.786
3 3 :	3.800	3.795	3.787	3.778	3.771	3.765	3.759
2 2 2 2 2 :	3.758	3.754	3.746	3.737	3.729	3.723	3.717
4 :	3.749	3.745	3.737	3.729	3.722	3.716	3.711
3 2 2 :	3.724	3.720	3.712	3.704	3.696	3.690	3.685
2 2 2 2 2 :	3.639	3.634	3.627	3.619	3.612	3.606	3.600
3 2 :	3.591	3.587	3.580	3.573	3.566	3.561	3.556
2 2 2 :	3.481	3.477	3.470	3.463	3.456	3.451	3.446
3 :	3.408	3.405	3.399	3.392	3.386	3.382	3.377
2 2 :	3.251	3.247	3.242	3.235	3.230	3.226	3.222
2 :	2.835	2.833	2.829	2.824	2.821	2.817	2.814

Tafel 2.6. (Niveau Alpha = 0.05)

Freiheitsgrad:	90	100	120	150	200	300	∞
Zerlegung :							
10 :	4.588	4.577	4.560	4.542	4.525	4.508	4.474
9 :	4.495	4.484	4.468	4.451	4.435	4.419	4.387
8 2 :	4.413	4.402	4.386	4.371	4.355	4.339	4.308
8 :	4.389	4.379	4.363	4.348	4.332	4.317	4.286
7 3 :	4.348	4.338	4.322	4.307	4.291	4.276	4.245
6 4 :	4.307	4.296	4.281	4.266	4.250	4.235	4.205
7 2 :	4.297	4.287	4.272	4.257	4.242	4.228	4.198
5 5 :	4.292	4.282	4.267	4.252	4.236	4.221	4.191
7 :	4.265	4.256	4.241	4.227	4.212	4.198	4.170
6 3 :	4.229	4.219	4.204	4.190	4.175	4.161	4.132
6 2 2 :	4.200	4.191	4.176	4.162	4.148	4.134	4.106
5 4 :	4.192	4.183	4.168	4.154	4.140	4.125	4.097
6 2 :	4.161	4.152	4.138	4.124	4.111	4.097	4.070
5 3 2 :	4.136	4.127	4.113	4.099	4.085	4.071	4.043
6 :	4.118	4.109	4.096	4.083	4.069	4.056	4.030
4 4 2 :	4.114	4.104	4.090	4.076	4.062	4.049	4.021
5 3 :	4.092	4.083	4.069	4.055	4.042	4.029	4.002
4 3 3 :	4.085	4.076	4.062	4.049	4.035	4.021	3.994
4 4 :	4.067	4.058	4.044	4.031	4.018	4.004	3.978
5 2 2 :	4.054	4.045	4.032	4.019	4.005	3.992	3.966
5 2 :	4.000	3.991	3.978	3.966	3.953	3.941	3.916
4 3 2 :	3.994	3.986	3.973	3.960	3.947	3.934	3.908
3 3 3 :	3.959	3.950	3.937	3.925	3.912	3.899	3.874
4 2 2 2 :	3.949	3.941	3.928	3.915	3.903	3.890	3.865
5 :	3.937	3.929	3.917	3.905	3.893	3.881	3.858
4 3 :	3.933	3.924	3.912	3.900	3.887	3.875	3.851
3 3 2 2 :	3.911	3.903	3.890	3.877	3.865	3.852	3.828
4 2 2 :	3.881	3.873	3.861	3.849	3.837	3.825	3.801
3 3 2 :	3.837	3.830	3.818	3.806	3.794	3.782	3.759
4 2 :	3.801	3.793	3.782	3.770	3.759	3.748	3.726
3 2 2 2 :	3.776	3.768	3.757	3.745	3.734	3.722	3.699
3 3 :	3.750	3.742	3.731	3.720	3.709	3.698	3.676
2 2 2 2 2 :	3.707	3.700	3.689	3.677	3.666	3.655	3.633
4 :	3.702	3.695	3.685	3.674	3.664	3.654	3.633
3 2 2 :	3.676	3.668	3.658	3.647	3.636	3.626	3.604
2 2 2 2 2 :	3.592	3.585	3.574	3.564	3.553	3.543	3.523
3 2 :	3.547	3.541	3.531	3.521	3.512	3.502	3.483
2 2 2 :	3.438	3.432	3.423	3.414	3.404	3.395	3.377
3 :	3.370	3.365	3.356	3.348	3.339	3.331	3.314
2 2 :	3.215	3.210	3.202	3.194	3.186	3.178	3.163
2 :	2.810	2.806	2.800	2.794	2.789	2.783	2.772

EDV in Medizin und Biologie 18 (2-3), 50-54, ISSN 0300-8282  
 © Verlag Eugen Ulmer GmbH & Co., Stuttgart; Gustav Fischer Verlag KG, Stuttgart

## A table of one- and two-tailed fourfold chisquare limits for unconventional alphas

R. Kohnen, R. Lotz and H. Busch

### Zusammenfassung

Über die verfügbaren Tafeln der Schranken für zweiseitige Chi-Quadrat-Tests hinausgehend werden Schranken für einseitige Vierfeldertests vertafelt. Die ein- und zweiseitigen Schranken werden von den konventionellen zu unkonventionellen Signifikanzniveaus ausgedehnt, um sie für simultane Bonferroni-Tests nutzen zu können. Bedarf und Gebrauch der Tafel werden anhand von Beispielen aus der klinischen Forschung illustriert.

### Summary

Beyond existing tables with critical  $\chi^2$ -limits for twotailed chisquare testing of homogeneity or association in fourfold tables, critical values for one-tailed testing are tabulated. The one- and two-sided critical  $\chi^2$ -limits are extended from conventional to unconventional significance levels in order to be used for simultaneous Bonferroni tests. The need for and the use of the table are illustrated by examples from clinical research.

### Introduction

Evaluating Pearsonian fourfold chisquare tests (see e.g., LIENERT, 1986, Chapter 5) is always made by conventional significance tables whose critical limits correspond to two-tailed testing. Two-tailed testing is mostly indicated in comparing two samples of one binary variable (X); however, testing against association between two binary variables (X, Y), is usually one-tailed, in that only positive associations are considered as an alternative ( $H_1$ ) to the nullhypothesis ( $H_0$ ) of independence. Since global homogeneity or contingency tests are increasingly replaced by simultaneous local and regional tests (see HAVRANEK et al., 1986), a table of one-tailed critical chisquare for both conventional and unconventional levels of significance is desirable but not available (see KRÜGER et al., 1981).

In performing  $r$  simultaneous chisquare tests in place of one global test, the conventional significance level has to be adjusted to become an unconventional level (cf. LIENERT et al., 1982)  $\alpha^* = \alpha/r$  as in Bonferroni-testing (see MILLER, 1981). Loss in power by unconventional simultaneous testing is often negligible compared to gain in information which sometimes

justifies the saying that 'Bonferroni wins again' (FUCHS and KENETT, 1980).

### Tabulation and reading of chisquare limits

In Table I, the critical limits of chisquare with 1 degree of freedom were calculated for one-sided  $\chi^2$  (P/1) by setting  $\chi^2 = z^2$  (2P) and for two-sided  $\chi^2$  (P/2) by setting  $\chi^2 = z^2$  (P) with  $z$  as the standard normal deviate. In order to cover unconventional as well as conventional levels of P, the P's were chosen to range from 0.10 down to 0.00000010 with  $\chi^2$ -values in two-digit spacing. The calculation of critical limits made use of Algorithm AS 70 (ODEH & EVANS, 1974).

From Table I we read that the one-sided chisquare limit CHIQ (P/1) for  $P = 0.0500000 = 5\%$  is 2.71 while the two-tailed limit is CHIQ (P/2) = 3.84 as to be read from text book tabulation of chisquare with 1 df (see FISHER & YATES, 1981, Table IV).

For testing  $r=4$  fourfold associations in a  $3 \times 3$ -contingency table with ordered categories simultaneously, we read the unconventional Bonferroni-limit for a conventional significance level  $P = 0.0010000 = 0.1\%$  to be  $\chi^2$  (0.001/4) =  $\chi^2$  (0.0002500) = 12.12 in one-tailed fourfold testing and 13.41 for two-tailed testing.

Testing significance of  $r = 20 \times (20-1)/2 = 190$  Pearsonian Phi-coefficients  $\phi$  in an  $20 \times 20$  interassociation matrix of 10 binary variables at level  $P = 0.01$  against associations, the CHIQ (P/1) limit of  $P = .01/190 = 0.0000526$  reads to be 15.04 approximately. Phi-coefficients exceeding  $\chi^2 = N \times \pi^2 = 15.04$  are considered to be significant at the conventional alpha level of  $P = .01$ , in one-tailed testing against positive interassociations.

Formally, Table I is also a means of finding  $\chi^2$ -values corresponding to one-tailed cumulated probabilities P which may have been arrived at from exact hypergeometric testing of small sample fourfold tables. That way, large and small fourfold tables may be made directly comparable as to heterogeneity or association effects, measured in terms of Phi-square coefficients,  $\phi^2 = \chi^2/N$ , which range from 0 to 1 indicating low to high practical significance, given statistical significance.

Reading of Table I may also be made inversely: A one-tailed fourfold chisquare of 6.17, e.g., is significant at the unconventional level alpha of 0.0065. Analogously, the same

$\chi^2$  is, in two-tailed testing, significant at the level 0.013 only. Reading Table I inversely allows the researcher to find, for given fourfold  $\chi^2$ -values, the corresponding alpha levels down to  $P = 0.0000010$ .

**Applications in biomedical research**

Table I reading of conventional and unconventional fourfold chisquare testing allows for some specified applications in

biomedical and psychosocial research, as are the following ones:

- a. Asymptotic testing for homogeneity of 2 samples, say a treatment versus control sample of patients as to improvement and no improvement from a disease (see LIENERT, 1986, Chap. 5);
- b. Asymptotic testing of association between 2 variables, say main and side effects of a treatment in one sample of patients (see LIENERT, 1986, Chap. 9);
- c. Asymptotic testing for outlying cells in a  $r \times c$ -contingency

One- and two-tailed fourfold ChiQ-limits for 2-Digit Alphas

Alpha	ChiQ (P/1)	ChiQ (P/2)	Alpha	ChiQ (P/1)	ChiQ (P/2)	Alpha	ChiQ (P/1)	ChiQ (P/2)
0.100	1.64	2.71	0.0400	3.06	4.22	0.0070	6.04	7.27
0.099	1.66	2.72	0.0390	3.11	4.26	0.0069	6.06	7.30
0.098	1.67	2.74	0.0380	3.15	4.31	0.0068	6.09	7.33
0.097	1.69	2.75	0.0370	3.19	4.35	0.0067	6.12	7.35
0.096	1.70	2.77	0.0360	3.24	4.40	0.0066	6.14	7.38
0.095	1.72	2.79	0.0350	3.28	4.45	0.0065	6.17	7.41
0.094	1.73	2.80	0.0340	3.33	4.49	0.0064	6.20	7.43
0.093	1.75	2.82	0.0330	3.38	4.55	0.0063	6.22	7.46
0.092	1.77	2.84	0.0320	3.43	4.60	0.0062	6.25	7.49
0.091	1.78	2.86	0.0310	3.48	4.65	0.0061	6.28	7.52
0.090	1.80	2.87	0.0300	3.54	4.71	0.0060	6.31	7.55
0.089	1.81	2.89	0.0290	3.59	4.77	0.0059	6.34	7.58
0.088	1.83	2.91	0.0280	3.65	4.83	0.0058	6.37	7.61
0.087	1.85	2.93	0.0270	3.71	4.89	0.0057	6.40	7.64
0.086	1.87	2.95	0.0260	3.78	4.96	0.0056	6.43	7.67
0.085	1.88	2.97	0.0250	3.84	5.02	0.0055	6.47	7.71
0.084	1.90	2.99	0.0240	3.91	5.09	0.0054	6.50	7.74
0.083	1.92	3.01	0.0230	3.98	5.17	0.0053	6.53	7.77
0.082	1.94	3.02	0.0220	4.06	5.25	0.0052	6.57	7.81
0.081	1.96	3.04	0.0210	4.14	5.33	0.0051	6.60	7.84
0.080	1.97	3.06	0.0200	4.22	5.41	0.0050	6.63	7.88
0.079	1.99	3.09	0.0190	4.31	5.50	0.0049	6.67	7.92
0.078	2.01	3.11	0.0180	4.40	5.60	0.0048	6.71	7.95
0.077	2.03	3.13	0.0170	4.49	5.70	0.0047	6.75	7.99
0.076	2.05	3.15	0.0160	4.60	5.80	0.0046	6.78	8.03
0.075	2.07	3.17	0.0150	4.71	5.92	0.0045	6.82	8.07
0.074	2.09	3.19	0.0140	4.83	6.04	0.0044	6.86	8.11
0.073	2.11	3.21	0.0130	4.96	6.17	0.0043	6.90	8.15
0.072	2.13	3.24	0.0120	5.09	6.31	0.0042	6.95	8.20
0.071	2.16	3.26	0.0110	5.25	6.47	0.0041	6.99	8.24
0.070	2.18	3.28	0.0100	5.41	6.63	0.0040	7.03	8.28
0.069	2.20	3.31	0.0099	5.43	6.65	0.0039	7.08	8.33
0.068	2.22	3.33	0.0098	5.45	6.67	0.0038	7.13	8.38
0.067	2.25	3.36	0.0097	5.47	6.69	0.0037	7.17	8.43
0.066	2.27	3.38	0.0096	5.48	6.71	0.0036	7.22	8.48
0.065	2.29	3.40	0.0095	5.50	6.73	0.0035	7.27	8.53
0.064	2.32	3.43	0.0094	5.52	6.75	0.0034	7.33	8.58
0.063	2.34	3.46	0.0093	5.54	6.76	0.0033	7.38	8.63
0.062	2.37	3.48	0.0092	5.56	6.78	0.0032	7.43	8.69
0.061	2.39	3.51	0.0091	5.58	6.80	0.0031	7.49	8.75
0.060	2.42	3.54	0.0090	5.60	6.82	0.0030	7.55	8.81
0.059	2.44	3.57	0.0089	5.62	6.84	0.0029	7.61	8.87
0.058	2.47	3.59	0.0088	5.64	6.86	0.0028	7.67	8.93
0.057	2.50	3.62	0.0087	5.66	6.88	0.0027	7.74	9.00
0.056	2.53	3.65	0.0086	5.68	6.90	0.0026	7.81	9.07
0.055	2.55	3.68	0.0085	5.70	6.92	0.0025	7.88	9.14
0.054	2.58	3.71	0.0084	5.72	6.95	0.0024	7.95	9.22
0.053	2.61	3.74	0.0083	5.74	6.97	0.0023	8.03	9.29
0.052	2.64	3.78	0.0082	5.76	6.99	0.0022	8.11	9.37
0.051	2.67	3.81	0.0081	5.78	7.01	0.0021	8.20	9.46
0.050	2.71	3.84	0.0080	5.80	7.03	0.0020	8.28	9.55
0.049	2.74	3.88	0.0079	5.82	7.06	0.0019	8.38	9.64
0.048	2.77	3.91	0.0078	5.85	7.08	0.0018	8.48	9.74
0.047	2.80	3.95	0.0077	5.87	7.10	0.0017	8.58	9.85
0.046	2.84	3.98	0.0076	5.89	7.13	0.0016	8.69	9.96
0.045	2.87	4.02	0.0075	5.92	7.15	0.0015	8.81	10.08
0.044	2.91	4.06	0.0074	5.94	7.17	0.0014	8.93	10.21
0.043	2.95	4.10	0.0073	5.96	7.20	0.0013	9.07	10.34
0.042	2.99	4.14	0.0072	5.99	7.22	0.0012	9.22	10.49
0.041	3.02	4.18	0.0071	6.01	7.25	0.0011	9.37	10.65

- table, say for predicting at least one of  $c$  outcomes of treatment from at least one of  $r$  treatments, as in predictive configural frequency analysis (cf., HEILMANN et al., 1979);
- d. Asymptotic testing for outlying cells in a multidimensional contingency table, say for identifying symptoms forming a syndrome, as in one-sample configural frequency analysis (KRAUTH & LIENERT, 1973, Chap. 1, 2; LEHMACHER, 1981);
- e. Asymptotic testing for partial association in fourfold tables arrived at from collapsing  $r \times c$  - contingency tables, say association between main and side effects within treatment responders neglecting non-responders, as in association configural frequency analysis (LIENERT & NETTER, 1984; BAUER et al., 1986);
- f. Asymptotic testing against interassociations of  $R = r \times (r-1)/2$  pairs of  $r$  binary variables using the unconventional level  $\alpha^* = \alpha/R$  for each one of the  $R$  simultaneous fourfold tests. This procedure is equivalent to testing against significance of 2-way marginals in  $r$ -way contingency cubes (see LIENERT, 1978, Chap. 17);
- g. Comparing  $r-1$  treatments with 1 control within a  $r \times 2$  - YES-NO improvement table pairwise using the unconven-

Alpha	ChiQ (P/1)	ChiQ (P/2)	Alpha	ChiQ (P/1)	ChiQ (P/2)	Alpha	ChiQ (P/1)	ChiQ (P/2)
0.00100	9.55	10.83	0.000400	11.24	12.53	0.000070	14.50	15.81
0.00099	9.57	10.85	0.000390	11.29	12.58	0.000069	14.53	15.84
0.00098	9.59	10.86	0.000380	11.34	12.63	0.000068	14.56	15.87
0.00097	9.61	10.88	0.000370	11.39	12.68	0.000067	14.58	15.89
0.00096	9.62	10.90	0.000360	11.44	12.73	0.000066	14.61	15.92
0.00095	9.64	10.92	0.000350	11.49	12.78	0.000065	14.64	15.95
0.00094	9.66	10.94	0.000340	11.54	12.84	0.000064	14.67	15.98
0.00093	9.68	10.96	0.000330	11.60	12.89	0.000063	14.70	16.01
0.00092	9.70	10.98	0.000320	11.66	12.95	0.000062	14.73	16.04
0.00091	9.72	11.00	0.000310	11.71	13.01	0.000061	14.76	16.07
0.00090	9.74	11.02	0.000300	11.78	13.07	0.000060	14.79	16.10
0.00089	9.76	11.04	0.000290	11.84	13.13	0.000059	14.82	16.13
0.00088	9.78	11.06	0.000280	11.90	13.20	0.000058	14.86	16.17
0.00087	9.81	11.09	0.000270	11.97	13.27	0.000057	14.89	16.20
0.00086	9.83	11.11	0.000260	12.04	13.34	0.000056	14.92	16.23
0.00085	9.85	11.13	0.000250	12.12	13.41	0.000055	14.96	16.27
0.00084	9.87	11.15	0.000240	12.19	13.49	0.000054	14.99	16.30
0.00083	9.89	11.17	0.000230	12.27	13.57	0.000053	15.03	16.34
0.00082	9.91	11.20	0.000220	12.35	13.65	0.000052	15.06	16.37
0.00081	9.94	11.22	0.000210	12.44	13.74	0.000051	15.10	16.41
0.00080	9.96	11.24	0.000200	12.53	13.83	0.000050	15.14	16.45
0.00079	9.98	11.26	0.000190	12.63	13.93	0.000049	15.17	16.49
0.00078	10.01	11.29	0.000180	12.73	14.03	0.000048	15.21	16.53
0.00077	10.03	11.31	0.000170	12.84	14.14	0.000047	15.25	16.57
0.00076	10.05	11.34	0.000160	12.95	14.25	0.000046	15.29	16.61
0.00075	10.08	11.36	0.000150	13.07	14.37	0.000045	15.34	16.65
0.00074	10.10	11.39	0.000140	13.20	14.50	0.000044	15.38	16.69
0.00073	10.13	11.41	0.000130	13.34	14.64	0.000043	15.42	16.73
0.00072	10.15	11.44	0.000120	13.49	14.79	0.000042	15.47	16.78
0.00071	10.18	11.46	0.000110	13.65	14.96	0.000041	15.51	16.82
0.00070	10.21	11.49	0.000100	13.83	15.14	0.000040	15.56	16.87
0.00069	10.23	11.52	0.000099	13.85	15.16	0.000039	15.61	16.92
0.00068	10.26	11.54	0.000098	13.87	15.17	0.000038	15.66	16.97
0.00067	10.29	11.57	0.000097	13.89	15.19	0.000037	15.71	17.02
0.00066	10.31	11.60	0.000096	13.91	15.21	0.000036	15.76	17.07
0.00065	10.34	11.63	0.000095	13.93	15.23	0.000035	15.81	17.12
0.00064	10.37	11.66	0.000094	13.95	15.25	0.000034	15.87	17.18
0.00063	10.40	11.69	0.000093	13.97	15.27	0.000033	15.92	17.24
0.00062	10.43	11.71	0.000092	13.99	15.29	0.000032	15.98	17.30
0.00061	10.46	11.75	0.000091	14.01	15.31	0.000031	16.04	17.36
0.00060	10.49	11.78	0.000090	14.03	15.34	0.000030	16.10	17.42
0.00059	10.52	11.81	0.000089	14.05	15.36	0.000029	16.17	17.48
0.00058	10.55	11.84	0.000088	14.07	15.38	0.000028	16.23	17.55
0.00057	10.59	11.87	0.000087	14.09	15.40	0.000027	16.30	17.62
0.00056	10.62	11.90	0.000086	14.11	15.42	0.000026	16.37	17.69
0.00055	10.65	11.94	0.000085	14.14	15.44	0.000025	16.45	17.76
0.00054	10.69	11.97	0.000084	14.16	15.47	0.000024	16.53	17.84
0.00053	10.72	12.01	0.000083	14.18	15.49	0.000023	16.61	17.92
0.00052	10.75	12.04	0.000082	14.20	15.51	0.000022	16.69	18.01
0.00051	10.79	12.08	0.000081	14.23	15.53	0.000021	16.78	18.10
0.00050	10.83	12.12	0.000080	14.25	15.56	0.000020	16.87	18.19
0.00049	10.86	12.15	0.000079	14.27	15.58	0.000019	16.97	18.29
0.00048	10.90	12.19	0.000078	14.30	15.61	0.000018	17.07	18.39
0.00047	10.94	12.23	0.000077	14.32	15.63	0.000017	17.18	18.50
0.00046	10.98	12.27	0.000076	14.35	15.66	0.000016	17.30	18.61
0.00045	11.02	12.31	0.000075	14.37	15.68	0.000015	17.42	18.74
0.00044	11.06	12.35	0.000074	14.40	15.71	0.000014	17.55	18.87
0.00043	11.11	12.40	0.000073	14.42	15.73	0.000013	17.69	19.01
0.00042	11.15	12.44	0.000072	14.45	15.76	0.000012	17.84	19.16
0.00041	11.20	12.49	0.000071	14.48	15.78	0.000011	18.01	19.33

- tional level  $\alpha \times = \alpha/(r-1)$  for each of the  $r-1$  simultaneous comparisons;
- h. Comparing  $r$  treatments (including control) pairwise in a  $r \times 2$  - YES-NO improvement table using the adjusted level  $\alpha^* = \alpha/R$  with  $R = r \times (r-1)/2$  for each of the  $R$  simultaneous comparisons;
  - i. Decomposing a  $r \times 2$  - YES-NO improvement table with  $r$  ordered treatments (increasing dosages of one drug or  $r$  drugs of increasing potency) into additive  $\chi^2$ -components; these are based on fourfold tables from collapsing the  $r \times 2$ -table suitably (see KIMBALL, 1954) using  $\alpha^* = \alpha/(r-1)$  for each fourfold  $\chi^2$ -component;
  - j. Comparing YES-NO-improvement from a pretreatment observation to a posttreatment observation in two samples of patients (treatment, control; cf. BUCK, 1975); the resulting fourfold table is to be evaluated using a conventional alpha level;
  - k. Detecting heteropoetic treatment effects in a prepost observation design from association of improvements (+) and disimprovements (-) in two outcome variables as suggested by KOHNEN & LIENERT (1979);
  - l. Detecting multivariate treatment effects by two sample

Alpha	ChiQ (P/1)	ChiQ (P/2)	Alpha	ChiQ (P/1)	ChiQ (P/2)	Alpha	ChiQ (P/1)	ChiQ (P/2)
0.0000100	18.19	19.51	0.0000400	19.94	21.26	0.0000070	23.28	24.62
0.0000099	18.21	19.53	0.0000390	19.99	21.31	0.0000069	23.31	24.64
0.0000098	18.23	19.55	0.0000380	20.04	21.36	0.0000068	23.34	24.67
0.0000097	18.25	19.57	0.0000370	20.09	21.41	0.0000067	23.36	24.70
0.0000096	18.27	19.59	0.0000360	20.14	21.47	0.0000066	23.39	24.73
0.0000095	18.29	19.61	0.0000350	20.19	21.52	0.0000065	23.42	24.76
0.0000094	18.31	19.63	0.0000340	20.25	21.58	0.0000064	23.45	24.79
0.0000093	18.33	19.65	0.0000330	20.31	21.63	0.0000063	23.48	24.82
0.0000092	18.35	19.67	0.0000320	20.36	21.69	0.0000062	23.51	24.85
0.0000091	18.37	19.69	0.0000310	20.43	21.75	0.0000061	23.55	24.88
0.0000090	18.39	19.71	0.0000300	20.49	21.82	0.0000060	23.58	24.91
0.0000089	18.41	19.73	0.0000290	20.55	21.88	0.0000059	23.61	24.94
0.0000088	18.43	19.76	0.0000280	20.62	21.95	0.0000058	23.64	24.98
0.0000087	18.45	19.78	0.0000270	20.69	22.02	0.0000057	23.68	25.01
0.0000086	18.48	19.80	0.0000260	20.76	22.09	0.0000056	23.71	25.05
0.0000085	18.50	19.82	0.0000250	20.84	22.17	0.0000055	23.74	25.08
0.0000084	18.52	19.84	0.0000240	20.92	22.24	0.0000054	23.78	25.12
0.0000083	18.54	19.87	0.0000230	21.00	22.33	0.0000053	23.82	25.15
0.0000082	18.57	19.89	0.0000220	21.08	22.41	0.0000052	23.85	25.19
0.0000081	18.59	19.91	0.0000210	21.17	22.50	0.0000051	23.89	25.23
0.0000080	18.61	19.94	0.0000200	21.26	22.60	0.0000050	23.93	25.26
0.0000079	18.64	19.96	0.0000190	21.36	22.69	0.0000049	23.97	25.30
0.0000078	18.66	19.99	0.0000180	21.47	22.80	0.0000048	24.01	25.34
0.0000077	18.69	20.01	0.0000170	21.58	22.91	0.0000047	24.05	25.38
0.0000076	18.71	20.04	0.0000160	21.69	23.02	0.0000046	24.09	25.42
0.0000075	18.74	20.06	0.0000150	21.82	23.15	0.0000045	24.13	25.47
0.0000074	18.76	20.09	0.0000140	21.95	23.28	0.0000044	24.17	25.51
0.0000073	18.79	20.11	0.0000130	22.09	23.42	0.0000043	24.22	25.55
0.0000072	18.82	20.14	0.0000120	22.24	23.58	0.0000042	24.26	25.60
0.0000071	18.84	20.17	0.0000110	22.41	23.74	0.0000041	24.31	25.65
0.0000070	18.87	20.19	0.0000100	22.60	23.93	0.0000040	24.36	25.69
0.0000069	18.90	20.22	0.0000099	22.61	23.95	0.0000039	24.41	25.74
0.0000068	18.92	20.25	0.0000098	22.63	23.97	0.0000038	24.46	25.79
0.0000067	18.95	20.28	0.0000097	22.65	23.99	0.0000037	24.51	25.84
0.0000066	18.98	20.31	0.0000096	22.67	24.01	0.0000036	24.56	25.90
0.0000065	19.01	20.34	0.0000095	22.69	24.03	0.0000035	24.62	25.95
0.0000064	19.04	20.36	0.0000094	22.71	24.05	0.0000034	24.67	26.01
0.0000063	19.07	20.39	0.0000093	22.73	24.07	0.0000033	24.73	26.07
0.0000062	19.10	20.43	0.0000092	22.76	24.09	0.0000032	24.79	26.13
0.0000061	19.13	20.46	0.0000091	22.78	24.11	0.0000031	24.85	26.19
0.0000060	19.16	20.49	0.0000090	22.80	24.13	0.0000030	24.91	26.25
0.0000059	19.20	20.52	0.0000089	22.82	24.15	0.0000029	24.98	26.32
0.0000058	19.23	20.55	0.0000088	22.84	24.17	0.0000028	25.05	26.38
0.0000057	19.26	20.59	0.0000087	22.86	24.20	0.0000027	25.12	26.45
0.0000056	19.29	20.62	0.0000086	22.88	24.22	0.0000026	25.19	26.53
0.0000055	19.33	20.65	0.0000085	22.91	24.24	0.0000025	25.26	26.60
0.0000054	19.36	20.69	0.0000084	22.93	24.26	0.0000024	25.34	26.68
0.0000053	19.40	20.73	0.0000083	22.95	24.29	0.0000023	25.42	26.76
0.0000052	19.44	20.76	0.0000082	22.98	24.31	0.0000022	25.51	26.85
0.0000051	19.47	20.80	0.0000081	23.00	24.33	0.0000021	25.60	26.94
0.0000050	19.51	20.84	0.0000080	23.02	24.36	0.0000020	25.69	27.03
0.0000049	19.55	20.88	0.0000079	23.05	24.38	0.0000019	25.79	27.13
0.0000048	19.59	20.92	0.0000078	23.07	24.41	0.0000018	25.90	27.24
0.0000047	19.63	20.96	0.0000077	23.10	24.43	0.0000017	26.01	27.35
0.0000046	19.67	21.00	0.0000076	23.12	24.46	0.0000016	26.13	27.46
0.0000045	19.71	21.04	0.0000075	23.15	24.48	0.0000015	26.25	27.59
0.0000044	19.76	21.08	0.0000074	23.17	24.51	0.0000014	26.38	27.72
0.0000043	19.80	21.13	0.0000073	23.20	24.53	0.0000013	26.53	27.87
0.0000042	19.84	21.17	0.0000072	23.23	24.56	0.0000012	26.68	28.02
0.0000041	19.89	21.22	0.0000071	23.25	24.59	0.0000011	26.85	28.19
						0.0000010	27.03	28.37

point-symmetry testing in a fourfold table with 2 complementary outcome patterns (say  $X+Y-Z+$  and  $X-Y+Z-$ ) in rows and 2 independent samples (treatment, control) in columns (see CARL & LIENERT, 1979).

Beyond the above cited fourfold applications, Table I may also be used for one-tailed binomial testing via  $\chi^2 = (f-N \times p)^2 / (N \times p \times (1-p))$  with  $f$   $N \times p$  as the number of patients improved under a new treatment and  $p$  as the proportion improved under a standard treatment within a reference population.

Specifically, one-tailed binomial testing with  $p = 1/2$  is resulting in  $\chi^2 = (a-b)^2 / (a+b)$  with  $a$  as the number of patients improved versus  $b$  as the number of patients disimproved while omitting the number of  $N-a-b$  patients unchanged under some treatment considered to be effective. The suggested  $\chi^2$ -test with 1 df is an asymptotic version of the exact sign test (see LIENERT, 1986, Chap. 5) for testing against directional change in a one- or two-tailed manner.

In bivariate rather than in univariate observations, patterns of bivariate change, like two-outcome improvements ( $++$ ), may be evaluated by bivariate sign testing via  $\chi^2 = (a-d)^2 / (a+d)$  where  $a=f(++)$  and  $d=f(--)$  in a fourfold table with row outcomes  $X=(X+, X-)$  and column outcomes  $Y=(Y+, Y-)$ . Sign tests applied to fourfold tables are discussed in LIENERT (1978, Chap. 19).

## Literature

- BAUER, P., P. HACKL, G. HOMMEL & E. SONNEMANN 1986: Multiple testing of pairs of one-sided hypotheses. *Metrika* **33**, 121-127.
- BUCK, W. (1976): Der Paardifferenzen-U-Test. *Arzneimittelforschung (Drug Res.)* **25**, 825-827.
- CARL, W. & G. A. LIENERT (1979): Eine Zweistichprobenversion von Sarris' bivariatem Vorzeichentest und ihre Anwendung in der Marktforschung. *Psychol. Beiträge* **21**, 401-409.
- FUCHS, C. & R. KENETT (1980): A test for detecting outlying cells in the multinomial distribution and two-way contingency tables. *JASA* **75**, 395-398.
- HAVRANEK, T., R. KOHNEN & G. A. LIENERT (1986): Nonparametric evaluation of ANOVA designs by local, regional, and global contingency testing. *Biom. J.* **28**, 11-21.
- HEILMANN, W.-R., G. A. LIENERT & V. MALY (1979) Prediction models in configural frequency analysis. *Biom. J.* **21**, 79-86.
- KIMBALL, A. W. (1954): Short-cut formulae for the exact partition of  $\chi^2$  in contingency tables. *Biometrics* **10**, 452-458.
- KOHNEN, R. & G. A. LIENERT (1979): Bivariate sign tests sensitive to homo- and heteropoetic treatment effects. *Biom. J.* **21**, 755-761.
- KRAUTH, J. & G. A. LIENERT: Die Konfigurationsfrequenzanalyse und ihre Anwendung in Psychologie und Medizin. Freiburg: Alber, 1973.
- KRÜGER, H. P., W. LEHMACHER & K.-D. WALL: The Fourfold Table up to  $N=80$  - Die Vierfeldertafel bis  $N=80$ . Stuttgart: Fischer, 1981.
- LIENERT, G. A.: Verteilungsfreie Methoden in der Biostatistik. Band 2. Meisenheim: Hain, 1978.
- LIENERT, G. A., O. LUDWIG & K. ROCKENFELLER (1982): Tables of the critical values for simultaneous and sequential Bonferroni-z-tests. *Biom. J.* **24**, 239-255.
- LIENERT, G. A. & P. NETTER (1984): Katecholaminreaktionen bei männlichen und weiblichen Hyper- und Normotonikern. *Psychol. Beitr.* **26**, 1-22.
- LEHMACHER, W. (1981): A more powerful simultaneous test procedure in configural frequency analysis. *Biom. J.* **23**, 429-436.
- MILLER, R. G.: Simultaneous statistical inference. New York: Mac-GrawHill, 1981.
- ODEH, R. E. & J. O. EVANS (1974): Algorithm AS 70. *Applied Statistics* **23**, 96-97.

Date of receipt: February 13th, 1987.

The authors' addresses: PD Dr. Ralf Köhnen, Universität Hamburg, Fachbereich Medizin, Abteilung für Medizinische Psychologie, Martinistr. 20, D-2000 Hamburg 20.  
Helmuth Busch, Regina Lotz, Max-Planck-Institut für physiologische und klinische Forschung, W. G. Kerckhoff Institut, Arbeitsgruppe Statistik, Parkstr. 1, D-6350 Bad Nauheim.

## Verteilungsfreie Cluster-Analyse mit statistisch äquivalenten Blöcken

H. Ackermann

### Zusammenfassung

Die vorliegende Arbeit befaßt sich mit der Frage einer mehrdimensionalen nichtparametrischen Cluster-Analyse und beschreibt die Grundlagen für eine Definition von nichtparametrischen »Cluster-Bereichen«, die sich aus Methoden der Konstruktion von skalierungsinvarianten, multivariaten Toleranz- und Diskriminanzbereichen ableiten. Das Verfahren verwendet auf der Grundlage des GOWERSchen Ähnlichkeitskoeffizienten  $S_G$  und einer »binären MAHALANOBIS-Distanz« eine adaptive Modifikation von LINGs  $k$ -Clustering zur sequentiellen Umordnung einer gegebenen Stützstellenmenge.

### Summary

The present paper concerns the problem of multidimensional, non-parametric cluster-analysis and describes the principles to define non-parametric »cluster-regions« derived from methods to construct scale-independent, multivariate tolerance- and discriminant-regions. Based on GOWER's coefficient of similarity and on a »binary MAHALANOBIS-distance«, the method described in this paper uses an adaptive modification of LING's  $k$ -cluster for a sequential reordering of a given set of observations.

### 0. Einleitung

Die Cluster-Analyse befaßt sich mit dem Problem, eine zunächst ungeordnete Menge von Objekten in Klassen möglichst ähnlicher Objekte zu unterteilen. Eine typische Fragestellung findet sich zum Beispiel in der Medizin, wenn eine Gruppe von Patienten auf Grund der Werte mehrerer Laborvariablen in – vielleicht bislang unbekannte – Diagnosegruppen oder Krankheitstypen unterteilt werden soll.

In der Literatur findet sich eine Fülle von unterschiedlichen Cluster-Analyseverfahren; die bekanntesten Verfahren sind sicher das von WARD (1963), der mit dem Ziel der Minimierung von Varianzen schrittweise jeweils ähnlichste Klassen zusammenfaßt, und, als zweite Methode, das Minimaldistanzverfahren nach FORGY (1965) und MCQUEEN (1967), das kleinste Distanzen zu Cluster-Zentren zur Grundlage hat. Durch Lektüre der Lehrbücher von zum Beispiel SPÄTH (1983), GORDON (1981), STEINHAUSEN und LANGER (1977), BOCK (1974) oder SNEATH und SOKAL (1973) kann man versuchen, sich einen Überblick über die Methodenvielfalt zu verschaffen.

Im folgenden wird auf der Basis multivariater Diskriminanzbereiche (ACKERMANN, 1986a/b) eine nichtparametrische, skalierungsinvariante Methode zur Definition von »Cluster-Bereichen« vorgestellt. Das iterative Verfahren bestimmt in jedem Schritt eine Anzahl von Diskriminanzbereichen, die zur Basis einer nichtparametrischen »binären MAHALANOBIS-Distanz« gemacht werden. Aufbauend auf diesem »adaptiven« Maß kann mit Hilfe einer Modifikation der LINGschen Methode der  $k$ -Cluster (LING, 1972) eine adaptive Metrik definiert werden, die eine Umordnung der Stützstellenmenge gestattet. Es ergeben sich dabei – möglichst disjunkte – Bereiche (!) und deren gemeinsames Komplement, denen die Stützstellen, aber auch künftige Objekte zugeordnet werden können. Das Verfahren kann somit in nichtexhaustiver Weise, aber bei erneuter Zuordnung von Objekten im Komplement der Cluster-Bereiche auch in exhaustiver Weise interpretiert werden.

Auf die genannten und einige weitere Charakteristika des vorgestellten Verfahrens wird weiter unten, auch im Vergleich zu anderen Methoden, näher eingegangen. Zur graphischen Veranschaulichung finden sich dort auch einige Beispiele.

### 1. Toleranz- und Diskriminanzbereiche

Betrachtet man einen Wahrscheinlichkeitsraum  $(\Omega, \mathfrak{A}, \mu) = (\mathbb{R}^N, \mathfrak{A}^N, \mu)$ , so wird ein Norm- bzw. Toleranzbereich durch eine Funktion  $S: \Omega^n \rightarrow \mathfrak{A}$  von Zufallsvariablen  $X_i; i=1, 2, \dots, n$  definiert.  $S$  nimmt Werte in der  $\sigma$ -Algebra  $\mathfrak{A}$  an und besitzt eine »Überdeckung«

$$(1.1) \quad U(S) = \mu \{S(X_1, \dots, X_n)\}.$$

Ist  $U(S)$  unabhängig von dem  $W$ -Maß  $\mu$ , so nennt man  $S$  verteilungsfrei oder nicht-parametrisch.

Auf der Basis verschiedener Arbeiten von WILKS (1941, 1942), WALD (1943), TUKEY (1947), FRASER (1951, 1953) und KEMPERMAN (1956) wurden von ABT (1982) und ACKERMANN (1983) Methoden zur Konstruktion von skalierungsinvarianten, nichtparametrischen Toleranzbereichen beschrieben; in der Monographie von ACKERMANN (1985) findet sich eine synoptische Darstellung der Problematik. Die zuletzt genannten Methoden gestatten bei Vorliegen von  $n$  Realisationen  $x_i \in \mathbb{R}^N; i=1, \dots, n$  die Konstruktion eines  $N$ -dimensionalen, nichtparametrischen Bereiches, der einen gewünschten Populationsanteil  $\pi$ , eventuell mit einer Sicherheit  $\gamma$ , überdeckt. Leser/innen, die an weiterführender Literatur interessiert

sind, können auf die Bibliographien von JÍLEK (1981) und JÍLEK und ACKERMANN (1987) verwiesen werden.

Eine Anwendung der skizzierten Toleranzbereiche auf diskriminanzanalytische Fragestellungen findet sich bei ACKERMANN (1986a/b). Dort wird bei Vorliegen von  $M$  Mengen von Beobachtungen  $x_m \in \mathbb{R}^N$ :  $i = 1, 2, \dots, n_m$ ;  $m = 1, 2, \dots, M$  und Vorgabe von Überdeckungen  $\pi_m$ :  $m = 1, \dots, M$  ein Verfahren beschrieben, das den Anteil von Zuordnungen in die Durchschnitte von  $M$  »Toleranzbereichen« bzw. in den Durchschnitt aller Komplemente minimiert. Diese ebenfalls verteilungsfreien und skalierungsinvarianten, multivariaten »Toleranzbereiche« werden deshalb in der zitierten Arbeit als Diskriminanzbereiche bezeichnet.

Charakteristisch für die erwähnten Toleranz- und Diskriminanzbereiche ist die Definition von beschränkten Teilmengen des Raumes  $\Omega = \mathbb{R}^N$ . Dadurch ergeben sich im Falle der Diskriminanzbereiche Zuordnungen in – möglichst disjunkte – Diskriminanzbereiche  $D_m$ :  $m = 1, 2, \dots, M$  und deren Komplement  $D_0 = \bigcup_m D_m$ . Es ist sicher naheliegend, eine solche Situation auch bei clusteranalytischen Fragestellungen herzustellen: In den nächsten Abschnitten wird versucht, bei Vorgabe einer Stützstellenmenge  $x_i$ :  $i = 1, 2, \dots, n$  iterativ gewisse »Cluster-Bereiche«  $C_m$ :  $m = 1, 2, \dots, M$  mit den Überdeckungen  $\pi_m$ :  $m = 1, 2, \dots, M$  zu definieren, die sich – im Sinne der Diskriminanzanalyse – »möglichst gut« unterscheiden, andererseits aber auch – im Sinne der Cluster-Analyse – »möglichst homogen« sind. Die Konstruktionsmethode benutzt ein Charakteristikum der zitierten Diskriminanzbereiche und geht in jedem Iterationsschritt von jeweils gleichen a-priori-Wahrscheinlichkeiten der jeweils aktuellen Cluster-Bereiche aus, was sich in der gegebenen Unkenntnis der »wahren« Verhältnisse auch ohne weiteres anbietet. Aufschluß darüber gibt die Arbeit von ACKERMANN (1986a) und der nachfolgende Abschnitt 2. Eine weitere Interpretation der Cluster-Bereiche  $C_m$ :  $m = 1, 2, \dots, M$  und des gemeinsamen Komplements  $C_0 = \bigcup_m C_m$  wurde andeutungsweise bereits in der Einleitung vorweggenommen und soll später an geeigneter Stelle präzisiert werden.

## 2. Cluster-Bereiche

Die Konstruktionsmethode zur Bestimmung von Cluster-Bereichen kann als ein heuristisches, den bekannten Minimal-Distanz-Verfahren ähnelndes Verfahren bezeichnet werden. Die Bezeichnung »heuristisch« soll hier bedeuten, daß im Verlauf der Konstruktion durch Austausch von Objekten eine »stabile« Gruppierung erzeugt wird, die unter den Austauschbedingungen keinen weiteren Wechsel von Gruppenelementen mehr zuläßt, des weiteren, daß keine explizite Zielfunktion zu definieren ist; der letztgenannte Mangel wird im 3. Abschnitt diskutiert und durch die Angabe eines »Beurteilungskriteriums« einer gegebenen Partition etwas kompensiert. Die »Austauschbedingungen« entsprechen denen der erwähnten Minimal-Distanz-Verfahren, tragen aber im Gegensatz zu den üblichen Verfahren adaptiven Charakter (das heißt, die verwendeten Distanzfunktionen werden einem aktuellen Gruppierungszustand der Objekte »angepaßt«) und sind nichtparametrischer Natur. Eine Skalierungsinvarianz gegen lineare Transformationen der betrachteten Variablen ergibt sich unmittelbar aus der Definition der Konstruktionsmethode.

Es sei nun  $x_i \in \mathbb{R}^N$ :  $i = 1, 2, \dots, n$  eine Menge  $\mathcal{E}$  von Punkten im  $\mathbb{R}^N$ , die in  $M$  möglichst »homogene« Klassen zerlegt werden soll, die ihrerseits  $M$  möglichst disjunkte Cluster-

Bereiche  $C$  definieren sollen. Der Begriff »disjunkt« ergibt sich aus der Verwendung von Diskriminanzbereichen als Teilmengen des  $\mathbb{R}^N$ , »homogen« sollte im üblichen Sinn der Cluster-Analyse verstanden werden und bezieht sich auf die verwendeten Austauschskriterien.

Als Anfangspartition  $\mathcal{E}_1^{(0)}, \dots, \mathcal{E}_M^{(0)}$  der  $n$  Punkte  $x_i \in \mathcal{E}$  wird die übliche initiale Partition verwendet, die den  $i$ -ten Wert der Klasse  $\mathcal{E}_m^{(0)} = \{x_m | j = 1, 2, \dots, n_m\}$  mit  $m = (i-1) \bmod(M) + 1$  zuweist. (Als alternative Methode zur Erzeugung einer Anfangspartition könnte, um dem nichtparametrischen Charakter des Verfahrens gerecht zu werden, auch SPÄTHS  $L_1$ -Clustering (SPÄTH, 1976) Anwendung finden, worauf aber hier nicht weiter eingegangen werden soll.) Zum Abschluß des  $O$ -ten Schrittes werden jetzt nach Vorgabe einer Überdeckung  $\pi = \pi_m$ :  $m = 1, 2, \dots, M$  initiale »Cluster-Bereiche«  $C_m^{(0)}$ :  $m = 1, 2, \dots, M$  erzeugt, die sich durch Anwendung der Diskriminanzbereiche aus ACKERMANN (1986a) auf die Mengen  $\mathcal{E}_m^{(0)}$ :  $m = 1, 2, \dots, M$  ergeben: Diese Mengen  $C_m^{(0)}$  sind per definitionem »möglichst disjunkt« und beruhen auf der  $O$ -ten Annäherung an eine »möglichst homogene« Klassifikation der  $n$  gegebenen Objekte. (Diese Situation erinnert an das Klassifikationsverfahren von DATTOLA (1969), der nichtdisjunkte Gruppierungen von Objektmengen erzeugt, bei der jedes Objekt von seiner zugehörigen Klasse höchstens eine gegebene Distanz  $d$  besitzt. DATTOLA ordnet alle Objekte, die in Schnittmengen liegen, allen Obermengen zu – die als Teilmengen des  $\mathbb{R}^N$  verstanden werden –, was mit den folgenden Zuordnungsregeln des vorgestellten Verfahrens vermieden werden soll. Untypisch entfernte Objekte werden bei DATTOLA als »unklassifizierbar« angesehen, wie dies bei nichtexhaustiver Interpretation der Cluster-Bereiche gleichermaßen der Fall ist: diese Objekte sind – für  $k > 0$  vgl. weiter unten – in der Menge  $C_0^{(k)} = \bigcup_m C_m^{(k)} \subseteq \mathbb{R}^N$  enthalten.)

Zur Definition der Austauschschritte wird eine Bezeichnungsweise von QUESENBERY und GESSAMAN (1968) verwendet: Sei

$$(2.1) \quad d: \Omega \rightarrow \Delta = \{\delta_{i_1 \dots i_\ell} | 1 \leq i_j \leq M\} \cup \{\delta_0\}$$

eine Entscheidungsfunktion mit

$$(2.2) \quad \begin{cases} d(x) = \delta_{i_1 \dots i_\ell} & \text{falls } x \in \bigcap_{j=1}^{\ell} C_{i_j}^{(k)} : 1 \leq i_j \leq M \\ d(x) = \delta_0 & \text{falls } x \in \bigcup_{m=1}^M C_m^{(k)} = C_0^{(k)}. \end{cases}$$

Für  $\ell = 1$  ist  $x$  eindeutig im Bereich  $C_{i_1}^{(k)}$  enthalten, für  $\ell > 1$  in der Schnittmenge von mindestens zwei Bereichen und im Falle von  $d(x) = \delta_0$  im gemeinsamen Komplement  $C_0^{(k)}$  aller  $M$  Cluster-Bereiche enthalten.

Mit Hilfe dieser Entscheidungsfunktion kann eine »nichtparametrische, binäre MAHALANOBIS-Distanz«  $D_M$  definiert werden, die trivialerweise adaptiv ist.  $D_M(x, m) = 0$  bedeutet, daß  $x$  zum Cluster-Bereich  $C_m^{(k)}$  gehört, eventuell aber auch in anderen Bereichen enthalten ist, und  $D_M(x, m) = 1$ , daß  $x$  »entfernt« liegt.

$$(2.3) \quad \begin{cases} D_M(x, m) = 0 & \text{falls } d(x) = \delta_{i_1 \dots i_\ell} \wedge \exists j : i_j = m \\ D_M(x, m) = 1 & \text{sonst.} \end{cases}$$

Als weiteres grundlegendes Distanzmaß wird das Komplement  $D_G$  des GOWERSCHEN Ähnlichkeitskoeffizienten  $S_G$  verwendet,

$$(2.4) \quad D_G(x_i, x_k) = 1 - S_G(x_i, x_k) = \frac{1}{N} \sum_{j=1}^N \frac{|x_{ij} - x_{kj}|}{R_j},$$



wobei  $R_j$  die Spannweite der  $j$ -ten Komponente der Stützstellenmenge  $\mathcal{E}$  bezeichnet; vgl. GOWER (1971) oder zum Beispiel SNEATH und SOKAL (1973).

Auf dieser Basis wird LINGs  $k$ -Clustering modifiziert: Ein Punkt  $x$  möge die Distanz  $D_{L(\lambda)}(x, \mathcal{E})$  zu einer Punktmenge  $\mathcal{E}$  besitzen, wenn für einen Anteil  $\lambda$  aller Werte  $x' \in \mathcal{E}$  gilt, daß  $D_G(x, x') \leq D_{L(\lambda)}(x, \mathcal{E})$ .

$$(2.5) \quad D_{L(\lambda)}(x, \mathcal{E}) = \max(a) : \left\{ \left\{ x' | x' \in \mathcal{E} \wedge D_G(x, x') \leq a \right\} \right\} \leq \lambda \cdot \|\mathcal{E}\|$$

Durch die Verwendung des Anteils  $\lambda$  an Stelle einer festen Anzahl  $k$  von nächstgelegenen Punkten aus  $\mathcal{E}$  kann die Distanz  $D_{L(\lambda)}$  wieder im adaptiven Sinne verstanden werden: Ist ein Cluster-Bereich  $C_m^{(k)}$  bei gleicher geometrischer Ausdehnung »dicht« oder »weniger dicht« mit Stützstellen besetzt, so erhält  $D_{L(\lambda)}$  für ein  $x$  und  $\mathcal{E} = C_m^{(k)}$  unabhängig davon prinzipiell den gleichen Wert.  $D_{L(\lambda)}$  kann somit auch als Maß für die Entfernung eines Punktes  $x$  von einem Cluster-Bereich (oder Teilmengen davon) aufgefaßt werden. Die Definition der Distanz  $D_{L(\lambda)}$  erweist sich offenbar auch als zweckmäßig, wenn nicht nur »kugelförmige«, sondern auch korrelierte, zum Beispiel ellipsoidförmige Cluster betrachtet werden. Beispiele finden sich im letzten Abschnitt.

Die Umordnung der Stützstellen im  $(k+1)$ -ten Iterationsschritt einer Partitionierung läßt sich jetzt leicht definieren. Die Umordnung erfolgt für die »klassifizierbaren« Realisationen  $x: \exists m > 0: D_M(x, m) = 0$  und für die »nichtklassifizierbaren« Punkte  $x: \forall m > 0: D_M(x, m) = 1$  in zwei Stufen:

In der ersten Stufe eines Iterationsschrittes werden die klassifizierbaren Punkte  $x: d(x) = \delta_{i_1, \dots, i_\ell}$  ( $\ell \geq 1 \wedge i_j \neq 0$ ) in den Schnittmengen  $\bigcap_i C_i^{(k)}$  der in Frage stehenden Cluster-Bereiche  $C_i^{(k)}$  betrachtet, und es wird auf Grund der Distanz  $D_{L(\lambda)}$  entschieden, welcher Menge  $\mathcal{E}_m^{(k+1)}: m \in \{i_1, \dots, i_\ell\}$  die Werte  $x$  zugeordnet werden. Bestimmt man die Mengen  $\mathcal{E}_i = \{x' | x' \in \mathcal{E}_i^{(k)} \wedge d(x') = \delta_{i_j}\}$  derjenigen Punkte  $x'$ , die in der Stützstellenmenge  $\mathcal{E}_i^{(k)}$  enthalten sind und auch eindeutig im Cluster-Bereich  $C_i^{(k)}$  liegen, so wird ein in Frage stehender Punkt  $x$  der Menge  $\mathcal{E}_m^{(k+1)}$  zugeordnet, wenn  $x$  zu der Menge  $\mathcal{E}_m$  den kürzesten Abstand  $D_{L(\lambda)}(x, \mathcal{E}_m) = \min_j D_{L(\lambda)}(x, \mathcal{E}_j)$  besitzt. Für Punkte  $x: (D_M(x, m) = 0 \wedge d(x) = \delta_m)$ , die also im Cluster-Bereich  $C_m^{(k)}$ , nicht aber in einer Schnittmenge enthalten sind, bedeutet dies stets eine Zuordnung zu der Menge  $\mathcal{E}_m^{(k+1)}$ . (Im ersten Schritt der Iteration ist es bei kleineren Stichprobenumfängen leicht möglich, daß  $\mathcal{E}_i = \emptyset$ . Darauf wird, wie auf andere eher programmtechnische Schwierigkeiten, nicht weiter eingegangen.)

Die zweite Stufe der Umordnung der Stützstellen betrifft die im  $k$ -ten Schritt nicht klassifizierbaren Punkte  $x: d(x) = \delta_0$  im Komplement  $C_0^{(k)}$  der Cluster-Bereiche  $C_m^{(k)}$ . Ein Punkt  $x: d(x) = \delta_0$  wird dem Cluster  $C_m^{(k)}$  und damit derjenigen im ersten Umordnungsschritt erzeugten Menge  $\mathcal{E}_m^{(k+1)}$  zugeordnet, zu der er den kleinsten Abstand  $D_{L(\lambda)}(x, \mathcal{E}_m^{(k+1)})$  besitzt.

Das Verfahren bricht ab, wenn im  $k$ -ten Zyklus der Partitionierung keine Umordnung der Stützstellen mehr vorgenommen werden kann, das heißt, wenn  $\forall m: \mathcal{E}_m^{(k)} \equiv \mathcal{E}_m^{(k+1)}$ . (Bei fast allen durchgerechneten Beispielen brach das Verfahren nach 4–7 Iterationen ab. Die Partitionierungen, die mehr als 10 Schritte in Anspruch nahmen, reproduzierten in der Regel über 2–3 Zyklen die gleichen Gruppierungen durch Umordnung weniger Punkte. Vgl. dazu auch zum Beispiel den Algorithmus von DATTOLA (1969)). Es sollte noch bemerkt werden, daß das beschriebene Verfahren im Gegensatz zum Beispiel zu der erwähnten MCQUEENSchen Variante des Minimal-

Distanz-Verfahrens nicht von der Reihenfolge der betrachteten Stützstellen abhängt, somit also ordnungsinvariant ist.

Eine Interpretation der Konstruktion und der beiden Umordnungsstufen wird im Rahmen der Diskussion einiger Beispiele im letzten Abschnitt dargestellt. Es muß sicher nicht eigens betont werden, daß die hier beschriebene Konstruktionsmethode wie alle anderen Cluster-Verfahren explorativen Charakter trägt und die Interpretation der Cluster-Bereiche – und insbesondere deren Überdeckungen  $\pi$  – in entsprechender Weise erfolgen muß.

### 3. Bestimmung der Clusteranzahl

Ein grundlegendes Problem der Cluster-Analyse liegt in dem Mangel einer verbindlichen, nichtoperationalen Definition des Begriffes »Cluster«: Daten können Gruppen unterschiedlicher Größe, Dichte und Form enthalten, so daß ein Cluster-Analyse-Algorithmus gelegentlich zu befremdlichen Resultaten führen kann. GORDON (1981) plädiert deshalb für die Verwendung von adaptiven Methoden und hält es für möglich, daß künftig von der »automatischen« Klassifikation abgegangen wird und »überwachte« bzw. interaktive Programme verwendet werden. Entsprechende Argumente sind bei der Frage der immanenten oder kontrollierten Festlegung der »optimalen« Klassenzahl denkbar.

Neben der letztgenannten visuellen, also subjektiv-pragmatischen Kontrolle, die besonders im mehrdimensionalen Fall nur durch unbefriedigende Projektionen auf Ebenen möglich ist, wird in der Literatur eine Bestimmung der »optimalen« Klassenzahl mit Hilfe der Zielfunktion des Algorithmus oder eines anderen Gütekriteriums vorgenommen.

Wie bereits oben diskutiert wurde, kann für das Verfahren aus Abschnitt 2 keine Zielfunktion explizit angegeben werden, so daß hier ein Beurteilungskriterium für Clusterlösungen angegeben wird, das heuristischer Natur ist und, wie bei der praktischen Anwendung beachtet werden sollte, keine mathematische Begründung in Anspruch nimmt.

Nach der Diskussion der Cluster-Bereiche ist es sicher naheliegend, ein Gütekriterium auf der Basis der Distanzen  $D_{L(\lambda)}$  zu konstruieren, da diese das Kriterium für die Umordnung der Stützstellenmenge darstellen; diese Gütefunktion lehnt sich an die SPÄTHsche Zielfunktion der  $L_1$ -Clustering an (vgl. SPÄTH, 1976, 1983). Definiert man nun wie in Abschnitt 2 für alle Cluster-Bereiche  $C_m$  – nach Abschluß der Konstruktion soll der obere Index weggelassen werden – einer gegebenen Partition die Mengen  $\mathcal{E}_m$  derjenigen Punkte  $x$ , die eindeutig in  $C_m$  enthalten sind,

$$(3.1) \quad \mathcal{E}_m = \{x | d(x) = \delta_m, x \in \mathcal{E}_m\},$$

so läßt sich zunächst ein »Kern«  $\mathcal{E}_m^{(0)}$  von  $k = \lambda \cdot \|\mathcal{E}_m\|$  »inneren« Punkten des Cluster-Bereiches  $C_m$  um den multiplen Median von  $\mathcal{E}_m$  als ein Complete-Linkage-Cluster aufbauen. Daran möge sich für die restlichen  $(1-\lambda) \cdot \|\mathcal{E}_m\|$  Punkte eine  $k$ -Clustering im Sinne LINGs anschließen, womit die Distanzen

$$(3.2) \quad a_{mj} = \min_{x \in \mathcal{E}_m^{(j-1)}} (D_{L(\lambda)}(x, \mathcal{E}_m^{(j-1)})) : j = 1, 2, \dots, (1-\lambda) \cdot \|\mathcal{E}_m\|$$

erklärt sind.  $\mathcal{E}_m^{(j)}$  ergibt sich durch Vereinigung des entsprechenden Punktes  $x$  mit der Menge  $\mathcal{E}_m^{(j-1)}$ .

Als Gütekriterium wird somit

$$(3.3) \quad G(M) = \sum_{m,j} a_{mj}$$

verwendet, das, wie empirische Berechnungen zeigen, ein im wesentlichen monotonen Verhalten in Abhängigkeit von der Cluster-Anzahl  $M$  zeigt. In Anlehnung an die Empfehlungen von STEINHAUSEN und LANGER (1977) wird – wie dies auch in den Beispielen des nächsten Abschnittes der Fall ist – die »optimale« Klassenanzahl  $M$  so gewählt, daß die Differenz  $G(M-1) - G(M)$  deutlich größer ist als die Differenz  $G(M) - G(M+1)$ , da in diesem Fall die Partition mit  $M-1$  Clustern vermutlich noch »heterogene« Gruppen enthält und eine geringere zweite Differenz auf eine vermutlich artificielle Weiteraufteilung hindeutet. Relativ gleichbleibende, kleinere Differenzen weisen im allgemeinen auf eine optimale Klassenanzahl von  $M=1$  hin, wie ebenfalls in durchgerechneten Beispielen bestätigt wurde.

**4. Beispiele und Interpretation**

Zur Definition von Cluster-Bereichen sind Werte von  $\lambda$  und  $\pi$  vorzuwählen: Eine Wahl eines kleinen Wertes von  $\lambda$  gestattet das Auffinden von nicht nur kugelförmigen, sondern auch von eventuell langgestreckten Clustern, was bei anderen Clusterverfahren nicht unbedingt der Fall ist; vgl. dazu auch Beispiel 3. Für praktische Anwendungen hat sich eine Wahl von  $\lambda$  mit  $\lambda = 1/3$  bewährt, was auch in allen Beispielen dieses Abschnittes beibehalten wird. Die Überdeckung  $\pi$  sollte so gewählt werden, daß zum Beispiel bei Bereichen mit  $\pi$ -Erwartung, für die also (vgl. GUTTMAN, 1970)

$$(4.1) \quad \pi_i = 1 - \frac{m_i}{n_i+1} \quad ; i = 1, 2, \dots, M$$

gilt, bei  $m_i$  eliminierten Blöcken neben den die initialen Rechteckbereiche definierenden  $2 \cdot N$  Schnitten noch »einige«, die Korrelationen der Daten erfassende Schnitte vorgenommen werden können; das heißt,  $m_i$  sollte mit  $m_i > 2 \cdot N$  gewählt werden. Da bei clusteranalytischen Fragestellungen die Cluster-Umfänge  $n_i : i = 1, 2, \dots, M$  natürlich nicht abzusehen sind, ergibt sich eine angemessene Abschätzung nach

$$(4.2) \quad \pi < 1 - \frac{2 \cdot N}{n/M+1}$$

Für praktische Zwecke hat sich die Abschätzung (4.3) bewährt:

$$(4.3) \quad \pi < 1 - 1.5 \cdot \frac{2 \cdot N}{n/M+1}$$

Leser/innen, die an Einzelheiten der Stichprobenumfangsbestimmung bzw. an der Konstruktion von Toleranzbereichen interessiert sind, können auf die Monographie von ACKERMANN (1985) verwiesen werden.

Die Daten des ersten Beispiels entstammen der Arbeit von HERMANS und HABBEMA (1975). Die Autoren betrachten im Rahmen einer hämatologischen Fragestellung zwei Gruppen von Patientinnen: Eine erste Gruppe bestand aus  $n_1 = 22$  bekannten Überträgerinnen und eine zweite Gruppe aus  $n_2 = 30$  Nichtüberträgerinnen der Hämophilie A. Zu einer diskriminanzanalytischen Trennung der beiden Gruppen wurden die beiden klinischen Parameter  $X_1 =$  Faktor-VIII-Aktivität und  $X_2 =$  Faktor-VIII-assoziiertes Antigen untersucht. Bei ACKERMANN (1986b) wurden diese Daten zur Demonstration von Diskriminanzbereichen verwendet, und es ist sicher reizvoll, diese Ergebnisse mit entsprechenden Cluster-Bereichen zu vergleichen. Die Daten wurden deshalb zu einer Stichprobe zu  $n = 52$  Werten vereinigt und die Fragestellung somit künstlich in eine clusteranalytische umgewandelt.

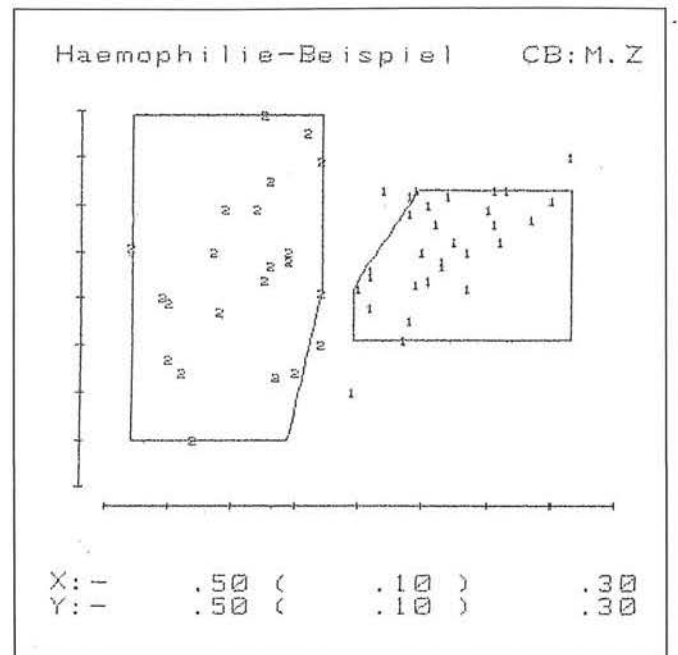
Aus Vergleichbarkeitsgründen wurde eine Überdeckung von  $\pi = \pi_1 = \pi_2 = 0,75$  gewählt und der Anteil  $\lambda$ , wie oben diskutiert, mit  $\lambda = 1/3$  festgesetzt. Das Verfahren aus Abschnitt 2 wurde für  $M = 2, \dots, 5$  durchgeführt; das Beurteilungskriterium aus Abschnitt 3 weist auf die abgebildete Partition mit zwei Clustern hin.

Abbildung 1 zeigt die  $n = 52$  Punkte, als ein Ergebnis des Partitionierungsverfahrens mit den Symbolen »1« und »2« bezeichnet, außerdem zwei Cluster-Bereiche, die per definitionem beide eine Überdeckung  $\pi = 0,75$  besitzen. Die beiden Bereiche  $C_1$  und  $C_2$  sind entsprechend ihrer diskriminanzanalytischen Komponente disjunkt und überdecken jeweils den Anteil  $\pi$  der den mit »1« resp. »2« bezeichneten Werten zugrundeliegenden Populationen. Das Ergebnis der Cluster-Bereiche kann damit zwanglos interpretiert werden:

Die Werte  $x : d(x) = \delta_m : m = 1, 2$  werden jeweils als »Kern« des Clusters zu  $C_m$  aufgefaßt und entsprechend die Werte  $x : d(x) = \delta_0$  im Komplement  $C_0$  als »nicht klassifizierbar« bezeichnet. Eine Zuordnung eines zukünftigen Wertes  $x_0$  kann deshalb ohne weiteres in gleicher Weise erfolgen, wobei die Zuordnung in Abhängigkeit von  $\pi$  mit entsprechenden Fehlklassifikationswahrscheinlichkeiten beurteilt werden kann. Einzelheiten dazu können der Arbeit von ACKERMANN (1986b) entnommen werden; dort findet sich auch eine Abbildung mit der »wahren« Aufteilung der beiden Gruppen unter diskriminanzanalytischen Gesichtspunkten (die im übrigen mit der hier gefundenen sehr gut übereinstimmt).

Die eben formulierte nichtexhaustive Interpretation, die durch die Zuordnungen in den Bereich  $C_0$  zustande kommt, kann natürlich auch in eine exhaustive umgewandelt werden, indem man gemäß (3.1) für  $m = 1, 2$  die Abstände  $D_{L(\lambda)}(x_0, \mathcal{E}_m)$  bestimmt und einen fraglichen Wert  $x_0$  dem nächstgelegenen Clusterbereich zuweist. Dies geht ebenfalls aus Abbildung 1 hervor und kann auch prinzipiell für eine zukünftige Beobachtung  $x_0$  durchgeführt werden.

Abb. 1. Graphische Darstellung zweier Cluster-Bereiche.  $X_1 =$  Faktor-VIII-Aktivität,  $X_2 =$  Faktor-VIII-Assoziiertes Antigen ( $\log(U/ml)$ )



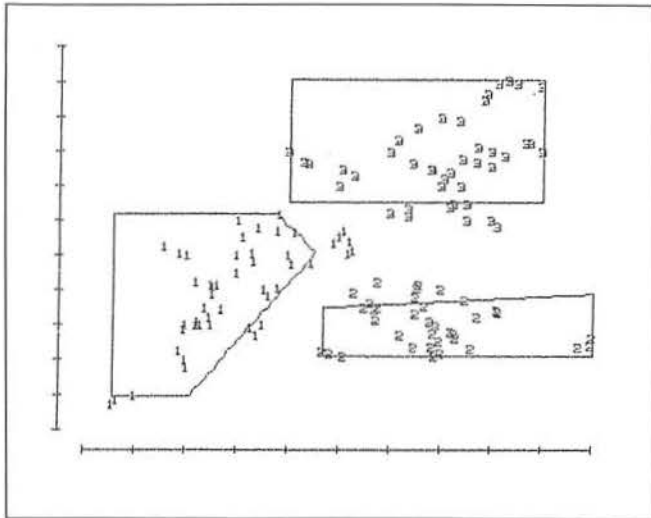


Abb. 2. Cluster-Bereiche mit  $\pi = 0,70$ ,  $\lambda = 1/3$  und  $n = 136$

Abb. 3. Cluster-Bereiche mit  $\pi = 0,70$ ,  $\lambda = 1/3$  und  $n = 80$

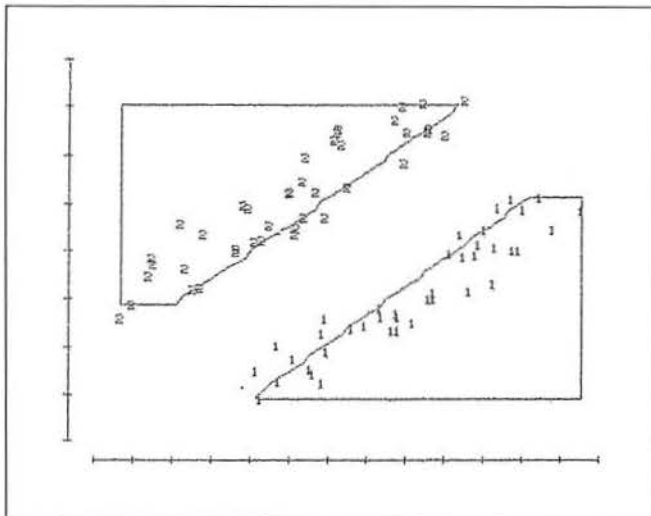
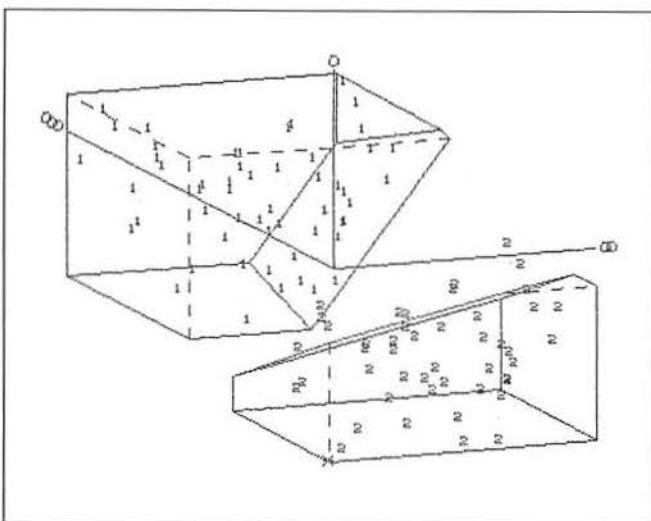


Abb. 4. Zwei dreidimensionale Cluster-Bereiche mit  $\pi = 0,70$ ,  $\lambda = 1/3$  und  $n = 100$



Als nächstes Beispiel zeigt Abbildung 2 eine etwas diffuse, künstlich erzeugte Punkteverteilung von  $n = 136$  Werten, die auf Grund des Beurteilungskriteriums  $G$  in  $M = 3$  Klassen eingeteilt wurden. Eine Interpretation könnte entsprechend den Kommentaren zu Abbildung 1 erfolgen, worauf aber sicher verzichtet werden kann.

Die beiden langgestreckten Cluster in Abbildung 3 demonstrieren – nach Meinung des Autors – den adaptiven Charakter des vorgestellten Cluster-Verfahrens: Wendet man zum Beispiel das »klassische« k-means-Verfahren nach MCQUEEN auf die abgebildete Punktmenge an, so ergeben sich zwei Cluster, von denen das eine die beiden linken, das andere die beiden rechten Hälften der »richtigen« Cluster umfaßt. Vgl. dazu auch GORDON (1981) bzw. die entsprechenden Bemerkungen in Abschnitt 3.

Abbildung 4 zeigt eine dreidimensionale Verteilung von  $n = 100$  Werten und die resultierenden Cluster-Bereiche. Bei ACKERMANN (1986a) finden sich für die gleichen Daten, die dort für diskriminanzanalytische Zwecke verwendet wurden, entsprechende graphische Darstellungen für Diskriminanz- und Toleranzbereiche. Es zeigt sich auch hier, daß die Methode der Cluster-Bereiche die ursprüngliche Situation recht gut reproduziert und nur wenige – allerdings sinnvolle! – Unterschiede zu erkennen sind.

**Literatur**

ABT, K. (1982): Scale-independent non-parametric multivariate tolerance regions and their application in medicine. *Biom. J.* **24**, 27–48.

ACKERMANN, H. (1983): Multivariate non-parametric tolerance-regions: a new construction technique. *Biom. J.* **25**, 351–359.

ACKERMANN, H. (1985): Mehrdimensionale nicht-parametrische Normbereiche. Springer-Verlag, Berlin – Heidelberg – New York – Tokyo.

ACKERMANN, H. (1986a/b): Verteilungsfreie Diskrimination mit statistisch äquivalenten Blöcken. Teil I: Mehrdimensionale skalierungsinvariante Diskriminanzbereiche. Teil II: Praktische Anwendung mehrdimensionaler Diskriminanzbereiche. *EDV in Med. und Biol.* **17**, 112–119, 120–124.

BOCK, H. H. (1974): Automatische Klassifikation. Vandenhoeck und Ruprecht, Göttingen, Zürich.

CZEKANOWSKI, J. (1909): Zur Differentialdiagnose der Neandertalgruppe. *Korrespondenzblatt Deutsch. Ges. Anthropol. Ethnol. Urgesch.* **40**, 44–47.

DATTOLA, R. T. (1969): A fast algorithm for automatic classification. *J. Library Automation* **2**, 31–48.

FRASER, D. A. S. (1951): Sequentially Determined Statistically Equivalent Blocks. *Ann. Math. Stat.* **22**, 372–381.

FRASER, D. A. S. (1953): Nonparametric Tolerance Regions. *Ann. Math. Stat.* **24**, 44–55.

FORGY, E. W. (1965): Cluster Analysis of Multivariate Data: Efficiency versus Interpretability of Classifications. (Abstract): *Biometrics* **21**, 768.

GORDON, A. D. (1981): Classification. Chapman & Hall, London, New York.

GOWER, J. C. (1971): A general coefficient of similarity and some of its properties. *Biometrics* **27**, 857–871.

GUTTMAN, I. (1970): Statistical Tolerance Regions: Classical and Bayesian. Griffin's Stat. Monographs and Courses. No. 26, London.

HERMANS, J., J. D. F. HABBEMA (1975): Comparison of five methods to estimate posterior probabilities. *EDV in Med. und Biol.* **6**, 14–19.

JILEK, M. (1981): A bibliography of statistical tolerance regions. *Math. Operationsforsch. und Statistik. Ser. Stat.*, **12**, 441–456.

- JILEK, M., H. ACKERMANN (1987): A bibliography of statistical tolerance regions II. *Math Operationsforsch. und Statistik, Ser. Stat.*, im Druck.
- KEMPERMAN, J. H. B. (1956): Generalized tolerance regions. *Ann. Math. Stat.* **27**, 180-186.
- LING, R. F. (1972): On the theory and construction of k-clusters. *Comp. J.* **15**, 326-332.
- MCQUEEN, J. (1967): Some methods for classification and analysis of multivariate observations. *Proc. 5th Berkely Symp. Math. Stat. Prob.* 1965/66, Berkely 1967, 21-297.
- QUESENBERRY, C. P., M. P. GESSAMAN (1968): Nonparametric discrimination using tolerance regions. *Ann. Math. Stat.* **39**, 664-673.
- SNEATH, P. H. A., R. R. SOKAL (1973): *Numerical Taxonomy*. W. H. Freeman and Company, San Francisco.
- SPÄTH, H. (1976): Algorithm 30: L1-Cluster-Analysis. *Computing* **16**, 379-387.
- SPÄTH, H. (1977): *Cluster-Analyse-Algorithmen*. R. Oldenbourg Verlag, München, Wien.
- SPÄTH, H. (1983): *Cluster-Formation und -Analyse*. R. Oldenbourg Verlag, München, Wien.
- STEINHAUSEN, D., K. LANGER (1977): *Clusteranalyse*. Walter de Gruyter, Berlin, New York.
- TUKEY, J. W. (1947): Nonparametric Estimation II: Statistically Equivalent Blocks and Tolerance Limits - The Continuous Case. *Ann. Math. Stat.* **18**, 529-539.
- WALD, A. (1943): An Extension of Wilks' Method for Setting Tolerance Limits. *Ann. Math. Stat.* **14**, 45-55.
- WARD, J. H. (1963): Hierarchical grouping to optimize an objective function. *J. Amer. Statist. Assoc.* **58**, 236-244.
- WILKS, S. S. (1941): Determination of Sample Sizes for Setting Tolerance Limits. *Ann. Math. Stat.* **12**, 91-96.
- WILKS, S. S. (1942): Statistical Prediction with Special Reference to the Problem of Tolerance Limits. *Ann. Math. Stat.* **13**, 400-409.

Eingegangen am: 12. Juni 1987.

Anschrift des Verfassers: Dr. Hanns Ackermann, Klinikum der Universität, Abteilung für Biomathematik, Theodor-Stern-Kai 7, D-6000 Frankfurt am Main 70.

EDV in Medizin und Biologie 18 (2-3), 61-64, ISSN 0300-8282  
 © Verlag Eugen Ulmer GmbH & Co., Stuttgart; Gustav Fischer Verlag KG, Stuttgart

# Populationsgenetische Diskrimination zwischen oligo- und polygenem Erbmodus einfacher Schwelleneigenschaften unter Einbeziehung von Geschlechtsdifferenzen

G. Arndt und K. Schmidt

## Zusammenfassung

1. Mit Hilfe zweier Diskriminationsprogramme – GONOSOM2, POLSEX3 – wurde versucht, eindeutige Trennungen zwischen geschlechtsgebundenem oligenen und geschlechtsbeeinflusstem polygenem Erbmodus bei einfachen Schwelleneigenschaften herbeizuführen. Hierbei wurden multidimensionale Such-Algorithmen mit Optimierungsautomatik verwendet.
2. An 33 in Matrizenform generierten Testbeispielen konnten in Modellversuchen mit 3 verschiedenen Sätzen für die Anfangswerte klare Trennungen zwischen den beiden genetischen Kategorien erzielt werden, wenn die Manifestationshäufigkeiten von Verwandten 1. und 2. Grades einbezogen wurden.
3. Vereinzelt auftretende Ungenauigkeiten bei den genetischen Parameterschätzungen und deren Abstellung wurden erörtert.

## Summary

1. By means of two discrimination program – GONOSOM2, POLSEX3 – it has been tried to differentiate between sex-linked unifactorial and sex-influenced multifactorial mode of heredity in cases of ordinary threshold characters. Multidimensional search-algorithm involving minimum function had been employed.
2. When the frequencies of manifestation in relatives of first and second degree in matrix schemes had been applied, discrimination between both categories was successful in 33 test examples generated, whereby three different sets of start values had been used.
3. Sporadic appearing inaccuracies in genetic parameter estimating had been discussed as well as their removal possibilities.

## Problematik

Wenn eine bestimmte Eigenschaft durch nur einen Schwellenwert in 2 Klassen – manifest oder unauffällig – unterteilt werden kann, ist nach JAMES (1971), REICH, JAMES und MORRIS (1972) eine Diskrimination zwischen oligenem und polygenem Erbmodus nicht möglich, wenn Verwandte 1. Grades analysiert werden.

Eigene Versuche, zusätzliche Kriterien – z. B. Geschlechtsunterschiede bei Probanden und deren Verwandten – zur Erzielung unimodaler Optimierung in die Suchalgorithmen einzubeziehen, scheiterten. So wurde es notwendig, uni- und bilineare Verwandte 2. Grades in die Programme aufzunehmen.

## Genetische Modelle

### 1. Gonosomes Modell

In einer Arbeit über stochastische Matrizen publizierten LI und SACKS (1954) Übergangswahrscheinlichkeiten, aus denen hervorgeht, wie groß der Anteil der zu erwartenden Verwandten von bestimmten Probanden-Genotypen (G) in einer Population ist. Es wurden sowohl autosome als auch gonosome Verhältnisse berücksichtigt.

Probanden als auch Verwandte können sowohl homogam (Chromosomen XX, beim Menschen und bei den Säugetieren das weibliche Geschlecht) als auch heterogam (Chromosom XO, männliches Geschlecht) veranlagt sein. XX-Individuen ermöglichen 3 Genotypen (AA, Aa, aa) XO-Individuen nur 2 (A, a), wenn diallele Bedingungen vorliegen.

Für die vier Kombinationen des gonosomalen Probanden-Verwandten-Modells gelten die folgenden Matrizen der ITO-Übergangswahrscheinlichkeiten:

### Probanden (X0)

#### Verwandte (X0)

$$G: \begin{matrix} A \\ a \end{matrix} \quad I_{22} = \begin{bmatrix} 1 & 0 \\ 0 & 1 \end{bmatrix} \quad T_{22} = 0_{22} = \begin{bmatrix} p & q \\ p & q \end{bmatrix}$$

#### Verwandte (XX)

$$G: \begin{matrix} A \\ a \end{matrix} \quad T_{23} = \begin{bmatrix} p & q & 0 \\ 0 & p & q \end{bmatrix}; \quad 0_{23} = \begin{bmatrix} p^2 & 2pq & q^2 \\ p^2 & 2pq & q^2 \end{bmatrix}$$

### Probanden (XX)

#### Verwandte (X0)

$$G: \begin{matrix} AA \\ Aa \\ aa \end{matrix} \quad T_{32} = \begin{bmatrix} 1 & 0 \\ 1/2 & 1/2 \\ 0 & 1 \end{bmatrix}; \quad 0_{32} = \begin{bmatrix} p & q \\ p & q \\ p & q \end{bmatrix}$$

**Verwandte (XX)**

$$G: \begin{matrix} AA \\ Aa \\ aa \end{matrix} I33 = \begin{bmatrix} 1 & 0 & 0 \\ 0 & 1 & 0 \\ 0 & 0 & 1 \end{bmatrix}; T33 = \begin{bmatrix} p & q & 0 \\ p/2 & 1/2 & q/2 \\ 0 & p & q \end{bmatrix}; O33 = \begin{bmatrix} p^2 & 2pq & q^2 \\ p^2 & 2pq & q^2 \\ p^2 & 2pq & q^2 \end{bmatrix}$$

Die I-Komponente enthält den vollständigen Korrelationsanteil, die T-Komponente einen partiellen, die O-Komponente ein unabhängiges Kontingent (Frequenzen einer Zufallspopulation).

Wie aus der Übersicht zu ersehen ist, beläuft sich die Summe jeder einzelnen Matrizenzeile auf 1. Um ein summa summarum der Matrizen zur Einheit zu erreichen, müßte mit den Genotypenfrequenzen der Probanden eine zeilenweise Multiplikation erfolgen, d. h., in der oberen Hälfte mit den Genotypen der Probanden (XO) – A, a – mit p bzw. q, in der unteren Hälfte mit den Genotypen von (XX) – AA, Aa, aa – mit p<sup>2</sup>, 2pq, q<sup>2</sup>. Es gilt q = 1 – p.

Ehe jedoch diese zeilenweise Multiplikation erfolgen kann, soll eine gewogene Addition der ITO-Matrizen durchgeführt werden:

$$c_I I + c_T T + c_O O, \text{ wobei } c_I + c_T + c_O = 1$$

Die c-Skalare ergeben sich aus den Abstammungskoeffizienten. Für unlineare Verwandte 1. Grades ist z. B. c<sub>I</sub> = 0, c<sub>T</sub> = 1, somit c<sub>O</sub> = 0. Die Übergangsmatrizen für Söhne-Väter sind T22, für Töchter-Mütter T33, für Töchter-Väter T32, für Söhne-Mütter T23. Als Beispiel für bilineare Verwandte seien die 4 Vollgeschwister-Kombinationen angeführt:

$$\begin{aligned} \text{♀ ♀: } S33 &= \frac{1}{2} I33 + \frac{1}{2} T33; \text{ ♂ ♂: } S22 = \frac{1}{2} I22 + \frac{1}{2} T22 \\ \text{♀ ♂: } S32 &= \frac{1}{2} T32 + O32; \text{ ♂ ♀: } S23 = \frac{1}{2} T23 + \frac{1}{2} O23 \end{aligned}$$

Tabelle 1. Übergangswahrscheinlichkeiten insgesamt

Probanden	Verwandte			
	X0		XX	
X0	Väter	T22	Mütter	T23
	Brüder	1/2 I22+1/2 O22	Schwestern	1/2 T23+1/2 O23
	GV-M*	1/2 I22+1/2 O22	GM-M	1/2 T23+1/2 O23
	GV-V	O22	GM-V	O23
	Onkel-M	1/4 I22+3/4 O22	Tante-M	3/4 T23+1/4 O23
	Onkel-V	O22	Tante-V	O23
XX	Väter	T32	Mütter	T33
	Brüder	1/2 T32+1/2 O32	Schwestern	1/2 T33+1/2 O33
	GV-M	1/2 T32+1/2 O32	GM-M	1/2 T33+1/2 O33
	GV-V	O32	GM-V	T33
	Onkel-M	1/4 T32+3/4 O32	Tante-M	3/4 T33+1/4 O33
	Onkel-V	1/2 T32+1/2 O32	Tante-V	1/2 T33+1/2 O33

\* GV-M: Großväter mütterlicherseits  
GM-V: Großmütter väterlicherseits

**X-Matrizen**

Um X = [x<sub>ij</sub>] zu erhalten, erfolgt in den ♂ ♂ und ♂ ♀ Kombinationen eine reihenweise Multiplikation mit dem Vektor VQ2 = (p, q), in den ♀ ♂ und ♀ ♀ Kombinationen mit dem Vektor VQ3 = (p<sup>2</sup>, 2pq, q<sup>2</sup>). So resultieren im linken oberen Quadranten X22, oben rechts X23, unten links X32 und unten rechts X33 unterschiedliche Zusammensetzungen.

**Penetranzen**

Unter Penetranz versteht man die (relative) Manifestationshäufigkeit eines Genotyps, welche von der Genotypenfrequenz unabhängig ist. Nach EMERY (1976) und JOHANSSON (1980) muß mit an Sicherheit grenzender Wahrscheinlichkeit angenommen werden, daß die Penetranz für rezessiv Homogame vollständig (f<sub>3</sub> = 1), für Heterozygote darunter liegt (f<sub>2</sub> < 1) und für Dominante noch geringer ist (f<sub>1</sub> < f<sub>2</sub>). Entsprechende Verhältnisse sollten dann auch für die Heterogamen gelten (f<sub>m</sub> < 1). Diese Vektoren erhielten die Bezeichnungen VF3 = (f<sub>1</sub>, f<sub>2</sub>, 1) bzw. VF2 = (f<sub>m</sub>, 1).

**Populationswerte**

Diese sind das innere Produkt von Genotypenfrequenzen und Penetranzen. Für

$$\text{XO-Probanden: } P_m = f_m p + q$$

$$\text{XX-Probanden: } P_w = f_1 p^2 + f_2 2pq + q^2$$

**Verwandtenfrequenzen**

Je nach Geschlechtskombination sind X-Matrizen mit 4, 6 und 9 Feldelementen x<sub>ij</sub> für die Verwandtenkontingente herleitbar. Subskripte von f... sind w<sub>i</sub> = w<sub>j</sub> = 1, 2, 3; m<sub>i</sub> = m<sub>j</sub> = 1, 2.

R<sub>w</sub> steht für gleichgeschlechtliche Verwandte (XX), R<sub>m</sub> ebenso für XO, entsprechend ist R'<sub>w</sub> ein Verwandter eines Probanden (XO), R'<sub>m</sub> ein solcher von Proband (XX). Für die bedingten Wahrscheinlichkeiten – Bedingung ist die Probandenpopulation – der Verwandten (SCHMIDT (1984), ARNDT und SCHMIDT (1984)) gelten folgende Formen:

$$\begin{aligned} \text{Proband (XO): } R_m &= \sum \sum f_{m_i} f_{m_j} x_{ij} \text{ (X22)} / 2P_m \\ R_w' &= \sum \sum f_{m_i} f_{w_j} x_{ij} \text{ (X23)} / 2P_m \end{aligned}$$

$$\begin{aligned} \text{Proband (XX): } R_w &= \sum \sum f_{w_i} f_{w_j} x_{ij} \text{ (X33)} / 2P_w \\ R_m' &= \sum \sum f_{w_i} f_{m_j} x_{ij} \text{ (X32)} / 2P_w \end{aligned}$$

Die Halbierung der Verwandtenhäufigkeiten ergibt sich aus dem vorausgesetzten Geschlechtsverhältnis von 1:1.

Die 12x2 Verwandtenhäufigkeiten werden zu einer Matrix zusammengefaßt. Eine 3. Spalte enthält die zugehörigen Komplemente der nichtbehafteten Verwandten ohne Geschlechtsdifferenzierung.

**2. Polygenes Modell**

Wenn eine normalverteilte Population nur 2 alternative Klassenanteile besitzt, wird die Abszisse durch eine Vertikale getrennt, die im Schwellenwert x<sub>p</sub> liegt (Definition der »Liability«). Nach FALCONER (1967) gelten für die »phänotypische Verwandtschaft« die Beziehungen

$$t = rh^2 = r \frac{V(A)}{V(P_h)} = \frac{SE}{SD} = \frac{x_p - x_R}{i_p}$$

mit i<sub>p</sub> = φ(x)/Φ(x), wobei φ(x) die Dichte, Φ(x) die Verteilungsfunktion von N(0,1) im Punkt x bezeichnen.

r steht für den Verwandtschaftskoeffizienten – 1/2 für Verwandte 1. Grades, 1/4 für solche 2. Grades – h<sup>2</sup> für »Heritabilität«, das Verhältnis der genetisch-additiven zur phänotypischen Varianz, SE für den Selektionserfolg bei den Verwand-

ten, SD für die Selektionsdifferenz zwischen manifesten Probanden und der Gesamtpopulation, so daß

$$SE = M_R - M_P = x_P - x_R,$$

$$SD = M_P - 0 = i_P.$$

Subskript P kennzeichnet die Probanden-Teilpopulation, R («Relatives») deren Verwandtenanteil unter N (0,1).

Erweitert man diese Grundform auf 2 Merkmale - i = 1,2 - so ergibt sich aus der allgemeinen Form die des korrelierten Selektionserfolges KSE (bzw. CR = Correlated response). Die geänderten Formulierungen lauten:

$$t_{12} = rh_1 r_A h_2 = \frac{KSE_{2,1}}{SD_1} = \frac{KSE_{1,2}}{SD_2}$$

Sind Merkmal 1 und 2 metrisch verschiedene Eigenschaften, wird  $r_A$  als genetische Korrelation den Anteil der Pleiotropie repräsentieren, das Komplement  $1-r_A$  den Anteil unterschiedlicher Genwirkung. Unter den geschilderten Gegebenheiten wird  $r_A < 1$  sein. - Wird diese Formulierung nach FALCONER (1967) nur für eine Eigenschaft, jedoch mit verschiedengeschlechtlichen Pfadkoeffizienten h. angewandt, müßte  $r_A = 1$  sein, da für die gleiche Eigenschaft der gleiche pleiotrope Geneffekt gilt ( $\sigma_{A_m} = \sigma_{A_w}$ ). Ist nun  $h_m \neq h_w$  und  $h. = \sigma_A / \sigma_{Ph.}$ , kann der Unterschied nur durch  $\sigma_{Phm} \neq \sigma_{Phw}$  erklärt werden.

Obwohl dieses Prinzip hypothetisch und nicht bewiesen ist (jedoch pragmatisch belegt), wurde es bei den Diskriminationsversuchen zunächst angewandt, zumal der innewohnende Fehler nur geringfügige Resultatsabweichungen erbringt. So gibt es zunächst 3 Formen von t., von denen die eine reziprok definiert werden kann:

$$\bar{t} = (t_m t_w)^{1/2} = t_{w/m} = t_{m/w}$$

Phänotypische Verwandtschaft	Wandlung nach x...
$t_m = \frac{X_{Pm} - X_{Rm}}{i_m}$	$X_{Rm} = X_{Pm} - t_m i_m$
$t_w = \frac{X_{Pw} - X_{Rw}}{i_w}$	$X_{Rw} = X_{Pw} - t_w i_w$
$\bar{t} = \frac{X_{Pm} - X'_{Rm}}{i_w}$	$X_{Rm}' = X_{Pm} - \bar{t} i_w$
$\bar{t} = \frac{X_{Pw} - X'_{Rw}}{i_m}$	$X_{Rw}' = X_{Pw} - \bar{t} i_m$

Die Verwandtenhäufigkeiten 1. und 2. Grades sind dann  $\Phi(X_{R.})/2$  bzw.  $\Phi(X_{R.}')/2$ . Sie werden zu einer  $12 \times 2$ -Matrix

zusammengefaßt. Auch hier enthält eine dritte Säule die Komplemente der normalen, jedoch nicht geschlechtsdifferenzierten Verwandten. Die Begründung für die Halbierung der Verwandtenhäufigkeiten war bereits früher gegeben worden.

8 Gleichungen bilden die Grundlage zur Schätzung des polygenen Modells. Davon sind die ersten vier Gleichungen den Verwandten 1. Grades, die restlichen 4 denen 2. Grades zugeordnet.

### 3. Parameter und Parameterschätzungen

#### 3.1 Parameter des gonosomen Modells

Hierbei werden 4 Parameter geschätzt, nämlich die Genfrequenz q sowie die Penetranzen  $f_m, f_1, f_2$ .

#### 3.2 Parameter des polygenen Modells

Ziel der Schätzung sind die unterschiedlichen phänotypischen Verwandtschaftskoeffizienten  $t_m$  und  $t_w$ , weiterhin die Populationsanteile der Manifestation  $P_m$  und  $P_w$ .

#### 3.3 Optimierungsprozeß

Für beide Modelle lassen sich die Erwartungswerte für die Verwandtenhäufigkeiten - manifest oder unauffällig - in Abhängigkeit von den Modellparametern berechnen und in einer  $12 \times 3$  Matrix entsprechend Tabelle 1 darstellen, wir wollen sie mit E bezeichnen. Die ersten beiden Spalten enthalten die geschlechtsdifferenzierten Anteile der manifesten Verwandten, die dritte Spalte deren nichtdifferenzierte Komplemente. Kennt man die empirischen Häufigkeiten (B) für alle im Modell berücksichtigten Verwandtenhäufigkeiten, so ist es Ziel der Programme, die Modellparameter so zu bestimmen, daß die Summe der quadratischen Abweichungen der empirischen und der erwarteten Häufigkeiten minimal wird. Also

$$\Sigma (E - B)^2 \stackrel{!}{=} \text{Min}$$

#### 3.3.1 Testergebnisse

Um die Programme zu testen, insbesondere um festzustellen, ob die Lösungen eindeutig werden, wurde folgendermaßen vorgegangen. Es wurden für beide Modelle Testbeispiele konstruiert (gonosomes Modell 17 Tests, polygenes Modell 16 Tests), indem für einen möglichst breiten, aber realistischen Bereich von Modellparametern die zugehörigen Erwartungswerte berechnet wurden. Jedes dieser Testbeispiele wurde mit drei verschiedenen Sätzen für die Anfangswerte durchgerechnet (Tab. 2).

Tabelle 2.

Das schlechteste Ergebnis brachte für das gonosome Modell folgender Test:

	$f_1$	$f_2$	$f_m$	q
Vorgegebene Parameterwerte	0.001	0.5	0.001	0.0005
Geschätzte Parameterwerte	0.000198	0.501	0.000298	0.0001
Summe der Abweichungsquadrate = 0.89 E-05				

Beim polygenen Modell handelt es sich um den Test:

	$P_m^{*})$	$P_w^{*})$	$t_m$	$t_w$
Vorgegebene Parameterwerte	0.005	0.005	0.1	0.05
Geschätzte Parameterwerte	0.00602	0.00635	0.07944	0.0095
Summe der Abweichungsquadrate = 0.65 E-05				

\*) Die Parameter  $P_m$  und  $P_w$  stehen für die Manifestationshäufigkeiten der getrennt behandelten männlichen und weiblichen Teilpopulationen. Unter Berücksichtigung der Gesamtpopulation als Einheit müssen deshalb die Werte von  $P_m$  und  $P_w$  halbiert werden.

In einigen Fällen kam es zum Programmabsturz, da zu große Exponenten auftraten. Dies konnte immer durch geänderte Anfangswerte behoben werden.

Die Optimierung wurde mit einem Programm der NAG-Bibliothek durchgeführt (E04UAF).

Zur Beurteilung der Schätzung wurde überprüft, wie oft die Schätzergebnisse von den Vorgaben abweichen, wenn man die Schätzwerte auf die Stellenzahl der Vorgabe rundet.

Beim gonosomen Modell weichen nach diesen Kriterien von den 204 Einzelschätzwerten 11 von den Vorgaben ab. Die Abweichungen sind minimal. Beim polygenen Modell sieht es nicht so positiv aus. Von den 192 einzelnen Schätzwerten weichen 33 ab. 6 Abweichungen sind unwesentlich. Die übrigen beziehen sich auf 7 Tests, bei denen drei oder vier der geschätzten Parameter so weit von den Ausgangswerten entfernt sind, ohne daß die Summe der Abweichungsquadrate sich offensichtlich vergrößert, daß der Eindruck besteht, daß lokale Minima existieren. In allen Fällen ist es gelungen, mit mindestens einem der Anfangswertvektoren die vorgegebenen Werte wieder aufzufinden.

In der Praxis sollten Modelle dieser Art sicherheitshalber mit mehreren Anfangswertvektoren gerechnet werden.

### 3.4 Trennfähigkeit der Modelle

Um zu überprüfen, wie gut die Modelle zu trennen sind, wurden jeweils die unter einem Modell konstruierten Häufigkeiten als empirische Werte in das andere Modell hineinsteckt.

Es zeigte sich, daß die Anpassung in allen 33 Testbeispielen wesentlich schlechter war, wenn die empirischen Häufigkeiten nicht mit dem »wahren Modell« übereinstimmen.

Werden Daten einer polygenen Population mit Hilfe des gonosomen Modells angepaßt, so liegen die Summen der Abweichungsquadrate zwischen  $0.1E-02$  und  $0.64E-04$ , während der maximale Wert dieses Kriteriums bei Anwendung des richtigen Modells bei  $0.65E-05$  lag.

Werden umgekehrt die Daten des gonosomen Modells mit Hilfe des polygenen Modells angepaßt, liegen die Summen der Abweichungsquadrate zwischen  $0.11$  und  $0.85E-02$ , während der maximale Wert beim richtigen Modell bei  $0.89E-05$  liegt. Ob diese Trennfähigkeit bei praktischen Problemen ausreicht, muß mit Hilfe von empirischen Daten untersucht werden.

### Literatur

- ARNDT, G., K. SCHMIDT (1984): Populationsgenetische Diskrimination zwischen oligo- und polygenem Erbmodus multipler Schwelleneigenschaften. *EDV in Medizin und Biologie* **15**, 57–61.
- EMERY, A. E. H. (1976): *Methodology in Medical Genetics*. Edinburgh, Churchill Livingstone.
- FALCONER, D. C. (1967): The inheritance of liability to diseases with variable age of onset, with particular reference to diabetes. *Ann. hum. Genet.* **31**, 1–20.
- JAMES, J. (1971): Incidence of an attribute in relatives of individuals with the trait. *Ann. hum. Genet.* **35**, 47–49.
- JOHANSSON, I. (1980): *Meilensteine der Genetik*. Hamburg – Berlin, P. Parey.
- LI, C. C., L. SACKS (1954) The derivation of joint distribution and correlation between relatives by use of stochastic matrices. *Biometrics* **10**, 347–460.
- REICH, T., J. JAMES, C. A. MORRIS (1972): The use of multiple thresholds in determining the mode of transmission of semicontinuous traits. *Ann. hum. Genet.* **36**, 163–184.
- SCHMIDT, K. (1984): Populationsgenetische Diskrimination zwischen oligo- und polygenem Erbmodus multipler Schwelleneigenschaften. Forschungsbericht, FU Berlin, FB 8, FG Biometrie.

Eingegangen am: 27. Mai 1987.

Anschrift der Verfasser: Prof. Dr. K. Schmidt, Dr. Gisela Arndt, Freie Universität Berlin, Fachbereich Veterinärmedizin, Fachgebiet Biometrie, Bachtelzenweg 29–31, D-1000 Berlin 33.



EDV in Medizin und Biologie 18 (2–3), 65–66, ISSN 0300-8282  
 © Verlag Eugen Ulmer GmbH & Co., Stuttgart; Gustav Fischer Verlag KG, Stuttgart

## Biometrie und moderne Informationstechnologie – einführende Bemerkungen

R. Haux und H. D. Unkelbach

### Zusammenfassung

*Die Aufgabengebiete von Biometrie und Informatik überschneiden sich, insbesondere bei rechnergestützten Informationssystemen. Es wird versucht, einige Konsequenzen aufzuzeigen, die sich hieraus für die Arbeit eines Biometrikers ergeben.*

*Schlüsselwörter: Biometrie, Informatik, rechnergestützte Informationssysteme.*

### Summary

*The fields of biometry and informatics overlap, especially for computer-supported information systems. It is intended to briefly point out some of the consequences for the work of a biometrician.*

*Keywords: biometry, informatics, computer-supported information systems.*

### Einleitung

Die Arbeitsweisen in der Biometrie werden durch die zur Verfügung stehenden technischen Möglichkeiten mit geprägt.

Obwohl die »automatische« Verarbeitung von Daten schon immer eng mit statistischen Verfahren verbunden war (3), waren auch zur Zeit R. A. Fishers noch Papier und Bleistift die wichtigsten »Datenverarbeitungswerkzeuge«. Man hatte sich deshalb auf diejenigen biometrischen Methoden zu beschränken, die mit geringem Rechenaufwand auskamen. Erst die Verfügbarkeit leistungsfähiger Datenverarbeitungsanlagen ermöglichte den Einsatz rechenintensiver Verfahren. Methoden und Werkzeuge zur Datenbankverwaltung lieferten die (aus der Sicht des Biometrikers) »technologischen« Voraussetzungen zur Analyse großer Datenbestände. Kommerziell verfügbare Statistische Auswertungssysteme ermöglichten zudem einem breiten Benutzerkreis den Einsatz biometrischer Methoden.

Heute ist zu beobachten, wie mit der verfügbaren Informationstechnologie Datenbestände verschiedener Herkunft mit vergleichsweise geringem Aufwand zusammengeführt werden können und wie sich »Informationslandschaften« mit ausgeprägten Kommunikationsinfrastrukturen bilden.

### Biometrie und Informatik

Innerhalb solcher Informationslandschaften, aber auch in zahlreichen anderen Bereichen, überschneiden sich nun die früher eher als disjunkt betrachteten Aufgabengebiete von Biometrie und Informatik. Man kann innerhalb dieser Schnittmenge zudem noch eine unterschiedliche Gewichtung zwischen den beiden Disziplinen beobachten (Abb. 1):

- a) Informatik für die Biometrie. Hier dominiert die Biometrie; die Methoden und Werkzeuge der Informatik tragen zur Beschreibung oder Lösung spezifisch biometrischer Probleme bei (7, 8). Hierzu gehören beispielsweise algorithmische (speziell numerische) Methoden für statistische Verfahren, Methoden des Datenbankentwurfs für klinische Studien oder Expertensysteme für die Biometrie. Dieses Gebiet bezeichnet man – in etwas allgemeinerem Rahmen – als Computational Statistics.
- b) Biometrie und Informatik. Die Aufgaben bei Informationssystemen – insbesondere bei rechnergestützten – lassen sich kaum der einen oder der anderen Disziplin schwerpunktmäßig zuordnen. Aus der Sicht der Informatik dominieren hier Aspekte wie Informationslogistik (3), aber auch der Informationswiedergabe oder der Datenanalyse. Aus der Sicht der Biometrie handelt es sich jedoch bei einer solchen Datenanalyse in vielen Fällen um nichts anderes als um die statistische Auswertung von Daten aus nichtrandomisierten Studien.
- c) Biometrie (und hier insbesondere Statistik) für die Informatik. Hier dominiert die Informatik; die Methoden der Biometrie tragen zur Beschreibung oder Lösung spezifischer Probleme der Informatik bei. Hierzu gehören beispielsweise statistische Ansätze für die Biosignal- und Bildverarbeitung, Ansätze aus der Warteschlangentheorie für die Modellierung von Betriebssystemen oder statistische Verfahren für Inferenzmechanismen in Expertensystemen.

Besonders bei b), das heißt bei der Konzeption von (rechnergestützten) Informationssystemen, sollte die Biometrie in ausreichendem Maße berücksichtigt werden: Eine nur auf den technischen Möglichkeiten basierende und nicht vorwiegend auf inhaltliche Fragestellungen ausgerichtete, auf Aggregation beruhende Auswertung von Datenbeständen führt nahezu zwangsläufig zu dem »garbage data in – garbage statistics out« Syndrom (das sich auch durch noch so umfangreiche und optisch gut aufbereitete Computerausdrucke nicht beheben läßt).

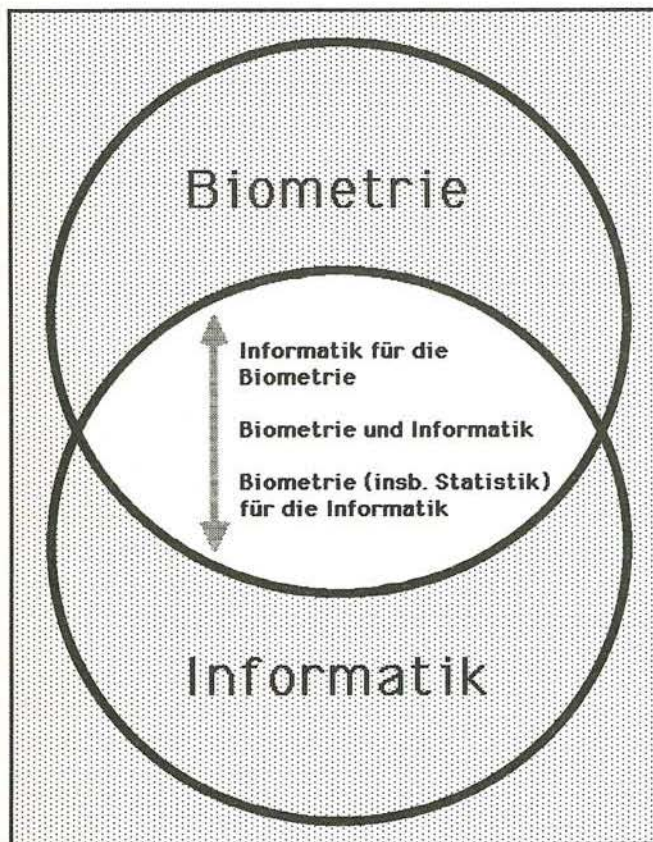


Abb. 1. Überschneidung der Aufgabengebiete von Biometrie und Informatik

Der Biometriker ist hier gefordert, Stellung zu nehmen und die methodischen Möglichkeiten und Grenzen statistischer Verfahren aufzuzeigen. Er darf sich in seiner Arbeit allerdings nicht nur auf die »klassischen« biometrischen Arbeitsgebiete beschränken – etwa auf Planung, Durchführung und Auswertung randomisierter Studien. Er sollte beispielsweise auch bei der Auswertung von im Routinebetrieb einer Klinik gewonnenen Daten und vor allem bei der Planung der Erhebung dieser Daten mitwirken. Gerade bei rechnergestützten Informationssystemen ergeben sich hier wichtige Aufgaben für den Biometriker.

### Biometrie und moderne Informationstechnologie

Auf dem 33. Biometrischen Kolloquium der Deutschen Region der Internationalen Biometrischen Gesellschaft wurde von den Autoren eine Sitzung organisiert, die sich mit »Biometrie und moderner Informationstechnologie« befaßte. Behandelt wurden einige ausgewählte Themen aus a) und b). Die überarbeiteten und erweiterten Vortragsausarbeitungen aus dieser Sitzung sind in diesem Heft zusammengefaßt.

– Zu b):

In zwei für den Biometriker wichtige Informationslandschaften – bzw. in Teile davon – wird exemplarisch Einblick gegeben:

- in die Informationslandschaft der Pharmaforschung (6) und
- in die Informationslandschaft einer Klinik (1). Hier werden Möglichkeiten für eine statistische Auswertung von Patientendaten aufgezeigt.

Die Autoren stellen Standort und Rolle der Biometrie in dem oben ausgeführten Sinn dar und weisen auf die Möglichkeiten der Synthese von biometrischen Methoden und von Methoden und Werkzeugen der Informatik hin.

– Zu a):

Während der Biometriker früher oft das eingesetzte statistische Verfahren selbst programmiert hat, bedient er sich heute zunehmend vorhandener Software, etwa der kommerziell verfügbaren Statistischen Auswertungssysteme. Die Verantwortlichkeit des Biometrikers für die von ihm ermittelten Ergebnisse bedingt dann, daß er sich auf die Richtigkeit der eingesetzten Software verlassen können muß. Er benötigt somit Kriterien, anhand derer er die Qualität der von ihm verwendeten Software beurteilen und sie gegenüber Dritten dokumentieren kann. Dieses Problem wird in (2) diskutiert.

Eine andere, für die Biometrie relevante (und unter Biometrikern kontrovers diskutierte) Entwicklung sind Statistische Expertensysteme. In (5) wird hierzu kritisch Stellung genommen und diskutiert, ob (und – wenn ja – wie) der Biometriker durch solche Werkzeuge bei einem Teil seiner Routinearbeiten unterstützt und entlastet werden kann.

In allen Arbeiten wird versucht, den Nutzen, den die moderne Informationstechnologie der Biometrie bringt, aufzuzeigen. Die Autoren versuchen darzustellen, welche Neuorientierung und welche Veränderung der Arbeitsweise eines Biometrikers dies mit sich bringt bzw. mit sich bringen sollte.

### Schlußbemerkung

Die Arbeitsweisen in der Biometrie verändern sich durch die heute zur Verfügung stehenden technischen Möglichkeiten.

Gleichgeblieben ist jedoch das Grundprinzip biometrischen Arbeitens: die Auswahl bzw. die Entwicklung des zum biologischen Problem passenden (mathematischen) Modells, des (statistischen) Verfahrens zur Datenanalyse sowie die (rechnergestützte) Auswertung der Daten und die Unterstützung bei der Interpretation der Ergebnisse.

### Literatur

- (1) GIANI, G., und W. MEYERS (1987): Verwaltung und Auswertung von klinischen Registern. EDV in Medizin und Biologie **18**, 75–83.
- (2) PASSING, H. (1987): Was bedeuten Software-Qualitätssicherung und -Validierung für den Biometriker? EDV in Medizin und Biologie **18**, 84–89.
- (3) REICHERTZ, P. L. (1982): Future developments of data processing in health care. Meth. Inform. Med. **27**, 55–58.
- (4) SINT, P. (1984): Roots of computational statistics. Havranek, T., Z. Sidak, M. Novak (Hrsg.). COMPSTAT 1984, 9–20. Wien: Physika.
- (5) STREITBERG, B. (1987): Expertensysteme in der Biometrie – Experten oder Ignoranten? EDV in Medizin und Biologie **18**, 90–94.
- (6) UNKELBACH, H. D. (1987): Biometrie in der Informationslandschaft der Pharmaforschung. EDV in Medizin und Biologie **18**, 67–74.
- (7) VICTOR, N. (1984): Computational statistics – tool or science? Stat. Software Newsl. **10**, 105–116.
- (8) ZELEN, M. (1983): Biostatistical science as a discipline – a look into the future. Biometrics **39**, 827–837.

Eingegangen am: 15. April 1987.

Anschriften der Verfasser: Priv.-Doz. Dr. Reinhold Haux, Universität Tübingen, Institut für Medizinische Informationsverarbeitung, Westbahnhofstraße 55, D-7400 Tübingen; Dr. Hans Dieter Unkelbach, Firma E. Merck, TWI/BD, Postfach 41 19, D-6100 Darmstadt.

## Biometrie in der Informationslandschaft der Pharmaforschung

H. D. Unkelbach

### Zusammenfassung

Für die Erforschung und Entwicklung neuer Arzneimittel sind in einem pharmazeutischen Unternehmen folgende organisatorische Einheiten eingerichtet: Pharmazeutische Chemie, Pharmakologie, Toxikologie und Klinische Forschung. Für jede dieser Einheiten hat die Biometrie (bei der Pharmazeutischen Chemie auch oft »Chemometrie« genannt) die Aufgabe, aus Rohdaten zusammenfassende Aussagen mit definierter Zuverlässigkeit zu ermitteln, also »Information« zu schaffen.

Zur Unterstützung verschiedener Teilaufgaben haben zunehmend DV-Systeme Eingang gefunden: Statistische Auswertungssysteme, Datenbanksysteme, Laborinformations- und Labormanagementsysteme, Text- und Grafiksysteme usw. Diese Systeme sind zunächst isoliert voneinander entstanden. Die rasche Entwicklung der modernen Informations- und Kommunikationstechnologie ermöglicht eine immer enger werdende Vernetzung der Systeme; das Zusammenwachsen der einzelnen Systeme zu einem die gesamte Pharmaforschung eines Unternehmens übergreifenden Informationsverbund ist zu erkennen.

In dieser entstehenden Informationslandschaft hat die Biometrie einen zentralen Standort, weil bei ihr Daten aus den verschiedenen Forschungseinheiten zusammenfließen. Die sich aus den Auswertungen dieser Daten ergebende Gesamtinformation bildet eine Basis für strategische Entscheidungen des Forschungsmanagements.

Man beginnt heute, Information als Ressource eines Unternehmens zu begreifen und spricht von ihr als dem »dritten Produktionsfaktor«. Dort, wo sich dieses Verständnis durchsetzt, erwächst der Biometrie eine neue Bedeutung im Rahmen des Informationsmanagements.

Schlüsselwörter: Biometrie, Informatik, Informationsmanagement, Informationsnetze, Pharmazeutische Forschung

### Summary

For research and development of new drugs in a pharmaceutical company the following organizational units exist: Pharmaceutical Chemistry, Pharmacology, Toxicology and Clinical Research. It is the function of biometry for each of these units to extract comprehensive results with defined reliability from a large body of raw data. In other words: The function of biometry is the production of information.

In the last 20 years data processing systems have been introduced which support various tasks of the biometrician: statistical analysis systems, data base systems, laboratory information and management systems, text and graphics systems, etc. At

first, these systems were installed independently of each other. But as modern information and communication technology progressed, these systems became more and more connected with each other.

The result of this ongoing integration process nowadays can be seen as an information network covering all R & D-divisions of a pharmaceutical company. Biometry holds a central domain within this growing information scenery, because it is there that the streams of raw data of the various R & D-departments flow together and are finished into refined information, which provides a basis for the decisions of the research management.

Today information is seen as a resource of a company, especially of a researching company, and information is called the »third production factor«. This understanding may result in a new significance of biometry within the scope of information management.

Keywords: biometry, informatics, information management, information networks, drug research

### 1. Einleitung

Informations- und Kommunikationstechnik dringen im Zuge ihrer rasanten Entwicklung in immer mehr Bereiche unseres täglichen Lebens ein. Nachdem die technischen Veränderungen massiv spürbar geworden sind, beginnt man nun, die Entwicklung auch geistig zu verarbeiten. Dabei hat sich die Information als solche als zentraler Begriff herausgeschält.

Um diesen Begriff ranken sich auch die Schlagworte, mit denen die Bedeutung dieser neuen Entwicklung ausgedrückt wird. Bei Unternehmen spricht man neuerdings von der Information als »Ressource« des Unternehmens und in demselben Sinne von der Information als dem »dritten Produktionsfaktor«. Auch in der Biometrie beginnt man, in diesen Kategorien zu denken. So veranstaltete im Januar 1987 die American Statistical Association (ASA) ein Kolloquium unter dem Motto »Statistics in the Information Age«.

Daß sich durch moderne Informations- und Kommunikationstechnologie vieles verändert hat, haben die Biometriker am eigenen Leibe erlebt. Ihre Arbeitsweisen und Arbeitsabläufe haben sich stark gewandelt. Die Frage ist: Haben sich die Dinge nur technisch verändert oder ist gar das Wesen der Biometrie betroffen? Wenn die Schlagworte über die Information nicht nur hohle Worte sind, was bedeutet das für die Biometrie in einem pharmazeutischen Unternehmen?

Im vorliegenden Aufsatz wird versucht, den Hintergrund zu diesen Fragen aufzuhellen. Dazu werden im folgenden zweiten Abschnitt zunächst die Aufgaben der Biometrie in der Pharmazeutischen Forschung aufgezählt. Über die Unterstüt-

zung dieser Aufgaben durch moderne Datenverarbeitungssysteme wird in Abschnitt 3 und in Abschnitt 4 über den Nutzen, den sie der Biometrie gebracht haben, referiert. In Abschnitt 5 wird versucht, die Tätigkeiten der Biometrie unter dem Aspekt des Informationsgeschehens darzustellen, wobei sich ein Bild der Informationslandschaft in der Pharmaforschung ergibt. Die Bedeutung der biometrischen Tätigkeiten wird schließlich im 6. Abschnitt durch Weiterführen des Gedankens von Forschung als Produktion von Information im Vergleich zur Produktion von Gütern aufgezeigt. Im letzten Abschnitt 7 werden die Konsequenzen für die zukünftige Entwicklung der Biometrie in der Pharmaforschung diskutiert.

## 2. Aufgaben in Pharmaforschung und -entwicklung

In einem forschenden pharmazeutischen Unternehmen sind Forschung und Entwicklung neuer Arzneimittel in folgende organisatorische Einheiten untergliedert:

- Pharmazeutische Chemie
- Pharmakologie
- Toxikologie
- Klinische Forschung

Jedem Bereich entspricht ein Aufgabengebiet. Aufgabe der Pharmazeutischen Chemie ist die chemische Synthese neuer Wirkstoffe für Arzneimittel. In der Pharmakologie wird anschließend die biologische Wirksamkeit der synthetisierten Substanzen geprüft. Dies geschieht zunächst nach einem groben Untersuchungsrastrer (»Screening«). Nachdem hierdurch Hinweise über die Wirkungsschwerpunkte gewonnen wurden, wird in der Allgemeinen und Speziellen Pharmakologie das Wissen über Art und Stärke der Wirkungen verbreitert und vertieft.

Anders als in der Pharmakologie, wo die »erwünschten Wirkungen« Ziel der Forschung sind, ist es die Aufgabe der Toxikologie, zu untersuchen, ob und welche Nebenwirkungen eine Substanz besitzt. Die Toxikologie soll die Unbedenklichkeit und Verträglichkeit von Arzneimitteln – soweit dies im Tierversuch möglich ist – abklären.

Die Erprobung des neuen Wirkstoffes am Menschen ist Aufgabe der Klinischen Forschung. In den Phasen I–IV der klinischen Prüfungen werden – beginnend mit sehr niedrigen Dosen an gesunden Freiwilligen bis zu groß angelegten Studien an Kranken mit therapeutischen Dosen und begleitenden Untersuchungen nach der Zulassung – Wirkungen und Nebenwirkungen des Arzneimittels am Menschen erforscht. Für eine ausführlichere Darstellung sei z. B. auf (15) verwiesen.

Ein Aufgabenbereich im Rahmen der Forschung und Entwicklung neuer Arzneimittel, der eigenständigen Charakter hat und deshalb oft als ein unabhängiger Bereich organisiert ist, ist die Galenik. Hier wird erforscht, welches die zweckmäßigste Darreichungsform (Tablette, Dragee, Zäpfchen, Salbe, Spritze, etc.) im Hinblick auf die Pharmakokinetik und Verträglichkeit des Wirkstoffes ist.

In all den genannten Forschungs- und Entwicklungsbereichen treten Teilaufgaben auf, die vom Biometriker zu erfüllen sind. Die Aufgabe der Biometrie in einem pharmazeutischen Unternehmen ist »die Anwendung mathematischer Methoden auf biologische Vorgänge und die Quantifizierung der Sicherheit bei Aussagen über Unbedenklichkeit und Wirksamkeit chemisch-pharmazeutischer Produkte« (1).

In der Pharmakologie, der Toxikologie und der Klinischen Forschung wird die Biometrie hauptsächlich im Rahmen von tierexperimentellen Studien und klinischen Prüfungen tätig:

- Studienplanung (mathematische Modellbildung, -überprüfung, Studiendesign mit Stichprobenumfängen und Randomisierungsplänen, Planung der Datenerfassung und -dokumentation)
- studienbegleitende Tätigkeiten (Plausibilitäts- und Vollständigkeitsprüfungen, Datenspeicherung, Bestandsaufnahmen, Zwischenberichte)
- Studienauswertung (statistische Auswertung der Daten, Datenpräsentation, Dokumentation, Interpretation der Ergebnisse, Bericht)

Die Tätigkeiten der Biometrie für den Bereich der Pharmazeutischen Chemie werden oft als »Chemometrie« bezeichnet. Sie bestehen darin, formelmäßige Zusammenhänge zwischen der chemischen Struktur von Substanzen und ihren biologischen Eigenschaften herzustellen. Pharmakologische, toxikologische und klinische Kenntnisse vorhandener Substanzen setzt man miteinander in Beziehung und untersucht, inwieweit ähnliche biologische Eigenschaften mit verwandter chemischer Struktur zusammenhängen. Hier schließt sich der Kreis in der Pharmaforschung: Durch frühere Forschung in Pharmakologie, Toxikologie und Klinischer Forschung erhält die Pharmazeutische Chemie Hinweise, von welcher chemischen Struktur neue Substanzen sein könnten, die bestimmte interessierende Eigenschaften erwarten lassen.

## 3. DV-Systeme

Der Begriff Datenverarbeitung (DV) wird im folgenden in üblicher pauschaler Bedeutung verwendet. Insbesondere wird nicht unterschieden, ob es sich um Methoden der DV (z. B. Datendesign) oder um Werkzeuge der DV (z. B. Datenbanksysteme) handelt.

Wenn man heute Biometrie-Abteilungen betrachtet und ihre Tätigkeiten beobachtet, so entsteht oberflächlich der Eindruck, als handle es sich um reine DV-Abteilungen, da praktisch für alle Tätigkeiten DV-Systeme eingesetzt werden. Es sollte deshalb festgehalten werden: Alle obengenannten Aufgaben der Forschungs- und Entwicklungsbereiche sowie die der Biometrie sind **inhaltlich** definiert und damit **unabhängig** von der **technischen** Durchführung, insbesondere also von jeglicher Art von DV-Unterstützung. In früheren Zeiten wurden diese Aufgaben ohne DV erfüllt. Auch wenn heutige Biometrie nicht mehr ohne DV denkbar ist, so sind doch die inhaltlichen Aufgabenstellungen die gleichen geblieben, während sich lediglich ihre technische Durchführung gewandelt hat, diese allerdings in einem solchen Ausmaß, daß man von einer grundlegenden Änderung der Arbeitsweisen in der Biometrie sprechen kann. Außerdem sind durch die technischen Möglichkeiten auch neue Teilaufgaben hinzugekommen.

Anfangs, beginnend in den 50er Jahren, hatte der Einsatz von Datenverarbeitung zum Ziel, den enormen **Rechenaufwand** bei der statistischen Auswertung, der nur durch großen Personaleinsatz geleistet werden konnte, dem Rechner zu übertragen. Die Ergebnisse dieser Auswertungen wurden mit der Schreibmaschine in Form von Tabellen gestaltet und dem ebenfalls maschinengeschriebenen Bericht beigelegt. Man erkannte jedoch bald, daß auch die **Dokumentation** der Versuchsdaten und der Auswertungsergebnisse recht gut vom Rechner übernommen werden konnte, fielen doch die Tabelleninhalte ohnehin fast als Nebenprodukt bei der Auswertung an.

Viel später, erst in der zweiten Hälfte der 70er Jahre, begann man, auch die Erzeugung von **Grafiken** der DV zu übertragen und die Berichte mit einem **Textsystem** zu schreiben.

Einen weiteren wesentlichen Fortschritt brachten **Datenbanksysteme**. Zu Beginn der DV standen die Rechenverfahren, also die Programme, im Vordergrund, die Daten wurden für das jeweilige Verfahren eigens zusammengestellt, die Struktur der Daten richtete sich nach dem Programm. Sollten die gleichen Daten nach einem weiteren Verfahren ausgewertet werden, so mußten sie i. a. in eine andere Form gebracht werden. Datenbanksysteme ermöglichten es, die Daten einmalig in einer programmunabhängigen Form zu verwalten und verfügbar zu halten.

Planung und Vorbereitung der Datenerfassung von Experimenten und Studien gehört – wie oben aufgeführt – zu dem Tätigkeitsbereich der Biometrie. Hierzu wurden in der Vergangenheit Erfassungsbelege (»Ablochbelege«) entworfen, in die die Daten eingetragen und in zentralen Datenerfassungsstellen von Datentypisten zuerst auf Lochkarten, später auf magnetische Datenträger übernommen wurden. Viele Daten entstehen auf Geräten wie Waagen, Zeitmeßgeräten, Analyseautomaten etc., die mittlerweile mit elektronischem Datenausgang ausgestattet werden. Diese Daten nimmt man heute **online** auf magnetische Datenträger, Band oder Platte, von wo aus sie direkt in ein Datenbanksystem oder ein Auswertungssystem übernommen werden können.

Zu einem Versuch oder einer Studie gehört eine Fülle von Daten, die i. a. an verschiedenen Orten entstehen. Alle diese Daten müssen für die Versuchsauswertung zusammengeführt werden. Aber auch bereits während der Versuchsdurchführung ist es wichtig, von Zeit zu Zeit den Zwischenstand überschauen zu können. Zu diesem Zweck haben in Laborbereichen sog. **LIMS-Systeme** (Labor-Information-Management- und -Steuerungs-Systeme) ab Ende der 70er Jahre Einzug gehalten. Diese Systeme unterstützen die Versuchsplanung, drucken gemäß der gespeicherten Versuchspläne Arbeitslisten für die verschiedenen Labors, fordern die fälligen Daten an, sammeln sie an den Entstehungsorten, führen sie in einem einheitlichen System zusammen und stellen sie für Auswertungen bereit.

Während sich für pharmakologische und toxikologische Versuche wegen der Nähe der beteiligten Labors LIMS-Systeme noch relativ einfach einrichten lassen, ist für große klinische Studien, an denen mehrere Krankenhäuser beteiligt sind, die Datenerfassung nicht so einheitlich zu gestalten. Hier beginnt man, **mobile Datensammelstationen** zu verwenden, mit deren Hilfe Daten vor Ort erfaßt und über das öffentliche Telefonnetz an den Rechner, der den Datenbestand der Gesamtstudie verwaltet, übermittelt werden können.

Es war typisch für den Einzug der Datenverarbeitung in die pharmazeutische Forschung und Entwicklung, daß die verschiedenen DV-Systeme zur Unterstützung der jeweiligen Teilaufgabe mehr oder weniger unabhängig voneinander eingeführt wurden. Dies hatte die Bildung von »Informationsinseln« zur Folge. Die in einem LIMS-System gesammelten Daten können oft nur mit Mühe in ein Datenbanksystem oder ein Auswertungssystem übertragen werden. Die heutigen Anstrengungen sind deshalb auf die **Integration** der Systeme gerichtet. Durch übergreifende konzeptionelle Planung verfolgt man das Ziel, die verschiedenen Systeme durchlässig zu gestalten und sie in einen Gesamtrahmen einzubinden. Die technischen Voraussetzungen hierzu bieten Netzwerke mit standardisierter Netzwerksarchitektur.

#### 4. Nutzen der DV-Systeme für die Biometrie

Die Biometrie hat von Anfang an den Einzug der Datenverarbeitung in der pharmakologischen Forschung und Entwicklung begleitet. Sie hat ihn nicht passiv über sich ergehen lassen, sondern war bei vielen Systemen der Initiator und hatte die Federführung bei der Beschaffung und Installation. Dies nicht zuletzt deshalb, weil sie selbst intensiver Anwender der Systeme ist. Es soll kurz zusammengestellt werden, welchen Nutzen die Biometrie durch moderne DV-Systeme erfahren hat. Dabei werde nach dem Daten-, Auswertungs- und Planungsaspekt unterschieden. Es wird dagegen nicht eingegangen auf Fragen der Software-Erstellung und dem damit verbundenen Nutzen von Software-Entwicklungs-Umgebungen (hierzu sei auf (9) verwiesen).

##### Datenaspekte

Datenbanksysteme entlasten von technischem Aufwand bei der Verwaltung der Daten. Sie stehen unabhängig vom Verwendungszweck für jede Art der Auswertung zur Verfügung. Durch die Vermeidung bzw. die Kontrolle der Datenredundanz (und deshalb z. B. Vermeidung der »Update-Anomalie«, s. (10)) ist weitgehende Konsistenz der Daten und damit – verglichen mit früher – eine höhere **Datenqualität** erreicht.

Datenqualität wird zusätzlich durch LIMS-Systeme verbunden mit online-Datenerfassung von Meßgeräten oder mobile Datenerfassungssysteme erhöht. Die Daten gelangen nämlich frühzeitig – oft unmittelbar nach ihrem Entstehen – ins System und werden dort vorbereiteten Plausibilitätsuntersuchungen unterworfen. Dadurch können falsche Meßwerte oder Fehleingaben noch am Ort des Entstehens aufgedeckt und möglicherweise die korrekten Werte beschafft werden, was später oft nicht mehr möglich ist. (Hierdurch entfällt ein großer Teil des mühseligen »Datenputzens« nach Abschluß einer Studie.)

Es sei angemerkt, daß die Einführung von LIMS-Systemen in toxikologischen Bereichen wesentlich durch die Forderungen der amerikanischen Gesundheitsbehörde (FDA – Food and Drug Administration) über Datenqualität und -integrität forciert wurde. Mit entsprechenden Systemen kann die Einhaltung bestimmter GLP-Regeln (Good Laboratory Practice, s. (4)) garantiert werden. Diese verlangen z. B., daß ein einmal erfaßter und durch die für die Erfassung zuständige Person akzeptierter Wert (Meßwert, Befund etc.) verwahrt werden muß, wobei jede Änderung nur mit Angabe des Datums, des Grundes und der verantwortlichen Person möglich ist und der ursprüngliche Wert nicht gelöscht werden darf.

Neben der Verbesserung der Datenqualität ist vor allem die starke Erhöhung der **Datenverfügbarkeit** durch Netzwerke, Datenbanken und Informationsdienste zu nennen.

Welchen Nutzen eine durchgängige Datenstruktur aller eingesetzten Softwarekomponenten bei klinischen Studien ermöglicht, wird in (14) aufgezeigt.

##### Auswertungsaspekte

Die Verfügbarkeit großer leistungsfähiger Rechner ermöglichte den Einsatz **rechenintensiver** statistischer Verfahren, man denke z. B. an exakte nicht-parametrische Teststatistiken oder Monte-Carlo-Verfahren.

Statistische Auswertungssysteme, die auf dem Markt verfügbar sind (z. B. SAS, SPSS, BMDP, RS/1), bieten eine breite Palette mathematisch/statistischer Verfahren. Für viele der Systeme ist mittlerweile auch eine PC-Version erhältlich. Daneben gibt es nach wie vor Unterprogrammssammlungen für numerische Verfahren, wie z. B. NAG-Library, IMSL.

Wegen der weiten Verbreitung und der großen Benutzerge-  
meinde kann man davon ausgehen, daß die käuflichen Aus-  
wertungssysteme im Laufe der Zeit einen hohen Qualitäts-  
stand erreicht haben, da der Großteil der Programmfehler  
aufgefallen und in neuen Releases behoben worden ist.

Die Biometrie ist dadurch von der Eigenprogrammierung  
der Basis-Verfahren entlastet, nur noch spezielle Verfahren  
müssen selbst entwickelt und können meist in die Auswer-  
tungssysteme eingebunden werden.

Die heutigen Statistischen Auswertungssysteme besitzen  
eine mehr oder weniger komfortable Benutzeroberfläche und  
bieten neben der Auswertung selbst die Möglichkeit, die  
Auswertungsergebnisse gemeinsam mit den zugehörigen  
Daten zu dokumentieren und tabellarisch oder grafisch zu  
präsentieren. Bei geeigneter Vorsorge beim Arbeitsablauf und  
der Vergabe von Dokumentationskennungen kann dann die  
Konsistenz zwischen den ausgedruckten Rohdaten, Auswer-  
tungsergebnissen, Tabellen und Grafiken automatisch garanti-  
ert werden.

Die einheitliche Datenschnittstelle eines Auswertungssy-  
stems ermöglicht es, denselben Datensatz nach verschiedenen  
Gesichtspunkten hin mit verschiedenen Verfahren explorativ  
auszuwerten. Dazu bieten leistungsfähige grafische Systeme  
bequeme Möglichkeiten, denselben Datensatz auf verschie-  
dene Arten zu visualisieren und die aussagefähigste Darstel-  
lungsform auszuwählen. Entsprechendes gilt für optimal  
gestaltete Tabellen durch leistungsfähige Reportgeneratoren.  
**Explorative Datenanalyse (EDA)** ist im Grunde erst durch  
leistungsfähige DV möglich geworden.

Bei der Interpretation der Ergebnisse einer Studie interes-  
siert neben den Auswertungsergebnissen der Studie selbst  
meist noch der Bezug zu früheren, vergleichbaren Studien. In  
der Vergangenheit konnte dieser Bezug nur durch die Erfah-  
rung des Fachwissenschaftlers hergestellt werden. Die Verfüg-  
barkeit »historischer« Daten erlaubt es nun, diese »Erfah-  
rung« des Fachwissenschaftlers zu objektivieren. Existiert eine  
durchgängige Datenstruktur und werden die Daten in einem  
Datenbanksystem gehalten, so wird der Rückgriff auf histori-  
sche Daten und der Vergleich mit den aktuellen Daten erleich-  
tert, studienübergreifende Analysen werden möglich. Man  
denke z.B. an die Berechnung von Normalbereichen für  
Blutwerte des im Unternehmen verwendeten Rattenstammes.  
(In diesem Zusammenhang sei auch auf (3) verwiesen; dort  
wird auf die analoge Problematik bei großen Krankenregistern  
eingegangen.)

Die historischen Datenbestände bieten darüber hinaus aus-  
gezeichnetes Material bei der Entwicklung biometrischer Aus-  
wertemethodiken. Die mathematischen Modellannahmen  
können anhand der Daten unbehandelte Tiere bzw. Proban-  
den auf Tauglichkeit geprüft werden, die Wahrscheinlichkeits-  
verteilung von Teststatistiken kann durch Simulationen über  
den empirischen Verteilungen ermittelt werden.

**Planungsaspekte**

LIMS-Systeme ermöglichen es, Studien vor ihrem Beginn im  
Detail zu planen. Dadurch kann nicht nur festgelegt werden,  
wann welche Daten erhoben werden müssen, auch alle Ran-  
domisierungsaktivitäten können soweit vorbereitet werden,  
daß die Randomisationslisten automatisch generiert werden  
können. Das gleiche gilt für die Auswertung. Vorab können  
alle Methoden und Verfahren bereitgestellt werden, so daß  
bei Abschluß der Studie die geplante konfirmative statistische  
Auswertung vorgenommen werden kann.

LIMS-Systeme unterstützen die Durchführung einer Studie  
durch die Vorgabe der Planung. Die Erarbeitung der Studien-

planung selbst unterstützen sie dagegen nicht. Aber auch für  
die Auswahl des Versuchsdesigns, die Berechnung geeigneter  
Stichprobenumfänge sowie die Ausarbeitung des Versuchspla-  
nes werden heute bereits Softwaresysteme angeboten, so z. B.  
DISCOVER von BBN (Bolt, Boranek and Newman, USA),  
das mit dem Auswertungssystem RS/1 arbeitet. Dieses  
System, das ursprünglich für statistische Versuchsplanung in  
nicht-biologischen Bereichen entwickelt wurde, ist für den  
biometrischen Bereich durchaus nützlich, wenn es um einfache,  
standardisierte Designs geht. Die Hoffnung auf befriedi-  
gende DV-Unterstützung bei der Versuchsplanung stützt sich  
auf **Statistische Expertensysteme**; hierzu sei auf (5, 13) ver-  
wiesen.

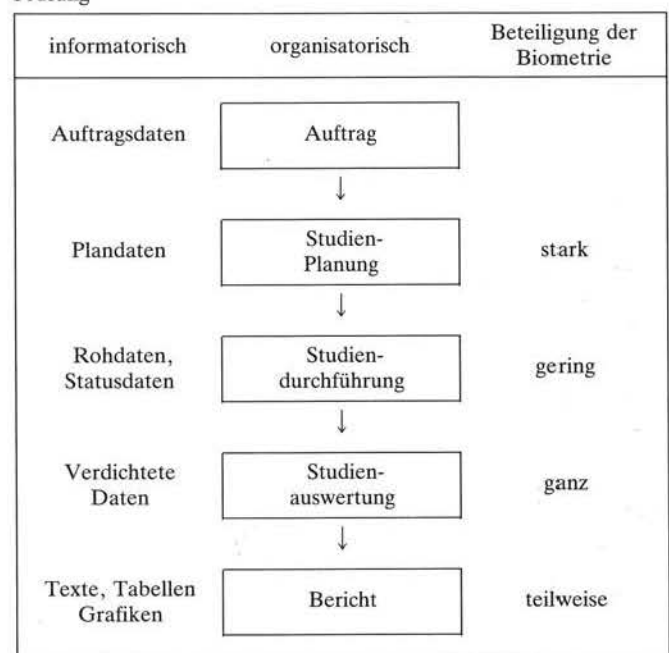
Zusammenfassend läßt sich feststellen, daß moderne DV  
die Arbeit des Biometrikers effizienter, qualitativ besser und  
bequemer gemacht hat und daß viele biometrische Leistungen  
erst durch DV möglich geworden sind.

**5. Informationslandschaft**

Auch wenn sich die generelle Aufgabenstellung der Biometrie  
im Rahmen der Pharmaforschung und -entwicklung nicht ver-  
ändert hat, so scheint sich dennoch das Verständnis dessen,  
was die Biometrie leistet, gewandelt zu haben und ihre Bedeu-  
tung im Unternehmen stärker geworden zu sein. Dies läßt sich  
sicherlich nicht dadurch erklären, daß heute viele ihrer Aufga-  
ben durch DV unterstützt werden – das ist an vielen anderen  
Stellen im Unternehmen auch der Fall. Der Grund scheint  
m.E. vielmehr darin zu bestehen, daß der Gegenstand der  
Biometrie, die »Informationsverarbeitung«, einen anderen  
Stellenwert erlangt hat. (Von einer parallelen Entwicklung im  
öffentlichen Gesundheitswesen wird in (11) berichtet.)

Im folgenden wird deshalb versucht, die Funktion der Bio-  
metrie aus »informatrischer« Sicht darzustellen. Dabei wer-  
den wir uns der Einfachheit halber hauptsächlich auf die  
Abläufe im operativen Betrieb beschränken und Tätigkeiten  
im Zusammenhang mit Methoden- oder Softwareentwicklung  
oder Know-How-Erwerb außer acht lassen.

Abb. 1. Entstehende Informationen beim Ablauf einer biologischen  
Prüfung



Verfolgen wir zunächst den Ablauf einer präklinischen oder klinischen Studie und betrachten, welche Informationen dabei entstehen (s. Abb. 1). Vom Forschungsmanagement wird der Auftrag zur Studie gegeben, dabei fallen Daten an, die den Typ, die Zielsetzung, den Zeitbereich usw. der Studie betreffen (Auftragsdaten). Der Auftrag wird unter der Federführung des verantwortlichen Studienleiters in die Studienplanung umgesetzt. Hierbei wird festgelegt, welche Aktionen nach welchen Methoden zu welchen Zeitpunkten von welchen Personen im Verlaufe der Studie vorgenommen werden (Plandaten). Bei der Erarbeitung der Plandaten wirkt die Biometrie wesentlich mit. Ein Teil der Plandaten betrifft die Biometrie selbst (Randomisationen, Plausibilitätsprüfungen, Auswertungen).

Bei der Durchführung der Studie entstehen die (eigentlichen) Daten, die Rohdaten (z. B. bei tierexperimentellen Studien Körpergewichte der Tiere, Blutwerte oder Befunde). Jedes Rohdatum ist nicht nur durch den Wert selbst, die Meßmethode, die Versuchsbedingungen sowie das Individuum, zu dem es gehört, gekennzeichnet, sondern auch durch die Angabe des genauen Zeitpunktes (Datum, Uhrzeit), die Person, die den Wert erhoben hat, sowie eine Aussage über Akzeptanz bzw. Validierung des Wertes. Zu letzterem werden meist Plausibilitätsprüfungen vorgenommen. Neben den Rohdaten fallen bei der Studiendurchführung auch »Statusdaten« an, die Auskunft über den Stand der Studie geben. Anhand dieser Daten kann die Bereichsleitung oder das Forschungsmanagement durch Vergleich mit den Plandaten erkennen, wie weit die Prüfung fortgeschritten ist.

Am Ende der Studiendurchführung sind alle Rohdaten erhoben, und von der Biometrie kann die Studienauswertung vorgenommen werden. Dabei entstehen aus Rohdaten verdichtete Daten (Mittelwerte, Streuungen, Mediane, Regressionskoeffizienten, ED<sub>50</sub>, Signifikanzaussagen usw.). Ein Teil dieser verdichteten Daten war bereits in der Studienplanung vorgesehen (konfirmative Analyse), ein anderer Teil entsteht »spontan« bei der explorativen Analyse nach Kenntnis der Prüfungsergebnisse. Neben den Rohdaten der aktuellen Prüfung werden für die Auswertung oft auch Daten aus früheren Prüfungen des gleichen Typs hinzugezogen, dies können Rohdaten oder verdichtete Daten sein. Man denke etwa an Normalbereiche bei hämatologischen Parametern oder an A-priori-Verteilungen bei Bayesischen Schätzungen.

Am Ende des Ablaufs einer Studie ist der Bericht zu verfassen. Neben der Formulierung der Ziele und der eingesetzten Methoden sind die Ergebnisse zu beschreiben und zu interpretieren sowie eine Bewertung vorzunehmen. Den Berichtstext formuliert der Fachwissenschaftler unterstützt durch den Biometriker bei der Methodenbeschreibung und der Ergebnisinterpretation, Tabellen und Grafiken erstellt der Biometriker.

Eine Entwicklungssubstanz durchläuft in den verschiedenen Bereichen der Pharmaforschung eine Vielzahl von Studien. Die Entscheidung, ob die Entwicklungssubstanz als Arzneimittel auf den Markt gebracht werden soll, hängt nicht vom Ergebnis der einzelnen Studien ab, es müssen vielmehr zusammenfassende Beurteilungen mehrerer Studien vorgenommen werden. Im Bestreben, diese Beurteilungen zu objektivieren, wird neuerdings die Biometrie in diesen Prozeß eingeschaltet. Praxis ist dies bereits in der Klinischen Forschung. Über entsprechende biometrische Methoden und zugrundeliegende mathematische Konzepte derartiger »Meta-Analysen« geben (7, 8) einen Einblick.

Der Wirkungsbereich der Biometrie reicht aber noch weiter über den Entwicklungszeitraum eines Arzneimittels hinaus

und umfaßt zunehmend den gesamten Zeitbereich des Einsatzes des Arzneimittels in der ärztlichen oder klinischen Praxis. Ein Beispiel sind die spontan gemeldeten Nebenwirkungen. Die WHO (World Health Organization) fordert von den Herstellern von Arzneimitteln, daß sie zu jedem Arzneimittel alle unerwünschten Nebenwirkungen sammeln, bewerten und berichten. Beim Aufbau der Nebenwirkungsdatenbanken, die Anforderungen innerhalb des Unternehmens als auch gegenüber Zulassungs- und Aufsichtsbehörden – national wie international – erfüllen, stehen anspruchsvolle Aufgaben für die Biometrie an; genannt sei nur die Erarbeitung geeigneter Datenschemata, die Klassifikation und Codierung von Befundungen (2) oder die Entwicklung geeigneter Auswertestrategien.

Als weitere Beispiele für studienübergreifende Datenverdichtungen sei die Dokumentation der in der Toxikologie oder Pharmakologie verwendeten Tierstämme genannt. Aus den Daten unbehandelter Tiere werden neben Normalbereichen für Körpergewichte, Organgewichte oder hämatologische und klinisch-chemische Parameter auch Tumorfrequenzen oder Mißbildungsraten ermittelt.

Verdichtete Daten sind nicht allein durch die zugrundeliegenden Rohdaten erklärt. Das Ergebnis einer Berechnung oder statistischen Auswertung hängt auch wesentlich von der Berechnungsmethode, dem statistischen Verfahren und dem zugrundegelegten mathematischen Modell ab. Wird die verdichtete Information in einem Informationssystem bereitgestellt, so muß dem Benutzer zur richtigen Interpretation auch die Information über die Berechnungsmethode zur Verfügung stehen. Dies hat dazu geführt, daß begonnen wird, auch die **Methodenbeschreibungen** der DV zu übergeben. In der Toxikologie, die nach GLP-(Good Laboratory Practice)Regeln arbeitet, werden die Methoden in sogenannten SOP's (»Standard Operating Procedures«) dokumentiert, die in SOP-Bibliotheken verfügbar gehalten werden.

Ziel der Pharmaforschung ist die Entwicklung wirksamer und sicherer Arzneimittel. Deshalb sind alle Aktivitäten der verschiedenen beteiligten Bereiche auf die in diesen Arzneimitteln enthaltenen chemischen Stoffe gerichtet. Dabei interessiert dann nicht mehr, welches Gewicht eine Ratte an einem bestimmten Tag in einem bestimmten Versuch hatte. Von Interesse sind vielmehr verdichtete Informationen über Wirksamkeit und Verträglichkeit. Dazu gehören die ED<sub>50</sub>-Werte (ED = effective dose) aus den pharmakologischen Versuchen, die LD<sub>50</sub>-Werte (LD = lethal dose) akuter Toxizitätsversuche, Angaben über eventuelle Lebertoxizität, Mutagenität, Teratogenität, Kanzerogenität ebenso wie Angaben über Bioverfügbarkeit im menschlichen Organismus usw. All dies sind Größen, die direkt die chemische Substanz charakterisieren. Diese biologischen Größen hält man heute zusammen mit den chemisch-physikalischen Größen einer chemischen Substanz (wie z. B. Schmelz- oder Siedepunkt, pH-Wert, Dampfdruck, Molekulargewicht etc.) in einer Datenbank, der »Stoffdatenbank« (15).

Die biologischen Eingabegrößen in diese Stoffdatenbank kommen mehr oder weniger direkt aus biometrischen Auswertungen der biologischen Prüfungen, die chemisch-physikalischen Größen aus der Analytik. Gemeinsam bilden sie für die Chemometrie die Basis für die Ermittlung von Struktur-Wirkungs-Beziehungen. In folgender Abbildung 2 ist schematisch dargestellt, wie die bei den biologischen Studien gewonnene Information durch Verdichtung in die Stoffdatenbank und von dort durch Ermittlung von Struktur-Wirkungs-Beziehungen wieder für die Synthese neuer Substanzen zurückfließt.

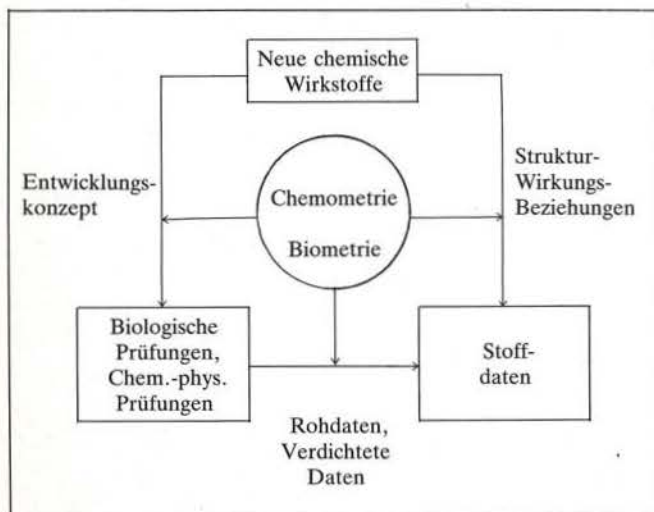


Abb. 2. Informationszyklus in der Pharmaforschung.

Während die Biometrie wesentlich beim Übergang von den biologischen Prüfungen zu den Stoffdaten und die Chemometrie beim Übergang von den Stoffdaten zu neuen chemischen Substanzen unterstützend tätig ist, so ist ihre Beteiligung bei der Aufstellung des Entwicklungskonzeptes für eine neu synthetisierte Substanz von Unternehmen zu Unternehmen unterschiedlich.

Es versteht sich sicherlich von selbst, daß der in Abbildung 2 skizzierte Informationszyklus kein abgeschlossenes System bildet. Zu den firmeneigenen Stoffdaten fließen aus der Literatur oder anderen externen Forschungsstellen weitere Daten ein. Für das Molecular Modelling und die Synthesepaltung werden eigene große Programmsysteme eingesetzt, die den Entwurf der chemischen Struktur und das Auffinden von Synthesewegen unterstützen. Auf der anderen Seite fließen Informationen nach außen. So gehen z. B. die Berichte der biologischen Prüfungen bei erfolgreichen Substanzen über die Registrierungsabteilung zu den Zulassungsbehörden.

Mit der Registrierungsabteilung bestand schon seit jeher eine enge Zusammenarbeit der Biometrie, da jedes Land, in dem die Zulassung des Arzneimittels angestrebt wird, eigene Vorschriften oder Empfehlungen für die Planung, Durchführung, Auswertung, Dokumentation, Präsentation und Archivierung von Prüfungen, insbesondere von klinischen Prüfungen, hat. Über diese Schiene fließen außerdem alle Informationen über den aktuellen Stand der Entwicklung einer Substanz zum Forschungsmanagement. Dieses erhält mit den verdichteten Ergebnissen aus den bis dahin durchgeführten Studien die Basis für die Entscheidung, ob die Entwicklung der Substanz weiter fortgeführt oder wegen mangelnder Wirksamkeit oder zu großer Nebenwirkungen eingestellt werden soll.

## 6. Informatorische Rolle der Biometrie

Ein pharmazeutisches Unternehmen lebt vom Ertrag aus dem Verkauf seiner Arzneimittel. Wenn die Summe von 150 Mio DM, die durchschnittlich für die Entwicklung eines Arzneimittels aufgewandt werden, ausgegeben sind, ist das Produkt noch nicht hergestellt. Die Funktion der Pharmaforschung ist lediglich, Wissen über ein zukünftiges Produkt zu schaffen. Wissen ist nichts anderes als Information. In diesem Sinne

kann man die Tätigkeit der Pharmaforschung als **Produktion von Information** verstehen. Besonders plastisch wird dies, wenn man sich die bis zur Herstellung eines Arzneimittels erarbeitete Information als LKW-Ladung von Ordnern vor Augen führt. Wie bei allen »intelligenten« Produkten, bei denen der Anteil an Wissen und Know-How im Verhältnis zum Herstellungsaufwand sehr hoch ist, spielt bei Arzneimitteln neben den Produktionsfaktoren Arbeit und Kapital der Faktor Information, der sogenannte »dritte Produktionsfaktor«, eine herausgehobene Rolle.

Wir wollen den Gedanken von Forschung als Produktion von Information weiterführen und einige Parallelen zur Produktion von Gütern ziehen.

Bei den biologischen Prüfungen können wir folgende Informationsobjekte unterscheiden:

- Versuchspläne
- Rohdaten
- Verdichtete Daten (Ergebnisse der Auswertung von Rohdaten)
- Darstellung von Ergebnissen (Tabellen, Grafiken)
- Interpretationen, Bewertungen
- Berichte

Die Basis des von der Pharmaforschung »produzierten« Wissens, die »Rohprodukte«, bilden die Rohdaten, die bei den biologischen Prüfungen in den verschiedenen Fachbereichen Pharmakologie, Toxikologie und Klinische Forschung sowie der Galenik erzeugt werden.

Die Rohdaten sind für das zu erzielende Ergebnis wenig informativ. Erst durch geeignete Zusammenfassung und Auswertung werden brauchbare Aussagen daraus. Diese Funktion der Biometrie, Rohdaten zu höherwertigen Informationsprodukten weiterzuverarbeiten, kann man somit als eine »Veredlung« von Rohinformation bezeichnen.

Damit das neue Informationsprodukt für den Gebrauch – das ist das inhaltliche Verstehen der Information – handlich ist, ist es in eine »zweckdienliche Form« zu bringen. Diese Funktion wird von der Biometrie durch die Präsentation der Ergebnisse in Form von Tabellen und Grafiken ausgeübt.

Bei der Planung der biologischen Prüfungen nimmt die Biometrie eine weitere informatorische Funktion wahr. Durch Erarbeitung des Studiendesigns wird festgelegt, wieviel Rohinformation von welcher Art für ein angestrebtes veredeltes Informationsprodukt gewonnen werden soll. Dadurch wird erreicht, daß die Rohprodukte in geeigneter Qualität und Menge hergestellt werden, damit sie für die anschließende Weiterverarbeitung brauchbar und ausreichend sind, aber andererseits kein unnötiger Abfall übrig bleibt.

Durch die Plausibilitätsprüfungen der Rohdaten (»Validierung«) erfüllt die Biometrie schließlich noch eine Teilaufgabe der »Qualitätskontrolle der Rohprodukte«.

Die informatorische Funktion der Registrierungsabteilung entspricht dem »Vertrieb« im Herstellungsbereich, da die produzierte Information entsprechend den Anforderungen der Behörden (das sind hier die »Kunden«) aufbereitet wird. Die Zusammenarbeit der Biometrie mit der Registrierungsabteilung entspricht somit einer »Vertriebsunterstützung«.

In folgender Abbildung 3 sind die Parallelen der Forschung mit der Herstellung noch einmal stichwortartig zusammengestellt. Anhand der im vorausgegangenen Abschnitt aufgezeigten informatorischen Funktionen ließen sich weitere Parallelen zum Herstellungsbereich ziehen. Die aufgezeigten Parallelen zeigen jedoch bereits, daß die Biometrie in der Pharmaforschung unter dem Gesichtspunkt der Produktion von Information eine zentrale Stellung einnimmt.



Pharmaforschung	Herstellung
Information	Produkt
Forschung	Produktion
Rohdaten	Rohprodukte
Auswertung	Veredlung
Präsentation	Formgebung
Studienplanung	Produktionsplanung
Validierung	Qualitätskontrolle
Registrierung	Vertrieb

Abb. 3. Parallelen von Forschung als Produktion von Information und Herstellung von Gütern

**7. Diskussion**

In den Abschnitten 3 und 4 wurde umrissen, welchen Nutzen moderne Informations- und Kommunikationstechnologie der Biometrie gebracht haben. Dabei könnte leicht der Eindruck entstanden sein, als sei der beschriebene Fortschritt bereits überall verwirklicht. Die Realität sieht jedoch viel nüchterner aus.

Jede der genannten Entwicklungen ist zwar bereits in einigen pharmazeutischen Unternehmen realisiert, und es gibt eine Reihe von Unternehmen, die sich in breitem Umfang die neuen Technologien zunutze gemacht haben. Es sind aber keineswegs überall alle neuen Möglichkeiten realisiert. Es handelt sich vielmehr um einen Entwicklungsprozeß, der heute im Gange ist und sich über die nächsten Jahre erstrecken wird. Es wird wohl noch 5 oder gar 10 Jahre dauern, bis die genannten Fortschritte allgemein erreicht sein werden.

Einer zweiten möglichen Fehlinterpretation der vorliegenden Ausführungen sei vorgebeugt. Bei der Darstellung der Funktionen der Biometrie aus informatorischer Sicht (Abschnitte 5 und 6) wurde aufgezeigt, daß die Biometrie in praktisch alle Abläufe der Informationsproduktion in der Pharmaforschung eingebunden ist und dabei wichtige Funktionen erfüllt. In diese Funktionen ist die Biometrie – wie der Rückblick zeigt – auf natürliche Weise, ohne sich darum zu drängen, hineingewachsen, weil sie sich bereits von Anfang an mit DV befassen mußte und weil für die speziellen Anforderungen des wissenschaftlichen Bereiches in den seltensten Fällen anderes Personal zur Verfügung stand.

Es stellt sich nun die Frage: Muß die Biometrie auch in Zukunft neben ihren zentralen Hauptaufgaben auch die verschiedenen Aufgaben, die mehr in das Fachgebiet der Informatik hineinragen, selbst ausüben, oder kann sie von einem Teil der rein technischen Aufgaben entlastet werden? In Frage kommen die Tätigkeitsbereiche

- Software-Entwicklung
- Datenadministration
- Rechnerbetreuung, Systembetreuung (z. B. LIMS-Systeme)

Hier wird in einigen Unternehmen eine gewisse Arbeitsteilung praktiziert. Auch wenn es derzeit noch keine einheitliche Meinung gibt, wie eine optimale Aufteilung aussehen könnte, so lassen sich dennoch einige Tendenzen erkennen.

Softwaresysteme werden in Zukunft nur noch ingenieurmäßig nach standardisierten Regeln (»Software-Engineering«) entwickelt werden (s. auch (9)). Es zeichnet sich ab, daß große Auswertungssysteme, wie sie in der Vergangenheit von den Biometrieabteilungen selbst entwickelt wurden, entweder durch gekaufte Systeme (z. B. SAS) ersetzt oder von Software-Entwicklungs-Abteilungen erstellt werden. Anders sieht

es aber mit der Programmierung spezieller statistischer Verfahren aus, die noch nicht in käuflichen Auswertungssystemen vorhanden sind. Diese Arbeit wird kaum jemand anderes als der über numerische Kenntnisse verfügende Biometriker leisten können.

Die Einrichtung und Betreuung von Auswertungssystemen in der biometrischen Anwendungsumgebung sowie die »Ausstattung« der gekauften Systeme und ihr Ausbau zum eigenen Auswertungssystem wird ebenfalls kaum von einem Außenstehenden befriedigend erfüllt werden können.

Beim Tätigkeitsbereich der Datenadministration ist wohl die vorherrschende Meinung, daß die Strukturierung und Verwaltung der Daten der biologischen Prüfungen bei der Biometrie liegen sollte, da diese schließlich die Weiterverarbeitung der Daten zu den Studienergebnissen vorzunehmen hat und sie außerdem auch die Anforderungen der Fachwissenschaftler bei der Planung und Durchführung der Studien kennt.

Anders sieht es bei den Stoffdaten aus. Diese haben für das Unternehmen einen ausgesprochen zentralen Charakter, der weit über die Pharmaforschung hinausreicht. Entsprechend sollten sie übergreifend modelliert und administriert werden. Es muß aber darauf hingewiesen werden, daß diese Modellierung nicht ohne Einbeziehung der Biometrie, einem der intensivsten Benutzer, geschehen darf, da sonst weder die Pflege der Stoffdaten noch deren Nutzbarkeit für chemometrische Analysen gesichert ist. (Über analoge Erfahrungen bei klinischen Registern wird in (3) berichtet.)

Für die Betreuung von Rechnersystemen einschließlich Anwendungssystemen (z. B. LIMS-Systemen) sind Spezialkenntnisse nötig, die meist in Kursen vermittelt werden. Bei den heutigen Ansprüchen der Benutzer an Betreuung und Komfort kann Systembetreuung nicht mehr »nebenbei« miterledigt werden. Üblicherweise wird eigens dazu ausgebildetes Personal eingesetzt. Ob dieses Personal organisatorisch der Biometrie unterstellt ist oder nicht, ist im Grunde unerheblich. Es muß aber gewährleistet sein, daß die Arbeit in enger Kooperation geschieht und die Vorgaben der Biometrie berücksichtigt werden.

Insgesamt kann man wohl festhalten:

- Soweit bestimmte DV-technische Aufgaben im Pharmaforschungsbereich nicht von der Biometrie selbst übernommen werden, muß gewährleistet sein, daß die Biometrie ihre Anforderungen als wichtiger Systemnutzer einbringen und durchsetzen kann.
- Werden DV-technische Aufgaben von der Biometrie selbst übernommen, so muß dies professionell geschehen, d. h. die notwendige Qualifikation (z. B. für Datendesign oder Software-Engineering) muß erworben bzw. erweitert werden.

Es wurde versucht, die Funktionen der Biometrie innerhalb der Pharmaforschung unter dem Gesichtspunkt der Informationsproduktion zu sehen. Dabei zeigte es sich, daß die Biometrie dort einen zentralen Platz einnimmt. Was dies für Konsequenzen für ihre Stellung im Unternehmen haben wird, darüber kann man derzeit nur spekulieren.

Das Verständnis von Information als Produktionsfaktor ist neu, und die Zukunft wird zeigen, inwieweit und wie schnell sich dieses Verständnis tatsächlich durchsetzen wird. Zu beobachten ist aber bereits heute, daß die klassische Datenverarbeitung abgelöst wird und in einem größeren Rahmen des Informationswesens im Unternehmen aufgehen wird. Dabei werden sich nicht nur – wie bisher – die Arbeitsweisen, sondern auch Organisationsabläufe und Organisationsstrukturen ändern. Wegen ihrer sachlich begründeten zentralen

Funktion innerhalb des Informationswesens der Pharmaforschung besteht die Chance, daß die Biometrie im Laufe dieses Prozesses die ihrer Funktion gemäße Bedeutung und Stellung erlangen wird.

## Literatur

- (1) AG »Biometrie in der chemisch-pharmazeutischen Forschung« in der Deutschen Region der Internationalen Biometrischen Gesellschaft (1983): Selbstverständnis des Biometrikers in der chemisch-pharmazeutischen Industrie. In: VOLLMAR, J. (Hrsg.): Biometrie in der chemisch-pharmazeutischen Industrie I, Gustav Fischer Verlag, Stuttgart und New York, 3–6.
- (2) BLOMER, R., H. NOWAK, J. VOLLMAR (1986): DV-gerechte Codierung von Daten aus klinischen Studien – Technik und Schlüsselssysteme. In: VOLLMAR, J. (Hrsg.): Biometrie in der chemisch-pharmazeutischen Industrie III, Gustav Fischer Verlag, Stuttgart und New York, 161–183.
- (3) GIANI, G., M. MEYERS (1987): Verwaltung und Auswertung von klinischen Registern. *EDV in Medizin und Biologie* **18**, 84–89.
- (4) HARSTON, J., E. PINTER (1979): Good Laboratory Practice Regulations der Food and Drug Administration, USA. *Pharm. Ind.* **41**, 259–268.
- (5) HAUX, R. (Hrsg.) (1986): *Expert Systems in Statistics*. Gustav Fischer Verlag, Stuttgart und New York.
- (6) HAUX, R., K.-H. JÖCKEL (1986): Database management and statistical data analysis: The need for integration and for becoming more intelligent. In: DE ANTONI, F., N. LAURO, A. RIZZI (eds.): *COMPSTAT, Proceedings in Computational Statistics, 7th Symposium held at Rome 1986*, Physika-Verlag, Heidelberg und Wien, 407–414.
- (7) HEDGES, L. V., I. OLKIN (1985): *Statistical Methods for Meta-Analysis*. Academic Press.
- (8) HÖBEL, W. (1987): Zusammenfassen von Ergebnissen klinischer Studien. Vortrag auf dem 33. Biometrischen Kolloquium der Deutschen Region der Internationalen Biometrischen Gesellschaft.
- (9) PASSING, H. (1987): Was bedeuten Software-Qualitätssicherung und -Validierung für den Biometriker? *EDV in Medizin und Biologie* **18**, 84–89.
- (10) SCHLAGETER, G., W. STUCKY (1983): *Datenbanksysteme: Konzepte und Modelle*. 2. Aufl., Teubner-Verlag, Stuttgart.
- (11) SELBMANN, H. K. (1986): Aktuelle Herausforderungen der Medizinischen Informationsverarbeitung. In: EHLERS, C. T., H. BELAND (Hrsg.): *Perspektiven der Informationsverarbeitung in der Medizin. Kritische Synopse der Nutzung der Informatik in der Medizin*. 31. Jahrestagung der GMDS, Göttingen, 1986, *Proceedings*, Springer-Verlag, Berlin, 3–6.
- (12) STEPHENS, M. D. B. (1985): *The detection of new adverse drug reactions*. Stockton Press, New York.
- (13) STREITBERG, B. (1987): Expertensysteme in der Biometrie – Experten oder Ignoranten? *EDV in Medizin und Biologie* **18**, 90–94.
- (14) VICTOR, N., R. HOLLE (1986): Database assisted conduct of clinical trials. In: DE ANTONI, F., N. LAURO, A. RIZZI (eds.): *COMPSTAT, Proceedings in Computational Statistics, 7th Symposium held at Rome 1986*, Physika-Verlag, Heidelberg und Wien, 368–377.
- (15) WOLF, T., H. D. UNKELBACH (1986): *Informationsmanagement in Chemie und Pharma*. Teubner Verlag, Stuttgart.

Eingegangen am: 15. April 1987.

Anschrift des Verfassers: Dr. Hans Dieter Unkelbach, Firma E. Merck, TWI/BD, Postfach 4119, D-6100 Darmstadt.

EDV in Medizin und Biologie 18 (2-3), 75-83, ISSN 0300-8282  
 © Verlag Eugen Ulmer GmbH & Co., Stuttgart; Gustav Fischer Verlag KG, Stuttgart

## Verwaltung und Auswertung von klinischen Registern

G. Giani und W. Meyers

### Zusammenfassung

Der zunehmende Einsatz von Rechnern bei der Bewältigung der vielfältigen Aufgaben im Bereich medizinischer Forschung und ihr wachsender Einfluß auf die medizinische Wissensbildung ist offensichtlich. Speziell Register, die detaillierter als je zuvor in großen Mengen sequentielle klinische Information auf Abruf bereitstellen, versprechen tiefere Einsichten in die Ätiologie der Krankheiten und besonders deren effiziente Behandlungen. Dennoch muß die Registeranalyse als methodisches Instrument zum Therapievergleich deutlich von der kontrollierten klinischen Studie abgegrenzt werden. Auch jede noch so sorgfältig durchgeführte retrospektive Stratifikation nach bekannten prognostischen Faktoren kann das wesentliche Element einer kontrollierten klinischen Studie, die Randomisation, nicht ersetzen. Stichworte wie vermengte Effekte und im besonderen die Patientenselektion sind hier anzuführen. Register lassen sich andererseits in explorativer Weise nutzen, um Krankheitsstadien zu definieren, prognostische Faktoren für Krankheitsverläufe zu ermitteln und a-priori Information zur Vorbereitung klinischer Studien zu gewinnen. Diese Aufgaben können sie nur dann effizient erfüllen, wenn sie nach den gleichen strengen Kriterien wie eine kontrollierte klinische Studie geplant und fortgeschrieben werden.

Unabdingbare Voraussetzung dafür ist zudem eine strukturierte Datenhaltung, die größtmögliche Flexibilität bei der statistischen Auswertung gewährleistet. Diese kann durch eine Rechnerkonfiguration mit einer Datenbankmaschine als Kernstück realisiert werden. Nicht zuletzt hängt die Benutzerfreundlichkeit von der Wahl eines geeigneten Datenmodells ab. Wegen seiner Einfachheit und weitgehenden physischen Datenunabhängigkeit wurde für das Aachener Harnwegstumregister (RUTTAC) das relationale Datenmodell präferiert.

Schlüsselwörter: Klinisches Register, rechnergestützte Datenhaltung und -analyse

### Summary

It is clearly distinguished between a clinical registry and a controlled clinical trial both being considered as tools for treatment comparisons. Any retrospective stratification according to known prognostic factors, however carefully it is carried out, cannot replace the essential feature of a controlled clinical trial, the randomization. The main points for this fact are patient selection and confounding factors. On the other hand registries can be efficiently used in an explorative manner to define disease states, to search for prognostic factors determin-

ing the course of a disease and to gain a-priori information for planning clinical trials. For these tasks a well-structured data management and storage is needed. This can be realized by a computer configuration with a database machine as the central element. Last not least the user-friendliness depends on the choice of an appropriate data model. Because of its simplicity and easy to use logical view on the data, the relational data model was preferred for the registry for urinary tract tumors at the Technical University of Aachen.

Keywords: Clinical Registry, computer-supported data storage and analysis

### 1. Einleitung

Den Einsatz von Rechnern bei der Bewältigung der vielfältigen Aufgaben im Bereich medizinischer Forschung und ihren wachsenden Einfluß auf die medizinische Wissensbildung zu bestreiten hieße, den Kopf in den Sand zu stecken. Speziell Register, die detaillierter als je zuvor in großen Mengen sequentielle klinische Information auf Abruf bereitstellen, versprechen tiefere Einsichten in die Ätiologie der Krankheiten und besonders deren effiziente Behandlungen. Zudem ist diese Art der Erkenntnisgewinnung kostengünstig, kaum zeitaufwendig und bereitet vor allen Dingen keine ethischen Probleme. Kann sie aber deshalb das bewährte Instrument zum Vergleich von Therapiemethoden, die kontrollierte klinische Studie (KKS), verdrängen, die sich ja geradezu dadurch auszeichnet, daß sie letztgenannte Eigenschaften nicht besitzt? Sie ist in den wenigsten Fällen besonders kostengünstig, besitzt z. B. im Bereich der chronischen Erkrankungen lange Follow-up-Phasen und bereitet nicht selten schon bei der Randomisation ethische Probleme. Anders als bei der KKS können zudem aus klinischen Registern gefolgerte Ergebnisse auf größere Fallzahlen gestützt werden bei gleichzeitig nicht so stark durch Ein- und Ausschlusskriterien eingeschränkten Studienpopulationen. Es ist nicht zu leugnen, daß von klinischen Registern angesichts dieser Argumente eine gewisse Faszination ausgeht und es viele Befürworter dafür gibt, die anstehenden medizinischen Probleme schwerpunktmäßig über Register zu lösen (STARMER et al., 1974, ROSATI et al., 1975, FRIES, 1976, BLUM & WIEDERHOLD, 1978).

Jeder weiß, daß die Randomisation, das ist die streng zufällige Zuteilung der Patienten zu verschiedenen Behandlungen, das wesentliche Unterscheidungsmerkmal zwischen einer KKS und einer Registeranalyse mit gleicher Zielsetzung ist. Es ist sogar das einzige, wenn das Register nach den gleichen strengen Kriterien angelegt und fortgeschrieben wird, nach denen

auch eine KKS geplant und durchgeführt wird. DAMBROSIA & ELLENBERG (1980) verwerfen suspekt Datensammlungen mit ungerichteter Zielsetzung, indem sie fordern:

»To ensure a maximum degree of data comparability, collaborators must agree upon, and insure the use of, a uniform vocabulary, common definitions and common patient-exclusion criteria. A protocol should be written, not to interfere in the process of patient management in the manner of a controlled clinical trial, but to make sure that, if certain observations are made, the mechanisms that generate the observations are technically the same both within and between centers and services.«

Noch deutlicher rücken HLATKY et al. (1984) die Registeranalyse in die Nähe der KKS:

»The methods used in the modern observational database and in a well-conducted clinical trial are similar, except for the means of selecting therapy. Both take care in defining terms, collecting data prospectively and completely, managing data in computerized system, following all patients and employ multidisciplinary research teams.«

Wenn sich aber die KKS nur durch die Randomisation von einem prospektiven Register absetzt, ist es um so wichtiger klarzustellen, welche Bedeutung der randomisierten Therapiezuweisung bei der Interpretation der Ergebnisse zukommt. Ganz zwangsläufig werden dabei auch die Grenzen des Einsatzes von klinischen Registern deutlich bzw. diejenigen Teilbereiche abgesteckt, zu denen sie sinnvolle Beiträge leisten können.

In den Abschnitten 2 bis 4 werden deshalb die bei der Auswertung klinischer Register zu berücksichtigenden statistischen Aspekte näher beleuchtet, bevor im fünften Abschnitt auf die Realisation eines Registers, insbesondere des RUT-TAC (Registry for Urinary Tract Tumors, Aachen) eingegangen wird.

## 2. Der Stellenwert der Randomisation

Vergleichende Therapiestudien haben zum Ziel, zwei (oder mehrere) für eine bestimmte Indikation konzipierte Therapien anhand festgelegter, für eine Erfolgshandlung relevanter Zielgrößen zu beurteilen. Das Ergebnis hat immer »statistischen« Charakter. Es kann nicht dazu verwendet werden, für den Einzelfall ganz konkret das Ausmaß des Erfolges unter einer Therapie vorherzusagen und somit immer die richtige Wahl zugunsten der besseren Behandlung zu ermöglichen. Neben der Therapie wirken nämlich eine Fülle von weiteren, nicht kontrollierbaren Einflußgrößen auf das Zielkriterium ein und bedingen die beobachtbare Unterschiedlichkeit der Ausprägungen des Zielkriteriums. Diese Variabilität wird modellmäßig durch eine Wahrscheinlichkeitsverteilung beschrieben. Beim Vergleich zweier Therapien sind es genauer gesagt zwei Wahrscheinlichkeitsgesetze, die einander gegenübergestellt werden und deren Grad an Verschiedenartigkeit die Größe des Effektunterschiedes der Behandlung bestimmt. Eine mögliche Abweichung der Verteilungen voneinander schlägt sich quantitativ in der Abweichung gewisser sie charakterisierender Kenngrößen nieder. Diese sind häufig Erwartungswerte, also beispielsweise die erwartete Überlebenszeit oder der erwartete Blutdruckabfall. Aussagen zur Überlegenheit einer Therapie können demnach auch nur solche Kenngrößen

betreffen und damit unter Bezug auf vorstehende Beispiele nur eine erwartete Verlängerung der Überlebenszeit oder einen erwarteten stärkeren Blutdruckabfall unter der besseren Therapie prognostizieren. Dies ist der oben angesprochene »statistische« Charakter von Studienergebnissen.

Information über die Wahrscheinlichkeitsverteilungen oder deren Kenngrößen liefern »Zufallsstichproben« aus der durch Ein- und Ausschlußkriterien im Studienprotokoll genau beschriebenen Grundgesamtheit von behandlungsbedürftigen Patienten. Der Begriff der »Zufallsstichprobe« bildet das Fundament jeglicher statistischer Inferenz und ist gleichzeitig auch eine der umstrittensten Vokabeln bei der Auslegung von Studienresultaten. Von einer Zufallsstichprobe spricht man, wenn jedes Individuum der Grundgesamtheit die gleiche Chance hat, in die Stichprobe aufgenommen zu werden. Nur über eine Zufallsstichprobe lassen sich bekanntlich Verteilungen und daraus abgeleitete Kenngrößen interessierender Merkmale verzerrungsfrei schätzen. Um zum Zwecke des Therapievergleichs die Verteilung des Zielkriteriums unter jeder der konkurrierenden Behandlungen schätzen zu können, geht man zweistufig vor. In der ersten Stufe versucht man, die interessierende Grundgesamtheit durch eine Zufallsstichprobe repräsentativ zu erfassen. Dann ordnet man in der zweiten Stufe diesen Patienten »ihre« Behandlung zu.

Dies geschieht in der Praxis durch Einbringung eines Zufallsmechanismus, der Randomisation. Ihre Bedeutung bei jeder Art von konfirmatorischer Schlußweise macht man sich am besten an folgender elementarer Überlegung klar.

In der Menge  $I := \{1, \dots, N\}$  seien die Patienten der betrachteten Grundgesamtheit in durch natürliche Zahlen kodierter Form zusammengefaßt. Mit der Wahrscheinlichkeit

$$Q(i) := \frac{1}{N}$$

wird Patient  $i$  in die Zufallsstichprobe aufgenommen. Beim Vergleich von o. B. d. A. nur zwei Therapien A und B werden im nächsten Schritt dem Patienten  $i$  mit der Wahrscheinlichkeit  $Q(A|i)$  die Behandlung A bzw. mit der Wahrscheinlichkeit  $Q(B|i) = 1 - Q(A|i)$  die Behandlung B zugewiesen. Durch

$$(1.1) \quad P(\{(i,c)\}) := \frac{1}{N} Q(c|i)$$

ist dann eine gemeinsame Verteilung über der Menge  $\Omega := \{(i,c) : i \in I, c \in \{A,B\}\}$  erklärt. Die Auswirkungen einer Therapie lassen sich nun durch eine Abbildung  $X: \Omega \rightarrow \mathcal{X}$  beschreiben, wobei  $X(i,c) \in \mathcal{X}$  als der am Patienten  $i$  unter Therapie  $c \in \{A,B\}$  erzielte Erfolg auf der Skala des Zielkriteriums interpretiert wird.

Wie oben schon angesprochen, betrifft die statistische Schlußweise den Vergleich der Verteilungen  $f_A: \mathcal{X} \rightarrow \mathbb{R}$  und  $f_B: \mathcal{X} \rightarrow \mathbb{R}$  des Zielkriteriums unter den beiden Therapien A und B. Diese berechnen sich aus (1.1) als bedingte Verteilungen der Merkmalswerte bei jeweils vorgegebener Therapie A oder B. Man erhält für jedes  $c \in \{A,B\}$  mit der Abkürzung  $E_c(x) := \{i \in I: X(i,c) = x\}$

$$(1.2) \quad f_c(x) = \frac{P(E_c(x) \times \{c\})}{P(I \times \{c\})} = \frac{\sum_{i \in E_c(x)} Q(c|i)}{\sum_{i \in I} Q(c|i)}$$

Die Verteilungen  $f_A$  und  $f_B$  hängen offensichtlich von der speziellen Wahl der Zuordnungsregel  $Q(A|i)$  und somit von subjektiven Vorgaben ab. Wählt man jedoch

$$(1.3) \quad Q(A|i) = \text{const. } (\neq 0),$$

so wird

$$(1.4) \quad f_c(x) = \frac{1}{N} \text{ card } E_c(x), \quad c \in \{A, B\}.$$

In dieser Darstellung sind  $f_A$  und  $f_B$  die eigentlich interessierenden, unverfälschten Verteilungen der Erfolgswerte unter beiden Therapien, deren Vergleich dann mit inferenzstatistischen Methoden angegangen werden kann. Häufig lassen sie sich noch wegen der meist großen zugrundeliegenden Grundgesamtheit durch Dichten der Normal- oder Binomialverteilung oder, falls das Zielkriterium die Überlebenszeit ist, durch Dichten der Exponential- oder Weibullverteilung approximieren. Ist jedoch die Bedingung (1.3) nicht gewährleistet und damit die Zuordnungsregel patientenabhängig, so sind die Verteilungen gemäß vorstehender Ausführungen verfälscht, und jegliche statistische Schlußweise muß aufgrund unzutreffender Modellannahmen kritisch überdacht werden. Die Bedingung (1.3) oder in üblicher Terminologie die (einfache) Randomisation ist also von zentraler Bedeutung. Sie sichert über eine unvoreingenommene, bewußte oder unbewußte Selektionen ausschließende Patientenzuweisung die Vergleichbarkeit der Therapiegruppen sowie die Reproduzierbarkeit der Ergebnisse. Das Niveau, auf dem sie durchgeführt wird, d. h. ob z. B.  $Q(A|i) = 0.5$  oder  $Q(A|i) = 0.75$  gewählt wird, beeinflußt nicht die Validität der aufgestellten Modellannahmen, sondern nur z. B. bei der Durchführung von Tests deren Trennschärfe. Der Verlust an Trennschärfe ist im allgemeinen um so größer, je ungleicher die zufällig erzielten Stichprobenumfänge der Behandlungsgruppen sind.

Um solche Defekte der Balance zu vermeiden, werden in der Literatur verschiedene Modifikationen der einfachen Randomisation vorgeschlagen, u. a. (vgl. KALISH & BEGG, 1984) das Verfahren der »Random permuted Blocks«, die »Minimization«-Methode (vgl. WHITE & FREEDMAN, 1978, POCKOCK & SIMON, 1975) oder die »Biased coin«-Methode von EFRON (1971). Mit einer ähnlichen Argumentation wie oben kann auch für diese Randomisationsverfahren der Nachweis der Unverfälschtheit der interessierenden Verteilungen erbracht werden.

Die fehlende Randomisation ist das wesentlichste Unterscheidungsmerkmal klinischer Register von kontrollierten klinischen Studien. Patienten, die in ein Register aufgenommen und prospektiv dokumentiert werden, erfahren eine individuelle, d. h. von den persönlichen Erfahrungen und Erkenntnissen des behandelnden Arztes oder der behandelnden Institution abhängige Therapie, die nicht selten noch einem Zeitrend unterworfen ist. Die Indikationen bei ein und demselben Krankheitsbild variieren beträchtlich. Will man die gespeicherte Registerinformation für einen retrospektiven Therapievergleich nutzen, ist große Vorsicht geboten. Es ist nicht auszuschließen, daß unbewußt gewisse Patientencharakteristika vorzugsweise die Applikation einer ganz bestimmten Therapie determinierten, während die Vergleichstherapie weniger häufig eingesetzt wurde. Hinsichtlich anderer Charakteristika mag sich die Situation dann wieder genau umgekehrt darstellen. Auf keinen Fall kann jedoch von einer patientenunabhängigen Zuordnungsregel der Form (1.3) ausgegangen werden; die entsprechenden Verteilungen der Zielkriterien sind gemäß (1.2) verfälscht. Durch unterschiedliche Patientencharakteristika bedingte Effekte auf die Zielgröße werden mit Therapieeffekten in unkontrollierbarer Weise vermengt. JESDINSKY (1977) und GREEN & BYAR (1980) betonen die Gefahr dadurch bedingter Fehlschlüsse, indem sie ein Register für Schilddrüsenkarzinome anführen, das der Strahlentherapie

beim Behandlungsvergleich mit nichtbestrahlten Patienten offensichtlich einen lebensverkürzenden Effekt zuwies. Weitere Analysen zeigten, daß mit größerer Priorität Patienten mit schlechter Prognose strahlentherapiert worden waren. Ob der beobachtete Negativeffekt allerdings diesem Umstand zuschreiben war oder tatsächlich allein mit der Strahlentherapie begründet werden mußte, konnte letztlich über das Register nicht beantwortet werden.

### 3. Kontrollgruppen aus Registern

Es braucht nicht besonders betont zu werden, daß der Wirksamkeitsnachweis einer Therapie nur durch Vergleich mit einer Referenzbehandlung erzielt werden kann. Zur Abschätzung des Einflusses muß der Prüftherapie eine Kontrollgruppe vergleichend gegenübergestellt werden. Dabei ist es zugegebenermaßen sehr verführerisch, die Kontrollgruppe retrospektiv einem geeigneten klinischen Register zu entnehmen. In der Literatur spricht man dann von einer historischen Kontrolle. Vordergründig sind die Vorteile einer solchen Vorgehensweise nicht zu übersehen. Zum einen läßt sich der Stichprobenumfang neu zu rekrutierender Patienten drastisch verringern, ohne gleichzeitig die errechnete Aussagesicherheit der geplanten Tests abzusenken. Verbunden damit ist zwangsläufig eine Verkürzung der Studiendauer, da der prospektive Studienarm nur die mit der Prüftherapie behandelten Patienten betrifft. Dies wiederum bedingt eine schnellere Verfügbarkeit und somit eine größere Aktualität der Studienergebnisse. Auf der anderen Seite werden aber auch ethische Gesichtspunkte für die Verwendung historischer Kontrollen angeführt. Es wird argumentiert, daß sich der Arzt nicht von seiner ethischen Verantwortung dem Patienten gegenüber befreien lassen dürfe, ja sogar verpflichtet sei, diejenige Behandlung vorzunehmen, die im konkreten Fall nach seiner Einschätzung das geringste Risiko in sich birgt und größten Nutzen verspricht (HELLMANN, 1979). Die Randomisation bringe dagegen ein willkürliches Element in das Behandlungsprozedere, sei deshalb methodisch für den Arzt schwer verständlich und trage zu einer verringerten Teilnahmebereitschaft sowohl der Ärzte wie der Patienten bei (GEHAN & FREIREICH, 1974). Die ausschließliche Behandlung mit der Prüftherapie, welche zu Studienbeginn nahezu immer einen Vertrauensvorschuß genieße und deshalb präferiert werde, und der anschließende Vergleich gegen eine historische Kontrolle komme dem ethischen Empfinden der Ärzte mehr entgegen als die randomisierte Therapiezuweisung. Besonders gewichtig seien diese Argumente, wenn sich schon in ersten Pilotstudien eine deutliche Überlegenheit über die herkömmliche Therapie abzeichne (GEHAN & FREIREICH, 1974).

Die ethischen Bedenken gegen eine Randomisation zugunsten einer arztorientierten Therapiezuweisung lassen sich weitgehend durch zwei Feststellungen ausräumen. Zum einen schwankt die Zuverlässigkeit ärztlicher Prognosen. Für den gleichen Patienten gibt es unter Ärzten mit gleicher Fachkompetenz nicht selten erheblich divergierende Meinungen über den relativen Nutzen alternativer Therapien (SHAPIRO, 1977). Zum anderen neigen gerade erfahrene, gut ausgebildete Kräfte nicht selten dazu, die Richtigkeit ihrer Aussagen zu überschätzen (GILBERT, McPEEK, MOSTELLER, 1977).

Dem Argument, große zu erwartende Unterschiede verbieten eine randomisierte Studie, ist entgegenzuhalten, daß zumindest im Tumorbereich und bei vielen chronischen Erkrankungen die auf prognostische Faktoren allein zurückzuführende Variabilität der Ausprägungen des Zielkriteriums

gewöhnlich weit größer ist als der therapeutische Effektunterschied (ZELEN, 1975, SIMON, 1984, SATHER, 1986).

Das Hauptproblem beim Verwenden von historischen Kontrollen liegt in der fehlenden Vergleichbarkeit mit der Behandlungsgruppe. Kontroll- und Therapiegruppe sollten in der Verteilung der wichtigsten prognostischen Faktoren übereinstimmen. Daß dies nicht automatisch durch Verwendung gleicher Ein- und Ausschlußbedingungen gewährleistet wird, und auch bei der Evaluation durchgeführte Adjustierungen noch zu Fehleinschätzungen von Therapieeffekten führen können, zeigen MICCIOLO et al. (1985) am Beispiel des Mammakarzinoms:

Um den Effekt einer adjuvanten Chemotherapie auf die Überlebenszeit nach radikaler bzw. erweitert radikaler Mastektomie zu evaluieren, wurden Patientinnen nach Randomisation entweder nur operiert ( $n = 179$ ) oder nach Operation zusätzlich in 12 Zyklen chemotherapiert ( $n = 207$ ). Zur Auswertung wurde das Cox-Modell (COX, 1972) für Überlebenszeiten mit schrittweiser Reduktion der Anzahl der Regressoren herangezogen. Auf der ersten Stufe wurden die wichtigsten prognostischen Faktoren, deren Wechselwirkung mit dem Behandlungsfaktor wie der Behandlungsfaktor selber berücksichtigt. Da die Autoren einen Faktor nur bei gleichzeitig auch nichtsignifikanten Wechselwirkungen mit der Behandlung aus der Modellgleichung nahmen, konnten auch nichtsignifikante Haupteffekte in das Endmodell gelangen. Tabelle 1 zeigt die Ergebnisse.

Tabelle 1. Einfluß des Behandlungsfaktors und anderer prognostischer Variablen auf die Hazard-Funktion, Ergebnisse der randomisierten Studie

	Regr.- Koeff.	Standard- fehler	P-Wert (zweiseitig)
Behandlung	-0,071	0,050	0,1527
Lymphknoten	0,168	0,050	0,0007
Tumorgröße	0,147	0,051	0,0039
menopausaler Status	0,067	0,050	0,1802
Beh. x meno. Status	0,103	0,050	0,0394

Ein Globaleffekt der adjuvanten Chemotherapie konnte nicht nachgewiesen werden. Deshalb ist die Wechselwirkung des Behandlungsfaktors mit dem menopausalen Status qualitativer Art (vgl. PETO, 1976). Das geschätzte relative Risiko der Chemotherapie im Vergleich zur reinen Operation betrug in der Prämenopause 0.5, während es auf 1.13 für die Patientinnen in der Menopause anstieg. Der zunächst positive Chemotherapieeffekt in der Prämenopause ließ sich für die Patientinnen in der Menopause nicht mehr verifizieren. Um den durch eine historische Kontrollgruppe induzierten Bias zu studieren, ersetzen MICCIOLO et al. (1985) die randomisierte Kontrolle unter Beibehaltung der Ein- und Ausschlußkriterien durch eine Gruppe von Patienten, die mindestens 5 Jahre vor Beginn der randomisierten Studie ohne nachfolgende Chemotherapie mastektomiert worden waren ( $n = 381$ ). Wie zu erwarten war, verteilten sich jetzt die prognostischen Faktoren ungleich auf beide Gruppen. Aber auch über das Cox-Modell waren keine vergleichbaren Ergebnisse zu erzielen, obwohl der Adjustierungseffekt des Regressionsansatzes dies erhoffen ließ (Tab. 2).

Tabelle 2. Einfluß des Behandlungsfaktors und anderer prognostischer Variablen auf die Hazard-Funktion, Ergebnisse nach Ersetzen der randomisierten durch eine historische Kontrolle

	Regr.- Koeff.	Standard- fehler	P-Wert (zweiseitig)
Behandlung	-0,276	0,080	0,0006
Lymphknoten	0,371	0,069	$<10^{-4}$
Tumorgröße	0,223	0,070	0,0014
menopausaler Status	0,519	0,113	$<10^{-4}$
Alter	-0,042	0,011	0,0001

Im Gegensatz zu den Resultaten der kontrollierten Studie ließ sich jetzt ein globaler Behandlungseffekt nachweisen, während die Wechselwirkung mit dem menopausalen Status verschwand. Die Verwendung einer historischen Kontrolle führte offensichtlich zu einem optimistischeren Resultat; auch die Patientinnen in der Menopause schienen jetzt von der Chemotherapie profitiert zu haben. Die Autoren erklären dieses Ergebnis durch einen Zeittrend:

»Clinical knowledge progressively accumulated during the trials appear to imply subtle improvements in the care of patients, tending to influence the survival and bias the comparisons in the historically controlled studies.«

Die Problematik des Zeittrends ist im Schrifttum eingehend diskutiert (BYAR, 1979, BYAR, 1980, GREEN, 1982, GREEN & BYAR, 1984, SCHUMACHER, 1986). Gerade weil er im nachhinein als Verzerrungen induzierender Effekt nur schwer zu identifizieren ist, sind Vergleiche mit historischen Kontrollen vorsichtig zu interpretieren. Um die Größenordnung des Bias abzuschätzen und gegebenenfalls die historische Kontrolle in die Endauswertung mit einzubeziehen, schlagen verschiedene Autoren (POCOCK, 1976, GEHAN & SCHNEIDERMAN, 1982) vor, die Patienten im Verhältnis 2:1 auf die Behandlungs- und Kontrollgruppe zu randomisieren. Auf diese Weise ließe sich unmittelbar durch Vergleich der randomisierten mit der historischen Kontrolle die Eignung letzterer als Ergänzung der Kontrollgruppe überprüfen. Die Zusammenfassung beider Kontrollgruppen hat dann den gleichen Stichprobenumfang wie die Behandlungsgruppe.

Nicht unerwähnt bleiben darf die Technik des Matchens als kontrovers diskutierte Möglichkeit zur Erzeugung von Referenzgruppen für den Therapievergleich (GEHAN & FREIREICH, 1974, DAMBROSIA & ELLENBERG, 1980). Beim Matchen wird jedem Patienten der Testgruppe ein hinsichtlich wesentlicher, prognostisch relevanter Charakteristika ähnlicher, schon früher behandelter Patient z. B. aus dem Register gegenübergestellt. Auf diese Weise soll die Vergleichbarkeit von Test- und historischer Kontrollgruppe erzwungen werden. Die rein technische Seite der Abwicklung des Matchens ist unter Zuhilfenahme eines Rechners heute kein Problem mehr. Der Registervorrat an potentiellen Vergleichspatienten muß nur hinreichend groß sein. Die Erfahrung zeigt, daß man bei drei bis vier zu matchenden Variablen in der Regel an die Grenze des Machbaren stößt. Fraglich ist, ob man mit diesen Variablen wirklich alle prognostisch relevanten Faktoren berücksichtigt hat. Zumindest läßt sich die Kalenderzeit nicht mit einbringen, was möglicherweise der oben erwähnten Verzerrung des Studienergebnisses durch einen Zeittrend Vor-schub leistet.

**4. Identifizierung prognostischer Faktoren aus Registern**

Die Kenntnis prognostischer Faktoren ist in vielerlei Hinsicht von großer Bedeutung.

- a) Bei kontrollierten klinischen Studien liefern prognostische Faktoren die Strata, innerhalb derer eine Randomisierung in etwa gleich große Vergleichsgruppen vorgenommen wird. Dadurch sollen Unbalanciertheiten in den Gruppen hinsichtlich wichtiger prognostischer Faktoren vermieden werden, die möglicherweise durch eine einfache Randomisation nicht hätten ausgeschlossen werden können und so vielleicht die Ursache für vermengte Effekte und im speziellen für Patientenselektion gewesen wären. Besonders wichtig ist die Stratifikation für Studien mit kleinen Fallzahlen. Bei Studien mit großen Patientenaufkommen hat die Stratifikation keinen nennenswerten Vorteil mehr vor der einfachen Randomisation (POCOCK, 1983, BROWN, 1980).
- b) Prognostische Faktoren werden benützt, um den Krankheitsverlauf zu beschreiben und zu verstehen. Dabei sind es gerade die Korrelationen und Wechselwirkungen zwischen ihnen, die von besonderem Interesse sind.
- c) Prognostische Faktoren werden zu Adjustierungszwecken bei der Auswertung klinischer Studien herangezogen. Von essentieller Bedeutung ist gerade dieser Aspekt für Studien mit historischen Kontrollen.
- d) Prognostische Faktoren werden verwendet, um Untergruppen von Patienten zu definieren, die von unterschiedlichen Behandlungen profitieren. Dies geschieht mit der Absicht, für jeden Patienten aufgrund seiner besonderen Konstellation prognostischer Faktoren die für ihn optimale Therapie zu ermitteln. Retrospektive Auswertungen großer Datenbestände wie Register können wertvolle Hinweise in dieser Richtung geben. Es muß jedoch davor gewarnt werden, die Resultate, die sich meistens aus einer Vielzahl von Tests ableiten, überzubewerten. Sowohl die Probleme des Multiplen Testens (SONNEMANN, 1982) wie die allein durch den retrospektiven Charakter der Auswertung bedingten Ergebnisverfälschungen sind im Auge zu behalten (GLASTEIN & MAKUCH, 1984, SATHER, 1985).

Zur Identifizierung prognostischer Faktoren werden multivariate statistische Verfahren herangezogen. Am häufigsten darunter kommen wohl Regressionsansätze wie das Cox-Modell bei Überlebenszeiten und das logistische Modell bei der Beurteilung von Erfolgswahrscheinlichkeiten zur Anwendung. Beide zeichnen sich durch ihre wenig restriktiven Verteilungsannahmen aus. Einige Bemerkungen zur Interpretation der Ergebnisse aus solchen retrospektiven Analysen an Registerdaten sind jedoch angebracht.

Ähnlich wie bei komparativen Therapiestudien mit einer historischen Kontrollgruppe ist Vorsicht geboten bei der Verallgemeinerung der Ergebnisse. Ein illustratives Beispiel dafür liefert das Hodentumorregister der Universität von Minnesota (GOLDMAN et al., 1980). Im Rahmen eines größeren Projektes sollte an einigen wenigen Patienten (n = 78) die Bedeutung und Wertigkeit des Stagings als prognostischer Faktor demonstriert werden. Entgegen aller Erwartung sank die Überlebenswahrscheinlichkeit nicht gleichmäßig mit fortgeschrittenem Stadium der Erkrankung zum Zeitpunkt der Hodenentfernung (Abb. 1). Nach genauerer Betrachtung fand sich eine mögliche Erklärung für dieses überraschende Ergebnis. 16 der Patienten im Stadium I hätten nach dem histologischen Lymphknotenbefund eigentlich dem Stadium II zugeordnet werden müssen. Die Reanalyse zeigte jedoch immer noch

keinerlei nennenswert günstigere Prognose für Stadium-I-Patienten (Abb. 2). Auch diesmal bot sich ein Argument zur Erklärung an. In dem Register waren die außerhalb der Universitätskrankenhäuser behandelten Patienten nicht erfaßt worden. Von diesen konnte ein Großteil in der Klasse der Stadium-I-Patienten vermutet werden, nämlich genau der Patiententeil, der nach der Hodenentfernung beschwerdefrei geblieben war. Vorwiegend waren dies Patienten mit Seminom als Primärtumor, deren Lebenserwartung bekanntlich größer ist als die der Patienten mit anderen Hodenkarzinomtypen. Die Vermengung von Faktoren wie Pathologie und Staging sowie Patientenselektion haben hier offensichtlich zu einer Unterschätzung der Überlebenswahrscheinlichkeit zumindest für Stadium-I-Patienten geführt.

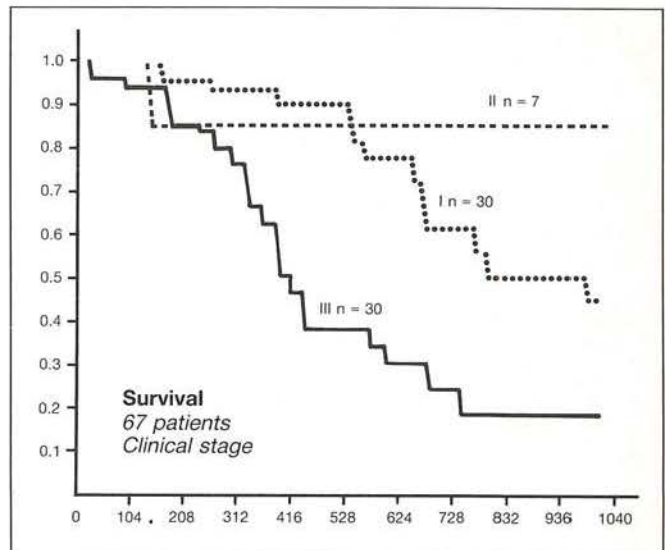
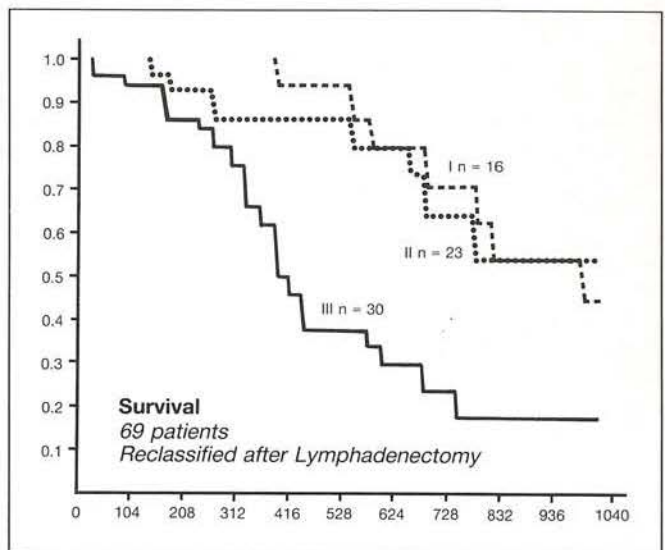


Abb. 1. Nach klinischem Staging stratifizierte Überlebenskurven von 67 Patienten mit Hodentumor (Ordinate: kumulierte Überlebensrate, Abszisse: Tage), Quelle: GOLDMAN et al. (1980)

Abb. 2. Überlebenskurven von 69 Patienten mit Hodentumor nach Reklassifikation aufgrund histologischer Befunde (Ordinate: kumulierte Überlebensrate, Abszisse: Tage), Quelle: GOLDMAN et al. (1980)



Die Wertigkeit prognostischer Faktoren wird häufig überschätzt. »Statistisch signifikante« Faktoren gefunden zu haben, bedeutet nicht in jedem Fall, daß damit auch der Krankheitsverlauf erklärbarer geworden ist. Der Anteil an der durch ein prädictives Modell erklärten residualen Variabilität sollte quantifiziert werden. Für das meistens bei binären Zielvariablen verwendete logistische Regressionsmodell z. B. ist nach EFRON (1978) der negative Logarithmus der Likelihoodfunktion  $L_1$  ein geeignetes Maß für die residuale Variabilität. Bezeichnet  $L_0$  dann die Gesamtvariabilität im Modell ohne jegliche Einflußvariablen, so kann  $L = (L_0 - L_1) / L_0$  als der Anteil an der Gesamtvariabilität interpretiert werden, der sich durch Einbezug prognostischer Faktoren erklären läßt. SIMON (1984) zitiert als Beispiel eine Arbeit von GEORGE & HOOGSTRATEN (1978) zur Identifizierung prognostischer Faktoren beim Mammakarzinom in der neun Kovariablen und deren Wechselwirkungsterme mit zwei Behandlungsfaktoren ein logistisches Regressionsmodell mit 30 zu schätzenden Parametern festlegt. Abgesehen von der geringen Fallzahl ( $n = 242$ ) im Vergleich zur Parameterzahl (overfitting) lassen sich mit diesem Ansatz nur  $L = (160.223 - 126.558) / 160.223 = 21\%$  der beobachteten Gesamtvariabilität der dichotomen Responsevariablen für den Erfolg einer Chemotherapie erklären.

Große Bedeutung kommt als explorative Methode zur Selektion prognostischer Faktoren der schrittweisen Regressionsanalyse zu. Zur Beurteilung der prädictiven Wertigkeit des Endmodells empfehlen HARRELL et al. (1985) eine Partitionierung der Daten in ein Test- und ein Trainingsset. Sie warnen davor, ein Modell mit zu vielen Variablen an zu wenig Patienten mit definierten Endpunkten (wie z. B. Tod) anzupassen und nennen als gerade noch vertretbares Verhältnis die Zahl 1:10. Eine interessante Alternative zu der vorgenannten Art der Modellvalidierung bietet das Bootstrapping. CHEN & GEORGE (1985) demonstrieren diese Methode an Daten von 224 Kindern mit akuter lymphozytischer Leukämie.

## 5. Der Aufbau eines klinischen Registers

Da ein klinisches Register im Gegensatz zu kontrollierten klinischen Studien meist keinen zeitlich eng begrenzten Rahmen besitzt, ist es wichtig, die Daten so zu organisieren, daß sie jederzeit für eine große Klasse von Fragestellungen auf möglichst einfache Art und Weise verfügbar gemacht werden können und größtmögliche Flexibilität bei der Aufbereitung für statistische Auswertungen gewährleistet ist.

Für den Biometriker ist es dabei unerlässlich, sich der Methoden und Werkzeuge der Informatik zu bedienen. Das Fachgebiet, welches die Anforderungen der Statistik im Rahmen der Informatik berücksichtigt, wird heute üblicherweise mit »Computational Statistics« bezeichnet. Nach VICTOR (1984) hat »Computational Statistics« . . . »die Untersuchung von Lösungs- und Erleichterungsmöglichkeiten der Aufgaben des Statistikers durch Einsatz des Rechners und/oder der Methoden der Informatik zum Inhalt«.

Da bei klinischen Registern die Benutzerfreundlichkeit im Hinblick auf den effizienten Einsatz statistischer Methoden von großer Bedeutung ist, muß der Biometriker in der Lage sein, die geeignete Informationstechnologie auszuwählen und sinnvoll einzusetzen. Wichtig ist dabei nicht nur der Einsatz eines einzelnen Rechners, sondern die Aggregation unterschiedlicher Komponenten zu einem Gesamtsystem. Dies können z. B. an einen Datenbankrechner als zentrale Einheit gekoppelte Rechner (PC's Minicomputer etc.) zur Eingabe,

Verwaltung und Auswertung der Daten sowie zur Erstellung von Graphiken sein.

Um die in ihrer Struktur sehr unterschiedlichen Aufgaben in den Teilbereichen Datenerfassung, -verwaltung und -auswertung klar voneinander trennen zu können, empfiehlt es sich, das Gesamtsystem in diese drei Moduln aufzuteilen, die jedoch untereinander kommunizieren müssen. HAWKINS und SINGER (1986) sagen dazu:

»We discovered from the survey that these three components are separate, both logically and operationally, that is, each one could be performed on an entirely different class of computer, if one desired to do so.«

Im folgenden soll kurz aufgezeigt werden, wie die Realisation dieser drei Moduln innerhalb des RUTTAC (Registry for Urinary Tract Tumors, Aachen) erfolgte, ohne dabei den Anspruch auf vollständige Behandlung der zugrundeliegenden Problematik zu erheben.

### 5.1 Die Datenerfassung

Zur Zeit werden Patientendaten, welche in klinischen Registern abgelegt werden sollen, üblicherweise auf Erhebungsbögen eingetragen und über eine Schnittstelle zur Datenbank erfaßt. Voraussetzung dafür ist, daß auf einer konzeptuellen Ebene in enger Zusammenarbeit mit den Klinikern alle inhaltlich relevanten Merkmale erfaßt wurden. Die Vollständigkeit sollte sich dabei an den mit diesen Daten zu beantwortenden Fragen orientieren, so daß sichergestellt ist, daß nur Daten zu solchen Merkmalen erfaßt werden, welche für spätere statistische Auswertungen von Bedeutung sind bzw. die zur Beantwortung der gestellten Fragen nützlich sein können. PROPPE (1960) weist u. a. darauf hin, daß der Erfolg modern konzipierter Dokumentationsmethoden in erster Linie von der ursprünglichen Fragestellung abhängt.

Da ein Hauptanliegen des RUTTAC darin besteht, eine umfassende Dokumentation des Krankheitsverlaufs von Patienten mit Harnwegstumoren mit einem möglichst lückenlosen Follow-up (Kontrolluntersuchungen in 3 Monatsabständen wurden angestrebt) bereitzustellen, wurde seinerzeit folgende inhaltliche Gliederung gewählt.

1. Anamnese
2. Angaben zu Behandlungen zwischen den einzelnen Kontrollen sowie Zustand des Patienten bei Kontrolle
3. Laborwerte
4. Untersuchungsmethoden
5. Metastasen
6. Pathologiebefund
7. Therapien
8. Komplikationen
9. Abschlußdokumentation.

Die Datenerhebung erfolgt auf 5 verschiedenen Bögen, die neben der Bogennummer als identifizierende Merkmale weiterhin enthalten:

- die Patientenidentifikation (I-Zahl)
- die Tumorkennziffer (TK)
- die Klinik-Nummer (KL-NR)
- die Registriernummer (REG)

Bis auf den Namen können alle Merkmale in kodierter Form eingegeben werden, so daß keine Klartextfassung nötig ist. Die Eingabe bzw. Korrektur der Daten erfolgt über Bildschirmmasken, die mit Hilfe eines Maskengenerators erzeugt werden. Die syntaktische und z. T. auch die semantische Integrität der Daten wird bereits während der Eingabe überprüft (Anbindung sogenannter »user action routines« an die Masken).



5.2 Die Datenverwaltung

Sind auf konzeptueller Ebene alle zu erfassenden Merkmale erst einmal festgelegt, so gilt es als nächstes, ein geeignetes Datenmodell zu definieren, welches die logische Struktur der Daten beschreibt. Dazu gehört sowohl die Beschreibung der einzelnen Attribute, ihrer Wertemengen und Beziehungen untereinander, als auch die Spezifikation semantischer Integritätsbedingungen (siehe z. B. SCHLAGETER/STUCKY, 1983).

Existieren bedingt durch eine Vielzahl statistischer Fragestellungen mehrere verschiedene externe Sichten auf die Daten, so ist es von Vorteil, ein Datenbankverwaltungssystem zu benutzen, welches es jedem Nutzer gestattet, durch einfache, überschaubare Manipulationen seine benötigten Daten zu erhalten. Dazu sagt MARTIN (1981) u. a.:

»Das Datenbanksystem ermöglicht diese mehrfachen (externen) Datensichten nicht just um der verschiedenen Sichten willen. Sondern diese Technik erlaubt, daß eine Sicht sich wenn nötig ändern kann, während alle anderen Sichten erhalten bleiben. Nur auf diese Weise kann verhindert werden, daß Änderungen an einer Stelle im System stets die ganze Vielfalt von Datensichten zerstören.«

Aufgrund seiner einfachen Struktur eignet sich zur Bewältigung genannter Anforderungen besonders gut ein relationales Datenbanksystem (CODD, 1970), das i. w. auf der mathematischen Theorie der Relationenalgebra beruht.

Die richtige Auswahl des Datenbanksystems muß allerdings unter vielen Gesichtspunkten erfolgen. Das Datenbanksystem sollte nicht nur an den Vorteilen, die das ihm zugrundeliegende Datenmodell selbst bietet, gemessen werden; vielmehr muß auch die Software beurteilt werden, welche dieses unterstützt. So erwähnen HAWKINS und SINGER (1986) als Güteaspekt u. a.

- Verfügbarkeit auf verschiedenen Rechnertypen,
- relativ leichter Transport einer Datenbank sowie angehängter Prozeduren zu einem anderen Rechner,
- gute Unterstützung von seiten des Anbieters,
- gute Dokumentation.

Ihre Wahl fiel dadurch auf SIR, ein auf dem hierarchischen Datenmodell basierendes Datenbankverwaltungssystem.

Da jedoch bei der Konzipierung des RUTTAC völlig andere Aspekte im Vordergrund standen, kamen die von HAWKINS und SINGER angeführten Argumente weniger zum tragen. Die Entscheidung wurde letztendlich beeinflusst durch die bereits oben erwähnten besonderen Vorteile des relationalen Datenmodells gegenüber dem hierarchischen bzw. dem Netzwerkmodell sowie der Tatsache, daß ein Datenbankrechner zur Verfügung stand, welcher dieses Modell unterstützt. Zu erwähnen bleibt allerdings, daß das Register erst vor kurzem bezüglich dieses Datenmodells neu strukturiert wurde. Vorher wurde es ebenfalls in SIR gehalten.

Für das Register RUTTAC wurde auf dem Datenbankrechner IDM 500, welcher das relationale Datenmodell unterstützt, eine Datenbank eingerichtet. Auf diesem Rechner steht als Datenmanipulationssprache die abbildungsorientierte Sprache IDL zur Verfügung. Der Datenzugriff erfolgt über den Host-Rechner VAX 11/780. Außerdem besteht über das lokale Netz des Klinikums der RWTH Aachen (LocalNet 20 von Sytek) die Möglichkeit, sich des Host-Rechners zu bedienen.

Obwohl die Struktur der Datenbank aufgrund des verwendeten relationalen Datenmodells auf den ersten Blick recht

einfach erscheint, ergaben sich bei der Definition der Relationen doch gewisse Schwierigkeiten als Folge funktionaler Abhängigkeiten zwischen Attributen. Beispielhaft sei hier nur die üblicherweise gestellte Forderung erwähnt, daß Relationen sich in dritter Normalform befinden sollten, um Insert-, Delete- bzw. Update-Anomalien zu vermeiden. Im RUTTAC haben wir jedoch an manchen Stellen bewußt auf diese Forderung verzichtet.

Beispiel: In der Relation ANAMNESE sind u. a. folgende Attribute enthalten:

ANDURO1: andere urologische Erkrankung ('X', '')

ANDUROJ1: Dauer dieser Erkrankung (in Jahren).

Zwischen diesen beiden besteht eine funktionale Abhängigkeit der Form

ANDUROJ1 → ANDURO1

wobei ANDUROJ1 kein Schlüsselkandidat für diese Relation ist und somit eine transitive funktionale Abhängigkeit vorliegt (der fehlende Wert im Attribut ANDUROJ1 sei dabei von der Betrachtung ausgeschlossen). Die Berechtigung, in die Datenbank aufgenommen zu werden, erhält die Variable ANDURO1 jedoch dadurch, daß es Patienten mit weit zurückliegenden urologischen Erkrankungen gibt, deren Zeitpunkte sich, aus welchen Gründen auch immer, jedoch nicht mehr feststellen lassen.

Diese Vorgehensweise ist durchaus gerechtfertigt, wenn mit der üblichen Methode einer Zerlegung von Relationen, die nicht in dritter Normalform sind, die Anzahl der benötigten Relationen zu groß wird. Aus Gründen der Übersichtlichkeit ist hier eine Grenze gesetzt. Bei Verletzung der dritten Normalform muß man andererseits dafür Sorge tragen, daß Insert-, Delete- und Update-Anomalien vermieden sowie semantische Integritätsbedingungen eingehalten werden. Eine Möglichkeit besteht darin, bereits bei der Eingabe Kontrollfunktionen einzubauen (siehe auch KARRISON, 1981). Dies kann jedoch die Benutzerfreundlichkeit des Systems durch z. B. erhöhte Laufzeiten stark einschränken. Eine Alternative wären Kontrollroutinen, die nach größeren vorbestimmten Zeitabständen die semantischen Integritätsbedingungen überprüfen. Dadurch werden jedoch möglicherweise wieder umfangreiche nachträgliche Recherchen notwendig, um bestehende Widersprüche zu beheben.

Im Rahmen der Datenverwaltung sind außerdem Datenschutzaspekte von besonderer Bedeutung. Gerade bei klinischen Registern, wo eine Vielzahl von Personen mit Erfassung, Eingabe und Korrektur der Daten beschäftigt sind, müssen gute Kontrollfunktionen zum Schutze der Daten eingebaut werden. Dies kann zum einen dadurch geschehen, daß dieser Personenkreis so klein wie möglich gehalten wird. Zum anderen können spezielle externe Sichten definiert werden, die keinen Rückschluß auf einzelne Personen zulassen. Im Register RUTTAC wird dies u. a. dadurch erreicht, daß bei der Ausgabe der Daten der Name nicht angezeigt wird. Außerdem ist der Datenzugriff vom Host-Rechner aus durch ein Kodewort geschützt, welches wöchentlich geändert wird.

5.3 Die Auswertung der Daten

In dieser Komponente spielt neben der statistischen Auswertung die Präsentation der Ergebnisse eine besondere Rolle. Aufgrund der Tatsache, daß unter der vorliegenden Rechner-

konstellation für RUTTAC kein integriertes System zur Verfügung steht, welches Verwaltung und Auswertung der Daten sowie Präsentation der Ergebnisse in einem ermöglicht (HAUX & JÖCKEL, 1986), ist es von besonderer Bedeutung, eine leistungsfähige Schnittstelle zu Statistischen Auswertungssystemen sowie zu Graphik-Programmpaketen zu besitzen.

Bei zugrundeliegendem relationalem Datenmodell können mit der Datenmanipulationssprache für spezielle externe Sichten durch Projektionen, Selektionen bzw. Verbunde die benötigten Daten relativ einfach über virtuelle Relationen aus der Datenbank extrahiert werden.

Die Daten für die Auswertungen werden dabei stets in sequentiellen Dateien bereitgestellt, welche durch spezielle Systembefehle leicht aus der Datenbank erzeugt werden können. Diese sind von den gängigen Statistischen Auswertungssystemen leicht zu verarbeiten. Einziger Nachteil bei dieser Vorgehensweise ist, daß die Datenbeschreibung nicht automatisch aus der Datenbank in das Auswertungsprogramm übergehen kann, sondern für jede Anwendung getrennt festgelegt werden muß.

Bisher wurden statistische Auswertungen nur mit SAS durchgeführt. Dabei beschränkten sich diese Auswertungen ausschließlich auf deskriptive Statistik. Die graphische Präsentation erfolgte zum Teil mit SAS, aber auch mit eigener Graphik-Software. Da in der VMS-Installation die Prozedur BMDP von SAS fehlt, ist geplant, eine Schnittstelle zum Statistischen Auswertungssystem BMDP zu schaffen, um insbesondere Cox'sche Regressionsmodelle für Überlebenszeiten einbeziehen zu können.

## 6. Der Stellenwert klinischer Register

Das Hauptziel klinischer Register ist es, zielorientierte Daten einer großen Patientenzahl bereitzustellen, um durch akkumulierte Informationen Erklärungen für unterschiedliche Krankheitsverläufe zu ermöglichen und Interventionsmaßnahmen abschätzen zu können. Dabei muß man sich sowohl der Möglichkeit von Fehleinschätzungen bei der Beurteilung der Wertigkeit prognostischer Faktoren wie der Wirkung von Behandlungsmethoden durch Patientenselektion, vermengte Effekte und Zeittrend bewußt sein. Da jeder Registerpatient eine individuelle, von den Erfahrungen und dem Kenntnisstand der Ärzte geprägte Behandlung erfährt, kann ein Register niemals die Datenquelle für einen objektiven Therapievergleich oder Wirksamkeitsvergleich sein. Register liefern häufig die Ideen für künftige randomisierte Studien. Sie lassen sich bei der Planung kontrollierter klinischer Studien einsetzen, indem sie die prognostischen Faktoren für die Stratifikation und Auswertung liefern und erste Schätzungen für Therapieeffekte vermitteln, an denen sich die Stichprobenkalkulationen orientieren können.

Diese Aufgaben können nur von einem modernen Register mit definierter Zielrichtung wahrgenommen werden, das nach einem vorgegebenen Protokoll nach standardisierten Methoden Befunde und Laborwerte erhebt und Patienten in einem kontrollierten Follow-up weiterverfolgt.

Für die Datenhaltung sowie -auswertung und -präsentation kann die Biometrie sich dabei der Methoden und Werkzeuge der Informatik bedienen. Dazu gehört sowohl die Auswahl einer geeigneten modernen Informationstechnologie (z. B. Datenbankrechner) als auch die Verwendung eines entsprechenden Datenmodells, um Auswertungen insbesondere hinsichtlich statistischer Anforderungen effizient durchführen zu können.

## Literatur

- BLOMER, S. (1980): Vergleich von Datenbankkonzepten und ihre Realisierung in praktisch verwendeten Datenbanksystemen im Hinblick auf medizinische Anwendungen. *Statist. Software Newsl.*, 84 (Suppl.).
- BLUM, R. L. und G. WIEDERHOLD (1978): Inferring knowledge from clinical data banks utilizing techniques from artificial intelligence. *Proceedings of the Second Annual Symposium on Computer Application in Medical Care*, 303-307. New York: IEEE.
- BROW, B. W. (1980): Designing for cancer clinical trials: selection of prognostic factors. *Cancer Treat. Rep.*, 64, 499-502.
- BYAR, D. P. (1979): The necessary and justification of randomized clinical trials. TAGNON, H. J. und M. J. STAGUET (Hrsg.): *Controversies in Cancer*, 75-82. New York: Masson.
- BYAR, D. P. (1980): Why data bases should not replace randomized clinical trials. *Biometrics*, 36, 337-342.
- CHEN, C.-H. und S. L. GEORGE (1985): The bootstrap and identification of prognostic factors via Cox's proportional hazards regression model. *Statist. Med.*, 4, 39-46.
- CODD, E. F. (1970): A relational model of data for large shared data banks. *Commun. ACM*, 13, 377-387.
- COX, D. (1972): Regression models and life tables. *J. R. Statist. Soc.*, B34, 187-220.
- DAMBROSIA, J. M. und J. H. ELLENBERG (1980): Statistical consideration for a medical data base. *Biometrics*, 36, 323-332.
- EFRON, B. (1971): Forcing a sequential experiment to be balanced. *Biometrika*, 78, 403-417.
- EFRON, B. (1978): Regression and ANOVA with zero-one data: measures of residual variation. *J. Amer. Statist. Assoc.*, 73, 113-121.
- FRIES, J. F. (1976): A data bank for the clinician? *N. Engl. J. Med.*, 294, 1400-1402.
- GEHAN, E. A. und E. J. FREIREICH (1974): Non-randomized controls in cancer clinical trials. *N. Engl. J. Med.*, 290, 198-203.
- GEHAN, E. A. und M. A. SCHNEIDERMAN (1982): Experimental design of clinical trials. HOLLAND, J. F. und E. FREI (Hrsg.): *Cancer Medicine*, 2. Aufl., Philadelphia: Lea & Febiger.
- GEORGE, S. L. und B. HOOGLSTRATEN (1978): Prognostic factors in the initial response to therapy by patients with advanced breast cancer. *J. Nat. Cancer Inst.*, 60, 731-736.
- GILBERT, J. P., B. MCPEEK und F. MOSTELLER (1977): Statistics and ethics in surgery and anesthesia. *Science*, 198, 684-689.
- GLASTEIN, E. und R. MAKUCH (1984): Illusion and reality practical pitfalls in interpreting clinical trials. *J. Clin. Oncol.*, 2, 488-497.
- GOLDMAN, A., G. BOSL, K. JOHNSON, E. E. FRALEY und B. J. KENNEDY (1980): Prognostic factors in cancer of the testis: an approach to identification. *Cancer Treat. Rep.*, 64, 487-496.
- GREEN, S. B. und D. P. BYAR (1984): Using observational data from registries to compare treatments: the fallacy of omnimetrics. *Statist. Med.*, 3, 361-370.
- HARREL, F. E., K. L. LEE, D. B. MATCHAR und T. A. REICHERT (1985): Regression models for prognostic prediction: advantages, problems, and suggested solutions. *Cancer Treat. Rep.*, 69, 1071-1077.
- HAUX, R. und K.-H. JÖCKEL (1986): Database management and statistical data analysis. The need for integration and for becoming more intelligent. De ANTONI, F., N. LAURO und A. RIZZI (1986) (Hrsg.) *COMPSTAT 1986*, 407-414.
- HAWKINS, B. S. und S. W. SINGER (1986): Design, development, and implementation of a data processing system for multiple controlled trials and epidemiologic studies. *Controlled Clinical Trials*, 7, 89-117.
- HELLMANN, S. (1979): Editorial: Randomized clinical trials and the doctor-patient relationship. *Cancer Clin. Trials*, 2, 189-193.
- HLATKY, M. A., K. L. LEE, F. E. HARRELL, R. M. CALIFF, D. B. PRYOR, D. B. MARK und R. A. ROSATI (1984): Tying clinical research to patient care by use of an observational database. *Statist. Med.*, 3, 375-384.
- JESDINSKY, H. J. (1977): Statistische Auswertung großer Datenmengen - nur ein technisches Problem? *Statist. Software Newsl.*, 3, 68-75.

- KALISH, L. A. und C. B. BEGG (1984): Treatment allocation method in clinical trials: a review. *Statist. Med.*, **4**, 129-144.
- KARRISON, T. (1981): Data editing in a clinical trial. *Controlled Clinical Trials*, **2**, 15-19.
- MARTIN, J. (1981): Einführung in die Datenbanktechnik. München: Hanser Verlag.
- MARUBINI, E. (1986): Data handling and computational methods in clinical trials. *COMPSTAT 1986*, 359-367. Heidelberg: Physica.
- MICCIOLO, R., P. VALAGUSSA und E. MARUBINI (1985): The use of historical controls in breast cancer. *Controlled Clin. Trials*, **6**, 259-270.
- POCOCK, S. J. (1976): The combination of randomized and historical controls in clinical trials. *J. Chronic Diseases*, **29**, 175-188.
- POCOCK, S. J. und R. SIMON (1975): Sequential treatment assignment with balancing for prognostic factors in the controlled clinical trial. *Biometrics*, **31**, 103-115.
- PROPPE, A. (1960): Der Primat der Fragestellung für eine wissenschaftlich nutzbare Dokumentation. *Med. Dok.*, **4**, 73-78.
- ROSATI, R. A., J. F. MCNEER, C. F. STARMER, B. S. MITTLER, J. J. MORRIS und A. G. WALLACE (1975): A new information system for medical practice. *Arch. Intern. Med.*, **135**, 1017-1024.
- SATHER, H. N. (1986): The use of prognostic factors in clinical trials. *Cancer*, **58**, 461-467.
- SCHLAGETER, G. und W. STUCKY (1983): Datenbanksysteme: Konzepte und Modelle. Stuttgart: Teubner (2. Aufl.).
- SCHUMACHER, M. (1986): Möglichkeiten und Grenzen der retrospektiven Auswertung von klinischen Studien. *FULGRAFF, G. M., H. LENAU, H. MAIER-LENZ und H. RODE (Hrsg.): Klinisch-Pharmakologisches Kolloquium II, Titisee 1986*, 36-48. Freiburg im Breisgau, CRF.
- SHAPIRO, A. R. (1977): The evaluation of clinical predictions: a method and initial application. *N. Engl. J. Med.*, **296**, 1509-1514.
- SIMON, R. (1984): Importance of prognostic factors in clinical trials. *Cancer Treat. Rep.*, **68**, 185-192.
- SONNEMANN, E. (1982): Allgemeine Lösungen multipler Testprobleme. *EDV in Medizin und Biologie*, **13**, 120-128.
- STARNER, C. F., R. A. ROSATI und F. MCNEER (1974): Editorial: Data bank use in management of chronic disease. *Comp. Biomed. Res.*, **7**, 111-116.
- VICTOR, N. (1984): Computational Statistics - tool or science? *Statist. Software Newsl.*, **10**, 105-116.
- WHITE, S. J. und L. S. FREEDMAN (1978): Allocation of patients to treatment groups in a controlled clinical study. *Brit. J. Cancer*, **37**, 849-857.
- WIESNER, B. (1980): Dokumentationssysteme in der Medizin - Probleme von Form und Inhalt. *KOLLER, S., P. L. REICHERTZ und K. ÜBERLA (Hrsg.): Medizinische Informatik und Statistik-Modelle in der Medizin*, **22**, 586-594.

Eingegangen am: 15. April 1987.

Anschrift der Verfasser: Dr. Guido Giani, Dipl.-Math. Wilfried Meyers, Abteilung Medizinische Statistik und Dokumentation der RWTH Aachen, Pauwelsstraße, D-5100 Aachen.

EDV in Medizin und Biologie 18 (2-3), 84-89, ISSN 0300-8282  
 © Verlag Eugen Ulmer GmbH & Co., Stuttgart; Gustav Fischer Verlag KG, Stuttgart

## Was bedeuten Software-Qualitätssicherung und -Validierung für den Biometriker?

H. Passing

### Zusammenfassung

*Bei der biometrischen Planung und Auswertung von Versuchen wird heute komplex vernetzte Software eingesetzt. Ihre Qualität kann die Arbeit des Biometrikers und auch das Ergebnis seiner Berichte beeinträchtigen. Jeder Biometriker ist für seine Berichte und damit auch für die Qualität der dabei eingesetzten Software verantwortlich.*

*Qualitätssicherung sorgt während der Entwicklung und Wartung von Software dafür, daß die vorgegebenen Qualitätsanforderungen erreicht werden. Dagegen macht der Biometriker durch eine Validierung transparent, daß die Software tatsächlich tut, was er konkret erwartet. In beiden Fällen spielen Pläne eine besondere Rolle – analog den Versuchsplänen in der Biometrie –, und es werden geeignete Nachweise erzeugt und aufbewahrt.*

*Diese Ideen sind nicht neu, sie sind aber vor allem im naturwissenschaftlichen Bereich nicht weit verbreitet. Deshalb ist eine entsprechende Bewußtseinsbildung und Weiterbildung für Biometriker erforderlich.*

*Schlüsselwörter: Software-Qualitätssicherung, Software-Validierung*

### Summary

*In biometrical planning and evaluation of experiments complexly netted software is employed. Its quality may impair the work of the biometrician and also the result of his reports. Any biometrician is responsible for his reports and therefore implicitly for the quality of the applied software, too.*

*Quality assurance provides that the predefined quality requirements are achieved during development and maintenance. By means of a validation however the biometrician gives documented evidence that the software really does what it is supposed to do. Planning is a very important aspect in both cases analogously to study plans in biometry. Records are produced and kept.*

*These ideas are not new but they are rarely used in science so that they must be brought home to biometricians. Further education is necessary.*

*Keywords: software quality assurance, software validation*

### 1. Einleitung

Der Biometriker setzt bei der Versuchsplanung und -auswertung seit Jahren schon Software ein. Anfangs waren das ganz

einfache Programme mit einfachen Algorithmen, während er heute komplexe Programmsysteme anwendet, die komplexe Algorithmen enthalten können. Die Versuchsdaten bekam er früher auf Papier; heute haben diese oftmals schon mehrere Programme durchlaufen, bevor sie überhaupt zum Biometriker gelangen: Sie wurden vielleicht online erfaßt und existieren als Rohdaten dann oftmals gar nicht auf Papier. Sie wurden vermutlich in einer Datenbank abgelegt und von dort extrahiert. Neuerdings erleben wir eine stärkere Vernetzung der Datenbestände und dadurch eine weitere Zunahme der Komplexität der Software.

Es stellt sich also für den Biometriker in zunehmendem Maße die Frage, ob die benutzte Software korrekt ist. Offenbar kann er diese Frage nicht einfach auf die von ihm angewandte Statistik-Software begrenzen.

Früher hat der Biometriker das eingesetzte statistische Verfahren oftmals selbst programmiert und getestet. Er lebte dann in dem Bewußtsein, daß sein Programm richtig ist. Allerdings wurden über die durchgeführten Tests üblicherweise keine Unterlagen aufbewahrt.

Heute kommt Software aus verschiedenen Quellen: Sie kann vom Biometriker selbst oder von einer fachfremden Person entwickelt sein. Sie kann im eigenen Hause entwickelt sein oder von einer Fremdfirma stammen. Bei PC-Software kann die Herkunft sogar unklar sein. Hinzu kommt, daß gerade der PC dazu verführt, Routinesoftware schnell einer Situation anzupassen. Alle diese Gesichtspunkte können die **Qualität von Software** beeinflussen. Leider gibt es im allgemeinen aber keine Unterlagen über die tatsächliche Qualität der Software.

Es gibt noch weitere für unser Thema bedeutsame Gesichtspunkte. Software kann für die routinemäßige Anwendung oder für einen einmaligen Einsatz bestimmt sein. Sie kann schon lange in Betrieb sein oder gerade erstmals eingesetzt werden. Sie kann auf einem zentralen System oder auf einem PC laufen.

Selbstverständlich ist der Biometriker für seine Ergebnisse verantwortlich. Er muß sich deshalb auf die eingesetzte Software verlassen können. Er kann sich gegebenenfalls nicht damit entschuldigen, daß die ihm verfügbare Software mangelhaft war. Er benötigt folglich **Kriterien**, mit denen er die Qualität der Software beeinflussen oder beurteilen kann. Außerdem wird er seine Beurteilung eingehend **dokumentieren**, damit er sich gegebenenfalls rechtfertigen kann.

Viele Biometriker sind in der pharmazeutischen Industrie tätig. Jeder Arzneimittelhersteller ist für seine Produkte verantwortlich, insbesondere für die Unterlagen, die zur Registrierung eingereicht werden. Deshalb verlangt die Food and

Drug Administration (FDA) in Anlehnung an bestehende Richtlinien (1, 2), daß Software validiert wird. Dadurch werden viele Biometriker gezwungen, sich von der Qualität der Software zu überzeugen und dies Dritten gegenüber zu belegen.

In den folgenden Abschnitten soll skizziert werden, wie man **Qualität von Software** definieren kann und was **Qualitätssicherung von Software** bedeutet. Dann wird der Begriff **Validierung von Software** geklärt. Diese Betrachtungen beziehen sich vor allem auf komplexe, routinemäßig einzusetzende Software. Sie haben nicht nur für die Biometrie Bedeutung und sind deshalb allgemein gehalten.

Im letzten Abschnitt sollen die Konsequenzen für den Biometriker diskutiert werden.

## 2. Qualität von Software

Ein **Qualitätsmerkmal** von Software ist eine Eigenschaft, bezüglich der sich Softwareprodukte unterscheiden lassen, die denselben funktionalen Umfang haben. Das Qualitätsmerkmal, für das man sich früher ausschließlich interessiert hat, ist Korrektheit. Man hat jedoch längst erkannt, daß z. B. Wartbarkeit, Benutzerfreundlichkeit und Sicherheit ebenfalls wichtige Qualitätsmerkmale von Software sind (3). Manche Qualitätsmerkmale, wie z. B. die Korrektheit, beziehen sich eher auf den Code, während sich andere, wie etwa die Wartbarkeit, eher auf die Dokumentation beziehen.

Jedes Qualitätsmerkmal hat mehrere mögliche Ausprägungen. Unter einer **Anforderung an ein Qualitätsmerkmal** versteht man die Vorgabe einer möglichen Ausprägung, die während der Entwicklung erreicht oder überschritten werden soll. Anforderungen werden also immer vor der Software-Entwicklung spezifiziert. Die tatsächliche Qualität der Software besteht darin, wie weit die Anforderungen an die Qualitätsmerkmale nachweislich erfüllt sind.

Als Anforderung an das Qualitätsmerkmal Korrektheit wird man intuitiv verlangen, daß die gesamte Software fehlerfrei ist. Aber was heißt „fehlerfrei“ bei einem komplexen System? Dazu müßte vollständig spezifiziert sein, was das System funktional leisten soll. Das schafft bei komplexen Systemen vermutlich niemand, und diese Arbeit kann nicht automatisiert werden. Aber selbst wenn man eine vollständige und sogar formale Spezifikation hätte, könnte man die Fehlerfreiheit der Software nicht nachweisen: Die Anzahl der zu testenden Situationen ist bei einem komplexen System unüberschaubar groß, und einer formalen Verifikation, d. h. einem Korrektheitsbeweis, sind technische Grenzen gesetzt. (Auf derartige Verifikationen wird im folgenden nicht weiter eingegangen.) Es wird heute in der Informatik als Tatsache akzeptiert, daß komplexe Systeme im allgemeinen nicht fehlerfrei sind.

Es erweist sich also als schwierig, eine Anforderung an das Qualitätsmerkmal „Korrektheit“ zu stellen. Das gilt erst recht für die anderen Qualitätsmerkmale, deren Bedeutung höchstens intuitiv klar ist.

Um schließlich die Qualität einer Software feststellen zu können, muß man die tatsächliche Ausprägung aller Qualitätsmerkmale messen können. Dazu wurden in der Informatik zahlreiche Software-Maße entwickelt (4), z. B. Lines of Code oder das Komplexitätsmaß von McCabe. Diese Maße haben für die Qualitätsmessung aber nur einen beschränkten Wert, da der Zusammenhang zwischen den Qualitätsmerkmalen und den Software-Maßen zu unklar ist. Außerdem beziehen sich die meisten Maße nur auf den Code und nicht auf die Dokumentation.

Ein Ausweg aus diesem Dilemma liegt darin, die Qualitätsmerkmale zu **operationalisieren** (5). Dabei werden für jede mögliche Ausprägung Entwicklungstätigkeiten definiert, die machbar, nachprüfbar und oftmals sogar quantifizierbar sind. Eine Qualitätsanforderung ist dann eine Liste derartiger Tätigkeiten, die während der Entwicklung durchgeführt werden müssen.

Anstelle einer Anforderung an die Korrektheit der Software wird man dann eine Anforderung an den Umfang der Korrektheitsprüfung, d. h. der Tests, stellen. Unter Testen wollen wir den jederzeit wiederholbaren Nachweis der Korrektheit eines Objektes relativ zu vorher festgelegten Testzielen verstehen (6). Anforderungen an die Korrektheit werden dann durch Testziele ausgedrückt. Diese Testziele definieren Tätigkeiten, die systematisch durchgeführt werden müssen.

Man könnte nun beispielsweise verlangen, daß jeder Modul einzeln und daß das System als Ganzes getestet wird. Zwei quantitative Testziele für den Test eines Moduls könnten etwa darin bestehen, daß für jeden Grenzwert eines jeden Eingabeparameters je ein Testfall unterhalb, auf und oberhalb des Grenzwertes genommen wird und daß mindestens 90 % aller Zweige des Moduls beim Testen durchlaufen werden. Testziele für den Systemtest könnten etwa darin bestehen, daß alle Funktionen in wenigstens drei Beispielen geprüft werden, daß alle vorgesehenen Fehlermeldungen erzeugt werden und daß die Grenzen der Anwendbarkeit geprüft werden.

Wie weit diese Anforderungen erreicht wurden, ist meßbar, gegebenenfalls unter Benutzung geeigneter Werkzeuge. Wenn alle Anforderungen erfüllt sind, folgt natürlich nicht, daß die Software insgesamt korrekt ist, man weiß aber, was geprüft wurde.

Wir wollen noch ein weiteres Qualitätsmerkmal betrachten, und zwar die Wartbarkeit. Ob Software mit vertretbarem Aufwand wartbar ist, hängt im wesentlichen von der Güte ihrer Dokumentation ab. Es ist deshalb naheliegend, einige Dokumente und ihren formalen Aufbau zu **standardisieren**. Es gibt derartige Standards, z. B. (7, 8, 9). Man wird dann dafür sorgen, daß diese Dokumente immer entwicklungsbegleitend erstellt werden. Die vorgegebene Gliederung bewirkt, daß sich im Wartungsfalle sogar eine Person in den Dokumenten zurecht findet, die sie nicht selbst erstellt hat. Wir kommen darauf noch zurück.

Weitere Qualitätsmerkmale lassen sich ebenfalls operationalisieren (5). Darauf soll hier nicht weiter eingegangen werden.

## 3. Qualitätssicherung von Software

Unter **Qualitätssicherung von Software** verstehen wir die Gesamtheit der geplanten und systematisch durchgeführten Maßnahmen, um die Qualitätsanforderungen während des gesamten Lebenszyklus zu realisieren. Ohne klar definierte Qualitätsanforderungen ist also eine Qualitätssicherung nicht möglich.

Qualitätssicherung von Software umfaßt **Maßnahmen zur Steuerung der Qualität** und **Maßnahmen zur Prüfung der Qualität** (3). Durch Steuerungsmaßnahmen wird vor und während der Entwicklung möglichst viel prophylaktisch getan, damit nach Abschluß der Entwicklung die Qualitätsanforderungen möglichst erfüllt sind. Durch Prüfmaßnahmen wird kontrolliert, ob die Anforderungen tatsächlich erfüllt sind. Prüfmaßnahmen finden an definierten Meilensteinen während der Entwicklung statt, also nicht nur am Ende der Entwicklung. Dadurch ist Qualitätssicherung in den gesamten Entwicklungsprozess **integriert**.

Die Qualität bereits vorhandener Software kann also nicht mehr gesichert, allenfalls noch retrospektiv geprüft werden. Qualitätssicherung hat damit prospektiven Charakter, und damit besteht eine gewisse Analogie zu prospektiven Studien in der Biometrie.

Der Begriff Qualitätssicherung mit Steuerungs- und Prüfmaßnahmen ist keinesfalls neu. Viele Prozesse umfassen eine Qualitätssicherung; man denke etwa an die Herstellung von Autos oder von Arzneimitteln oder an die Qualitätssicherung in der Laboratoriumsmedizin. Bei diesen technischen, chemischen oder biologischen Prozessen sind die Einflußgrößen allerdings leichter zu erkennen und zu kontrollieren als bei dem intellektuellen Prozeß der Entwicklung von Software.

Die Informatik hält schon seit längerem die geeigneten Steuerungsmaßnahmen unter dem Begriff des **Software-Engineering** bereit. Software-Engineering bedeutet, daß Software ingenieurmäßig, d. h. systematisch, **entwickelt** wird. Einige Ideen dazu sollen hier skizziert werden.

Ausgangspunkt ist die Erkenntnis, daß der Code ein Folgeprodukt vorangehender Entwicklungsschritte ist. Deshalb wird der gesamte Entwicklungsprozeß formal in Abschnitte zerlegt, in denen mit definierten Tätigkeiten und definierten Methoden vorgegebene Ziele erreicht werden. Bei dieser Formalisierung hat man aber sehr viel Spielraum für eigene Festlegungen.

Üblicherweise wird der Entwicklungsprozeß in zeitlich aufeinander folgende **Phasen** aufgeteilt, beispielsweise in Definition der Anforderungen, in Entwurf, in Kodierung und Test der Moduln, in Integration zum Gesamtsystem und in Installation in die Zielumgebung. Eine nachfolgende Phase soll dabei erst begonnen werden, wenn die vorangehende beendet ist. Die Erfahrung zeigt jedoch, daß derartige Phasenmodelle recht starr sein können, wenn z. B. einige Moduln schon integriert werden sollen, obwohl noch nicht alle übrigen Moduln fertig sind.

Dieser Nachteil läßt sich durch eine modifizierte Betrachtungsweise leicht umgehen. Denn das Ziel, das in einer Phase erreicht werden soll, kann man sich als Dokument vorstellen, und damit bietet sich eine dokumentenbezogene Einteilung an. Das bedeutet, daß ein Folgedokument erst dann begonnen werden kann, wenn die dafür vorauszusetzenden Dokumente bereits vorliegen. Wir haben auch gesehen, daß durch Standardisierung von Dokumenten das Qualitätsmerkmal Wartbarkeit operationalisiert werden kann. Durch diese Vorgehensweise wird die gesamte Dokumentation in den Entwicklungsprozeß integriert, und es kann nicht mehr passieren, daß sie erst nachträglich erstellt oder ganz ausgelassen wird. Außerdem wird durch Standards ermöglicht, daß der Rechner die Dokumentenerstellung unterstützt. Umgekehrt sorgt der Rechner dafür, daß Formalien automatisch eingehalten werden. Diese Unterstützung ist für die Akzeptanz der Vorgehensweise durch die Software-Entwicklung ganz wesentlich.

Wir schlagen eine Einteilung in folgende Dokumente vor:

- Pflichtenheft,
- Systembeschreibung,
- Benutzerhandbuch,
- Installationshandbuch (ggf.),
- Testdokumentation,
- Code.

Im **Pflichtenheft** definieren die späteren Anwender die funktionalen und die Qualitätsanforderungen detailliert aus ihrer Sicht. Das Pflichtenheft beschreibt, was die Software tun soll. Es ist Voraussetzung für die gesamte Entwicklung und für den Test des gesamten Systems. Es darf deshalb während der

Entwicklung nicht mehr geändert werden. Wenn kein Pflichtenheft vorliegt, kann keine Qualitätssicherung erfolgen.

Anschließend wird das DV-Konzept entworfen. In der **Systembeschreibung** werden die Bausteine, die Algorithmen und der Aufbau der Dateien und Datenbanken und ihre wechselseitigen Beziehungen untereinander dargestellt. Die Systembeschreibung enthält damit, was die Software im Detail leistet und wie sie das tut. Sie ist die wesentliche Voraussetzung für spätere Wartungen.

Das **Benutzerhandbuch** beschreibt, was die Software leistet und wie sie zu bedienen ist. Das Benutzerhandbuch ist die Voraussetzung dafür, daß die Software später sachgerecht und richtig benutzt werden kann.

Das **Installationshandbuch** beschreibt schließlich die Schritte, die bei der Installation der Software in der Zielumgebung des Benutzers (das ist möglicherweise ein anderer Rechner als der, auf dem die Software entwickelt wurde) durchzuführen sind. Das Installationshandbuch ist die Voraussetzung dafür, daß die Software dann auch in der Zielumgebung richtig funktioniert.

Wir haben bereits gesehen, daß Testziele die Grundlage des Testens sein sollen. Sie werden über die Qualitätsanforderungen im Pflichtenheft prinzipiell vorgegeben und in **Testplänen** ggf. weiter konkretisiert. Ein Testplan enthält außer den präzisierten Testzielen die Spezifikation der Testumgebung und der Randbedingungen, der einzusetzenden Werkzeuge, des Akzeptanzkriteriums und der verantwortlichen Personen. Aus den Testzielen werden die Testfälle und daraus die Testdaten abgeleitet. Während ein Testfall eine verbale Charakterisierung einer zu testenden Situation ist, sind Testdaten die Realisierung eines Testfalles. Jeder Testfall und seine Testdaten besitzen ein erwartetes Ergebnis. Beim Modultest leitet es sich aus der Modulbeschreibung ab, beim Systemtest dagegen aus dem Pflichtenheft.

Tests sollen auf diese Weise nicht nur systematisch, sondern auch reproduzierbar und nachvollziehbar sein. Reproduzierbarkeit bedeutet, daß jeder Testfall jederzeit wiederholbar ist. Das ist für spätere Wartungen bedeutsam. Nachvollziehbarkeit bedeutet, daß jeder Testfall selbsterklärend beschrieben ist und daß eine vollständige, auch für Außenstehende verständliche Dokumentation über die Testdurchführung vorliegt. Diese **Testdokumentation** enthält

- den Testplan,
- die Testfälle,
- die Zuordnung der Testfälle zu den Testzielen,
- die Testergebnisse.

Zum systematischen Vorgehen des Software-Engineering gehört wesentlich der Einsatz von Methoden und **Werkzeugen**. Die Werkzeuge sollten möglichst in einer **Software-Entwicklungsumgebung** aufeinander abgestimmt sein. Darauf soll hier nicht näher eingegangen werden.

Wir wollen nun annehmen, daß eine Vorgehensweise definiert wurde. Alle Arbeiten im Rahmen dieser Vorgehensweise können von einem Personenkreis durchgeführt werden. Prüfmaßnahmen im Rahmen der Qualitätssicherung müssen dagegen von **unabhängigen Personen** durchgeführt werden.

Für Prüfmaßnahmen muß ebenfalls ein Formalismus aufgebaut werden. Alle Prüfungen erfolgen an vorgegebenen **Meilensteinen** im Entwicklungsprozeß. Ein Meilenstein ist z. B. der Abschluß eines der genannten Dokumente oder eines Kapitels daraus.

Geprüft werden die genannten Dokumente im Hinblick auf die Qualitätsanforderungen im Pflichtenheft. Dabei unterscheidet man formale und inhaltliche Prüfungen. **Formale Prüfungen** bauen auf dem Formalismus auf, der bei der De-

definition der Vorgehensweise erstellt wurde. Sie sind ziemlich trivial, wenn die Dokumentenerstellung vom Rechner unterstützt wird und wenn geeignete Prüfwerkzeuge benutzt werden können. Dagegen sind **inhaltliche Prüfungen** aufwendig, da sie weitgehend intellektuell durchgeführt werden müssen. Je formaler jedoch ein Dokument aufgebaut ist, desto mehr wird seine inhaltliche Prüfung vereinfacht. Außerdem kann das gesamte System, insbesondere die Kombination von Funktionen, weitergehend getestet werden.

Alle diese Prüfungen werden natürlich ebenfalls dokumentiert; dazu haben wir das **Qualitätskontrollheft**. Es belegt die tatsächliche Qualität der Software.

Nach Abschluß aller Prüfungen wird die Software in einen definierten Zustand versetzt, nämlich **Version 1**. Sie wird dann in der Zielumgebung installiert und probeweise in Betrieb genommen. Der Anwender macht die Abnahme und den Akzeptanztest (s. u.). Wenn dabei Fehler auftreten, werden sie anschließend beseitigt. Erst danach wird Version 1 zum produktiven Einsatz freigegeben.

Während des Betriebs können Fehler auftreten, oder es können Erweiterungs- oder Anpassungswünsche durch die Anwender vorgebracht werden. Heute werden derartige Änderungen oftmals auf Zuruf gemacht; die Erfahrung lehrt jedoch, daß dabei neue Fehler in die Software eingebaut werden.

Qualitätssicherung soll dafür sorgen, die Qualitätsanforderungen während des gesamten **Lebenszyklus** zu realisieren. Analog zur Erstentwicklung muß deshalb auch die **Wartung** formalisiert werden. Dazu werden Fehlermeldungen gesammelt, und weitergehende Wünsche und zugehörige Qualitätsanforderungen werden in einem neuen Pflichtenheft definiert. Analog zu einer Neuentwicklung werden alle Dokumente durchlaufen und gegebenenfalls aktualisiert. Qualitätsprüfungen sind natürlich ebenfalls involviert. Dabei wird für eine sorgfältige Trennung zwischen dem Code, der in Betrieb ist, und dem in Wartung befindlichen Code gesorgt; Entsprechendes gilt auch für alle übrigen Dokumente. Dadurch wird sichergestellt, daß die bisherige Version bis zum Abschluß der Wartung unbeschädigt in Betrieb bleiben kann.

Besondere Aufmerksamkeit verdient bei einer Wartung das Testen: Es werden nämlich nicht nur neue Testfälle für die Änderungen getestet, sondern es wird auch ein **Regressionstest** mit den früheren Testfällen gemacht. Dadurch überzeugt man sich, daß die nicht geänderten Moduln die gleichen Ergebnisse wie bislang bringen.

Nach Abschluß der Wartung hat man die nächste Version. Ihre Qualität ist im aktualisierten Qualitätskontrollheft belegt. Sie wird analog zu Version 1 freigegeben. Der Lebenszyklus der Software stellt sich damit als Folge von definierten Versionen dar.

Es könnte jetzt der Eindruck entstanden sein, daß Qualitätssicherung einen immensen zusätzlichen Aufwand bedeutet. Publierte Erfahrungen (10) zeigen jedoch, daß der Aufwand für Entwicklung und Wartung während des gesamten Lebenszyklus durch Qualitätssicherung erheblich zurückgeht:

Entwicklungsbegleitende Qualitätsprüfungen sorgen nämlich dafür, daß viele Fehler, vor allem Entwurfsfehler, nicht erst beim Testen gefunden werden. Je früher aber ein Fehler gefunden wird, desto geringer ist der Aufwand, ihn zu beheben. Außerdem läßt sich durch den Einsatz von Werkzeugen viel Aufwand kompensieren. Deshalb wird sich der Aufwand für die Erstentwicklung nicht wesentlich erhöhen.

Andererseits wird der Aufwand für Wartungen sogar stark zurückgehen. Denn erstens enthalten die freigegebenen Ver-

sionen dank der systematischen Entwicklung und der projektbegleitenden Prüfungen weniger Fehler, so daß weniger Anlässe für eine Wartung bestehen. Zweitens entfallen viele nachträgliche Änderungswünsche, wenn der Leistungsumfang am Anfang klar definiert wird. Schließlich verzichten Anwender wegen der formalen Handhabung von Wartungen auf unnötige Extras, die sie sonst vielleicht gefordert hätten.

Diese Argumente erscheinen uns überzeugend. Eigene Erfahrungen haben wir aber bisher noch nicht.

#### 4. Validierung von Software

Software zu **validieren** bedeutet, evident zu machen und zu dokumentieren, daß sie tatsächlich tut, was sie tun soll (11). Eine Validierung soll also zwei Fragen beantworten:

- Was soll die Software tun?
- Tut die Software das tatsächlich?

Bei einer Validierung müssen Dokumente erstellt werden, aus denen sich später ersehen läßt, wie diese Fragen beantwortet wurden.

Im Gegensatz zur Qualitätssicherung ist Validierung eine retrospektive Tätigkeit und verursacht deshalb zusätzlichen Aufwand. Validierung setzt Qualitätssicherung nicht notwendigerweise voraus; wir werden jedoch sehen, daß vorherige Qualitätssicherung eine Validierung sehr vereinfacht.

Validierung ist nicht eine einmalige Prüfung. Damit der gesamte Lebenszyklus abgedeckt werden kann, unterscheidet man deshalb die **Erstvalidierung** und **Revalidierungen**. Bei den Revalidierungen wiederum kann man mehrere Anlässe unterscheiden:

Nach einer Änderung der Betriebsumgebung, z. B. des Betriebssystems, gibt es die **umgebungsbedingte Revalidierung**. Bei ihr wird eher stichprobenartig geprüft, ob das System noch lauffähig ist.

Nach einer Änderung der Software dagegen gibt es die **produktbedingte Revalidierung**. Dabei werden sowohl die geänderten als auch die unveränderten Teile geprüft. Wenn die Änderung geringfügig war, wird man jedoch auf eine produktbedingte Revalidierung verzichten wollen. Damit sich auf diese Weise nicht doch im Laufe der Zeit unbemerkt Fehler einschleichen können, gibt es von Zeit zu Zeit **routine-mäßige Revalidierungen**.

Es sollen nun einige **Validierungsprinzipien** angegeben werden.

Alle Validierungen basieren auf einer **internen Richtlinie**, die die Aktivitäten bei einer Validierung und die dabei zu erstellenden Dokumente formalisiert. Diese müssen so definiert werden, daß aus ihnen hervorgeht, was die Software tun soll, daß sie das tatsächlich tut und wie das geprüft wurde. Für eine konkrete Validierung wird dann aus dieser Richtlinie ein **Validierungsplan** abgeleitet.

Eine Validierung soll die **Qualität** der Software transparent machen. Der sicherste Weg zu qualitativ hochwertiger Software besteht in den beschriebenen Steuerungsmaßnahmen der Qualitätssicherung, die offenbar zu einer geeigneten Dokumentation führen. Derartige Software ist deshalb besonders einfach zu validieren. Deshalb wird man Software, die noch nicht existiert und die später validiert werden soll, in der beschriebenen Weise entwickeln. Wenn aber von bereits existierender Software die definierte Dokumentation fehlt oder noch unvollständig ist, muß sie nachträglich erstellt werden.

Bei einer Validierung muß die **Lauffähigkeit** der Software in der Betriebsumgebung des Benutzers nachgewiesen werden, falls diese von der Entwicklungsumgebung verschieden ist.

Jede Validierung erfolgt unter der **Gesamtverantwortung des Benutzers**. Denn für die Brauchbarkeit einer Software ist nicht nur ihre Qualität entscheidend, sondern auch, wie weit sie den praktischen Erfordernissen des Benutzers genügt. Es kann ja durchaus passieren, daß eine Software die Qualitätsanforderungen des Pflichtenheftes alle erfüllt, daß jedoch das Pflichtenheft unvollständig war. Der Benutzer kann aber natürlich die Verantwortung für Teilaspekte einer Validierung delegieren.

Der Aufwand einer Validierung hängt davon ab, wie **kritisch** die Anwendung der Software ist.

Damit der gesamte Lebenszyklus der Software durch einige Validierungen abgedeckt werden kann, muß die Software in definierten **Versionen** geführt werden. Jede neue Version wird produktbedingt revalidiert.

Sämtliche Dokumente einer Version müssen **aufbewahrt** werden, soweit das möglich oder sinnvoll ist. Außerdem müssen sämtliche Dokumente, die bei einer Validierung erstellt werden, aufbewahrt werden.

Aus diesen Prinzipien wird man seine eigene Vorgehensweise ableiten. Wir schlagen vor (11), prinzipiell drei verschiedene Aktivitäten zu unterscheiden, nämlich Qualitätsprüfung, Abnahme und Akzeptanztest. Außerdem schlagen wir vor, die bereits genannten Dokumente Systembeschreibung, Testdokumentation und Benutzerhandbuch zu definieren: Das Benutzerhandbuch enthält aus der Sicht des Anwenders, was die Software tun soll und wie sie bedient wird. Die Systembeschreibung enthält, was die Software im Detail tut und wie sie aufgebaut ist. Die Testdokumentation soll belegen, daß die Software das Gewünschte tatsächlich leistet und wie das geprüft wurde. Ein Qualitätskontrollheft belegt sogar die Prüfungen durch unabhängige Personen.

Bei der **Prüfung der Qualität** wird man mehrere **Qualitätstypen** unterscheiden. Die Zuordnung zu einem Qualitätstyp hängt dann davon ab, welche Dokumente bereits existieren:

Software, die nachweislich unter Qualitätssicherung erstellt wurde, bildet deshalb den besten Typ und erfordert keine Nachdokumentation. Für Software, die schon lange in Betrieb ist, liegt viel Erfahrung vor. Anstelle einer Testdokumentation reicht dann ein Erfahrungsbericht. Die Systembeschreibung und das Benutzerhandbuch werden erforderlichenfalls nachträglich erstellt. Software, für die es keinerlei Unterlagen und kaum Erfahrung gibt, bildet den schlechtesten Typ. Für sie müssen erforderlichenfalls Systemhandbuch, Testdokumentation und Benutzerhandbuch nachträglich erstellt werden.

Die zweite Aktivität ist die **Abnahme** der Software. Dabei wird die Lauffähigkeit der Software in der Betriebsumgebung des Benutzers mit vorgegebenen Beispielen stichprobenartig festgestellt.

Die dritte Aktivität ist der **Akzeptanztest**. Dabei macht sich der Anwender mit der Software vertraut und prüft unter Regie eines Testplanes im probeweisen Betrieb, ob oder wie weit sie den praktischen Bedürfnissen entspricht. Dadurch entsteht die Testdokumentation des Akzeptanztests.

Bei einer Validierung sind nicht immer alle Aktivitäten erforderlich. Im Validierungsplan wird alles Nähere festgelegt. Einige Beispiele sollen das verdeutlichen:

**Beispiel 1:** Ein Anwender läßt Software mit begleitender Qualitätssicherung entwickeln. Dazu gibt er ein Pflichtenheft vor und bestellt Systembeschreibung, Testdokumentation, Benutzerhandbuch und Qualitätskontrollheft nach einem vorgegebenen Standard. Bei der Validierung nimmt er das Qualitätskontrollheft ab, und er macht die Abnahme und den Akzeptanztest.

**Beispiel 2:** Ein Anwender läßt Software entwickeln. Er gibt

ein Pflichtenheft vor und bestellt Systembeschreibung, Testdokumentation und Benutzerhandbuch nach einem vorgegebenen Standard. Er läßt jedoch keine Prüfmaßnahmen durch unabhängige Personen durchführen. Bei der Validierung prüft er die ausgelieferten Dokumente, d. h. er erstellt selbst ein Qualitätskontrollheft. Außerdem macht er die Abnahme und den Akzeptanztest.

**Beispiel 3:** Ein Anwender entwickelt seine Software selbst. Dabei erstellt er die Systembeschreibung, die Testdokumentation und das Benutzerhandbuch. Eine Validierung ist dann nicht mehr notwendig.

**Beispiel 4:** Ein Anwender benutzt weitverbreitete Standardsoftware, wie z. B. SAS. Man kann dann annehmen, daß die Gesamtheit der Anwender implizit für eine ständige Prüfung der Qualität sorgt. FDA gesteht deshalb aus pragmatischen Gründen zu, daß die eigene Prüfung der Qualität nicht erforderlich ist. Die Lauffähigkeit wird stichprobenartig mit gelieferten Beispielen geprüft. Der Akzeptanztest kann unter Umständen entfallen, wenn primär die klar definierten mathematischen Algorithmen interessieren.

Die Validierungsdokumente werden natürlich aufbewahrt. Schließlich gibt ein **Lebenslauf der Software** einen Überblick über die bisherigen Versionen und Validierungen.

Wir haben den Zusammenhang zwischen Software-Entwicklung und Validierung gesehen. Wer im Hause Software entwickelt, muß deshalb eine geeignete Vorgehensweise dafür festlegen. Darüber hinaus müssen Auslösung und Durchführung von Wartungen formalisiert werden.

An dieser Stelle sei darauf hingewiesen, daß weder Qualitätssicherung noch eine Validierung verhindern können, daß Software falsch angewandt wird und deshalb falsche Ergebnisse liefert. Dieses Risiko läßt sich nur durch gute Benutzerhandbücher oder durch Schulung der Benutzer gering halten.

## 5. Konsequenzen für den Biometriker

In der Biometrie und in vielen anderen Naturwissenschaften sind Qualitätssicherung und Validierung von Software bis heute nicht verbreitet. Man ist sich der Problematik oftmals gar nicht bewußt. Insofern ist die Situation heute ähnlich wie zur Zeit der ersten Statistikprogramme, als man manche numerischen Probleme bei der Programmierung einfach nicht gesehen hat.

Dadurch, daß FDA Validierungen fordert, werden die Biometriker mit der Thematik konfrontiert. Es ist anzunehmen, daß andere Behörden folgen werden. Bei sicherheitsrelevanter Software (z. B. in Kernkraftwerken) wird Qualitätssicherung seit längerem praktiziert. Für kommerzielle Software beginnt man damit, es wird aber vor allem sehr viel davon geredet.

Qualitätssicherung und Validierung erfordern offenbar einen Formalismus, der manchem Biometriker gewiß nicht fernliegt, während er andere erschrecken mag. Bei kleinen Programmen oder bei Ad-hoc-Programmen reduziert sich der Formalismus auf einen kleinen Kern. Andererseits kann bei der Entwicklung eines neuen Verfahrens der Biometrie z. B. durchaus Kreativität in der Software-Entwicklung erforderlich sein. Der Biometriker wird sich deshalb fragen, welche Vorteile Qualitätssicherung bzw. Validierung bringen und wann entsprechende Maßnahmen angemessen sind.

Die Vorteile sind:

Der Biometriker wird sich über die Brauchbarkeit angebotener und eigener Software klar. Indem er systematisch vorgeht, wird die eigene Vorgehensweise zur Methode, und seine Beur-



teilung ist reproduzierbar. Er bekennt sich zu seiner Verantwortung und erzeugt Nachweise der eigenen Sorgfalt. Er erliegt selbst nicht der Versuchung, Software für den Augenblick zu ändern, und schließt damit eine berüchtigte Fehlerquelle aus. Wenn er alle Versionen aufbewahrt, sind alle statistischen Ergebnisse jederzeit reproduzierbar.

In folgenden Fällen sind Qualitätssicherung bzw. Validierung angemessen:

Software, die routinemäßig eingesetzt werden soll, sollte generell qualitätsgesichert bzw. validiert werden. Für Software, die an Laien weitergegeben werden soll, die also ihrerseits keine Plausibilitätsbetrachtungen mit den Ergebnissen machen können, gilt dasselbe. Software, die in irgendeiner Weise kritisch ist, sollte ebenfalls generell qualitätsgesichert bzw. validiert werden.

Für Software, die der Biometriker selbst entwickelt und die er an andere weitergeben oder gar publizieren möchte, sollte er die genannten Dokumente ebenfalls verfügbar machen. Das zeugt von Ehrlichkeit und Verantwortungsbewußtsein gegenüber den Anwendern.

Schließlich wird man sich fragen, wie sich Validierungen vereinfachen lassen:

Einerseits sollten wir die Anbieter von Software drängen, daß sie uns Dokumente zur Verfügung stellen, die unsere Validierung erleichtern.

Andererseits wurde bereits gezeigt, daß weitverbreitete Standardsoftware im Hinblick auf Validierungen erhebliche Vorteile bieten kann. Deshalb ist zu erwarten, daß in der Biometrie in Zukunft in noch stärkerem Maße Standardsoftware eingesetzt werden wird. Diese Entwicklung könnte die Verbreitung neuer Verfahren behindern. Allerdings wird die biometrische Forschung selbst durch die Notwendigkeit von Validierungen nicht behindert. Vielmehr dienen Validierungen der Konsolidierung.

Es gibt zu diesem Thema bestimmt einen Bedarf an Weiterbildung. Wir glauben, daß hierin auch eine Aufgabe der Biometrischen Gesellschaft und der Gesellschaft für Medizinische Dokumentation, Informatik und Statistik (GMDS) liegt.

## Literatur

- (1) Good Practices in the Manufacture and Quality Control of Drugs (GMP). WHO Official Records, No 176, Annex 12, Part 1, 99-104, 1969.
- (2) Nonclinical Laboratory Studies. Good Laboratory Practice Regulations (GLP). Federal Register 43, No 247, 59986-60025, 22. 12. 1978.
- (3) Deutsche Gesellschaft für Qualität (DGQ) und Nachrichtentechnische Gesellschaft im VDE (NTG): Software-Qualitätssicherung, Aufgaben, Möglichkeiten und Lösungen. Beuth Verlag, Berlin, 1986.
- (4) GILB, T.: Software Metrics. Winthrop Publishers Inc., Cambridge Mass., 1977.
- (5) WILLMER, H. und H. BALZERT: Fallstudie einer industriellen Software-Entwicklung. Bibliographisches Institut, Reihe Informatik, Band 39, Mannheim, 1984.
- (6) ABBENHARDT, H.: TUS - Methode und Werkzeug für den systematischen Software-Test. In: Bibliographisches Institut, Reihe Informatik, Band 44, Mannheim 1985.
- (7) ANSI/IEEE Standard 830-1983, IEEE Guide to Software Requirements Specifications.
- (8) ANSI/IEEE Standard 730-1981, IEEE Standard for Software Quality Assurance Plans.
- (9) ANSI/IEEE Standard 829-1983, IEEE Standard for Software Test Documentation.
- (10) SNEED, H.: Software-Qualitätssicherung für kommerzielle Anwendungssysteme. Verlag Rudolf Müller, Köln, 1983.
- (11) PASSING, H. und H.-D. UNKELBACH: Software-Validierung aus dem Blickwinkel der GLP- bzw. GMP-Richtlinien. Die Pharmazeutische Industrie 49, 6 (1987), 590-595.

Eingegangen am: 15. April 1987.

Anschrift des Verfassers: Dr. H. Passing, c/o Hoechst AG, Abt. Informatik und Kommunikation/Software, Postfach 800 320, D-6230 Frankfurt 808.

## Expertensysteme in der Statistik – Experten oder Ignoranten?

B. Streitberg

### Zusammenfassung

*Polemisch kann der Beitrag in folgenden Thesen zusammengefaßt werden:*

- (1) *Es gibt keine statistischen Expertensysteme*
- (2) *Es wird auch niemals solche Systeme geben*
- (3) *Da es sie nicht geben kann*
- (4) *Dennoch ist es sinnvoll, an den Problemstellungen derartiger Systeme zu arbeiten.*

*Interpretiert man eine Reihe von Theoremen über das Entscheidungsproblem als technische Aussagen über Möglichkeit und Komplexität von Inferenzkalkülen in Expertensystemen, so kann man zeigen, daß vollständige und korrekte Kalküle unmöglich sind, sobald der Ausdrucksreichtum der Wissensbasis ein gewisses Niveau übersteigt. Für die Formulierung statistischen Wissens ist dieser Ausdrucksreichtum jedoch unabdingbar.*

### Summary

*Somewhat polemically, the paper can be condensed to the following theses:*

- (1) *There are no statistical expert systems*
- (2) *There never will be any*
- (3) *Because there can't be any*
- (4) *Nevertheless it is worthwhile to work on the problems posed by them.*

*By interpreting certain theorems about the Entscheidungsproblem by Goedel, Church, Turing and others as technical statements about the feasibility of inference machines in expert systems, it is shown, that inference machines are impossible for knowledge bases, that contain rules of the type a mathematical statistician works with.*

### 0. Vorbemerkung

Wenn man über ein Teilgebiet der künstlichen Intelligenz spricht, ist es schwer, emotional und ideologisch geprägten Auseinandersetzungen aus dem Wege zu gehen. Es scheint mir daher fair, zunächst meinen Standort klarzumachen.

Die erste Übung, die ich an der FU Berlin mitveranstalten konnte, behandelte das Gebiet lernender Computerprogramme. Seither verwende ich Computer tagtäglich in der praktischen Arbeit. Neben eher prosaischen Anwendungen, wie statistisch-biometrischen Auswertungen oder der Verarbeitung des vorliegenden Textes, benutze ich den Rechner primär als eine Art mathematisches Laboratorium, in dem man Vermutungen über mathematische Sachverhalte gewin-

nen und tentativ testen kann, ob sich die Arbeit des Beweisens überhaupt lohnt. Auch wenn die folgenden Thesen also eine gewisse Skepsis gegenüber den Versprechungen der »artificial intelligentsia« erkennen lassen, handelt es sich bei deren Autor nicht um einen Computerfeind. Im Gegenteil.

Ich halte Computer für intelligenzverstärkende Maschinen, deren zunehmender Einsatz zu immer intelligenteren Problemlösungen, gerade in unserem Gebiet der Statistik, führen kann. Natürlich setzt dies eine Überwindung der heute immer noch dominanten Benutzeroberflächen der frühen 60er Jahre (unbequeme Batchorientierung wie SPSS oder m. E. stupidisierende Menueorientierung wie Statgraphics voraus) – die Entwicklung graphischer Oberflächen (Rank Xerox Palo-Alto-Projekt, teilweise im Macintosh realisiert) oder die zunehmende Akzeptanz des freien Dialogzugriffs zu mächtigen funktionellen Systemen wie APL2, S, NIAL lassen hier hoffen.

Allerdings meine ich, daß man sich klarmachen muß, daß es nicht der Verstärker ist, der die Musik macht. Computerprogramme können bereits vorhandene Intelligenz verstärken, sie ersetzen sollen sie nicht. Vielleicht basiert die augenblickliche Attraktivität des KI-Bereichs zum Teil auf einer gewissen Unmündigkeit der potentiellen Benutzer, d. h. auf deren, vielleicht auch nur eingebildeten, Unfähigkeit, sich ihres eigenen Verstandes ohne die Anleitung eines Expertensystems zu bedienen.

Wie die vorstehende Bemerkung, so sind bewußt auch die folgenden (Anti-)Thesen provokativ gehalten. Ihr Ziel wäre erreicht, wenn aus ihnen eine anregende Diskussion und, womöglich, eine Synthese der gegensätzlichen Positionen entstünde.

Aktuelle Informationen sind in den Proceedings der Konferenzen Princeton 1985 (GALE 1986), Aachen 1985 (HAUX, 1986) und Rom (COMPSTAT) zu finden. Der etwas feuilletonistische Stil mag durch den behandelten Gegenstand seine Entschuldigung finden.

### 1. Einführung und Bemerkungen zur Terminologie

Bereits das Wort Expertensystem ist geschickt gewählt. Allerdings, wie ich befürchte, wohl eher im Interesse des Marketings solcher Systeme als in dem einer sachlichen Einschätzung ihrer Möglichkeiten. Der potentielle Käufer stellt sich wohl ein Computerprogramm vor, das einen menschlichen Experten (z. B. einen statistischen Berater) ersetzen kann, also auf Knopfdruck ein Problem des Benutzers im Dialog erfaßt, mit Hilfe der gesammelten Expertise des Fachs eine mögliche Problemlösung erstellt und schließlich, wie ein menschlicher Experte, den Lösungsweg sinnvoll begründet, d. h. Fragen

beantworten kann. Das ganze Gebiet der KI wimmelt von Neologismen eines bestimmten Typs. Charakteristisch ist die Verknüpfung von Begriffen, die in der Psychologie oder Umgangssprache menschliche Intelligenzleistungen bezeichnen, mit maschinellen Assoziationen. Beispiele finden sich bereits bei der Bezeichnung der Module eines solchen Computerprogramms:

Die **Dialogkomponente** ist die Schnittstelle zum Benutzer. Das idealtypische Expertensystem sieht einen natürlichsprachlichen Dialog vor. Die angebotenen Computerprogramme sind davon allerdings zumeist weit entfernt. Während ein natürlichsprachlicher Monolog, d.h. nur der Computer schreibt Sätze, die der Benutzer mit Kurzeingaben beantworten muß, jedem Großrechnerbenutzer zur Genüge bekannt ist, ist bereits das Problem der syntaktischen Analyse frei formulierter Sätze ungelöst. Dies liegt wohl wesentlich daran, daß die natürliche Sprache ohne die Kenntnis des semantischen und pragmatischen Kontexts prinzipiell mehrdeutig ist. Ein Dialog in eingeschränkten Problembereichen ist allerdings realisierbar, wie das etwas kindische Beispiel von Textadventures (»take axe and kill dwarf«) oder besser ein fortgeschrittenes Datenbankabfragesystem wie Question & Answer (»show me names and addresses of all statisticians who published something about expert systems«) zeigt. Inwieweit für statistische Anwendungen ein natürlichsprachlicher Dialog überhaupt sinnvoll ist, ist fraglich. Alternativen sind z.B. kommando-orientierte Systeme (REX, A4) oder maskenorientierte Oberflächen (PANOS).

Die **Wissensbasis** ist die Datenbank des Systems. Sie soll zum einen die Expertise des Gebiets in einer geeigneten Datenstruktur zur Verfügung stellen, zum anderen Aussagen über das konkret vom Benutzer gestellte Problem speichern. Es herrscht ziemlicher terminologischer Wirrwarr, was die Feineinteilung der Wissensbasis und die konkret gewählten Formen der »knowledge representation« betrifft. Sinnvoll scheint mir, im Prolog-Stil, eine Unterscheidung zwischen Faktenbasis (»die vorliegenden Daten sind nicht normalverteilt«) und Regelbasis (»Wenn Daten nicht normalverteilt sind, dann wende nichtparametrische Verfahren an«). Hier aus Platzgründen nichtbehandelte Alternativen zu den üblichen regelbasierten Systemen sind frame- oder objektorientierte Systeme. Eine besonders gelungene Wortschöpfung bezeichnet den »Knowledge Engineer«. Seine Aufgabe ist es, das Expertenwissen etwa durch Befragung von Personen zu erheben und geeignet im Rechner abzubilden. Diese Abbildung ist sicher kein Isomorphismus und meine persönliche Erfahrung als Befragter ließ mich auch am Gelingen eines Homomorphismus zweifeln.

Die **Inferenzmaschine** ist das eigentliche Verarbeitungsprogramm. Ihre Aufgabe ist, mit Hilfe der Regeln, Schlüsse aus den vorliegenden Fakten zu ziehen. Ein mathematisches Modell einer simplen Inferenzmaschine wird in Abschnitt (3) vorgestellt. Die obige Kleinstwissensbasis würde den naheliegenden Schluß liefern, daß nichtparametrische Verfahren anzuwenden sind (ebenso fragwürdig wie die angewendete Regel – Exponentialverteilung?, robuste Verfahren?). Charakteristisch für Expertenwissen ebenso wie für Probleme der Praxis ist deren mangelnde Determiniertheit: »Wenn die Daten nicht normalverteilt sind, wende im Regelfall nichtparametrische Verfahren an« – »Die Daten sind vermutlich nicht normalverteilt«. Das Problem der Inferenz unter Unsicherheit in wissensbasierten Systemen ist ungelöst. Existierende Ansätze lassen im allgemeinen jedenfalls eine erschreckende Unkenntnis der hier vielleicht doch nicht ganz irrelevanten statistischen Theorie erkennen. Nach einem Satz von Hajek

(vgl. den Beitrag von Spiegelhalter in der Princeton-Konferenz) lassen sich die vorgeschlagenen Kalküle auf »Idiot Bayes« modulo einer monotonen Transformation reduzieren. Der Bayes-Ansatz stellt dabei die Unsicherheit von Fakten durch subjektive Wahrscheinlichkeit wie  $P(A)$  und die Unsicherheit von Regeln durch bedingte Wahrscheinlichkeiten wie  $P(C|A)$  dar. Eine Anwendung der Regel C if A and B kann dann in der Form  $P(C) = P(C|A,B) P(A,B)$  repräsentiert werden. Dies ist durchaus korrekt, erfordert jedoch die Festlegung von subjektiven Wahrscheinlichkeiten für alle Kombinationen von Fakten. Zum Idiot-Bayes Ansatz wird die Methode, wenn für die Wahrscheinlichkeit einer Konjunktion prinzipiell von der Unabhängigkeit der Einzelfakten ausgegangen wird, also etwa  $P(A,B) = P(A)P(B)$  unterstellt wird. Nach dem zitierten Satz lassen sich alle Unsicherheitsmaße Q, die bestimmten Plausibilitätsforderungen genügen, in der Form  $Q(A) = F(P(A))$  darstellen, wobei F monoton ist und die Wahrscheinlichkeiten  $P(A)$  aus einem Idiot-Bayes-Verfahren stammen.

Die **Erklärungskomponente** gehört ebenfalls zur Benutzerschnittstelle. Ihre Aufgabe ist, auf Grund eines Protokolls über die angewendeten Regeln dem Benutzer die gezogenen Schlußfolgerungen zu erläutern. Zahlreiche Systeme liefern auf die »Warum«-Frage nicht viel mehr als einen Dump dieses Protokolls.

Wie man auch in den folgenden Abschnitten noch sehen wird, stehen hinter den beeindruckenden Bezeichnungen oft theoretisch ungelöste Probleme und leider oft auch technisch wenig überzeugende Realisierungen. Ich würde mir eine vorsichtiger Terminologie wünschen, die nicht bereits in der Wortwahl suggeriert, daß die Probleme der Rechneremulation bestimmter menschlicher Intelligenzleistungen beim aktuellen Stand der Technologie gelöst seien.

## 2. Expertensystemshells – leere Gehäuse?

Ein Expertensystem ist nichts anderes als ein Computerprogramm. Bekanntlich läßt sich jedes überhaupt vom Rechner lösbare Problem im Prinzip in jeder beliebigen Programmiersprache realisieren, die zumindest die Mächtigkeit einer Turing-Maschine hat. Die Hobbycomputer-Zeitschriften wimmeln von Expertensystemen in BASIC, deren Wissensbasis etwa in DATA-Statements niedergelegt ist. Für professionelle Entwicklungen werden Sprachen wie LISP, PROLOG oder NIAL vorgezogen. In Aachen habe ich einige Argumente für APL2 als KI-Entwicklungssprache vorgebracht. Den meisten Anwendern dürften bei der Erstellung von Expertensystemen in diesen Sprachen jedoch noch zuviele Tools fehlen. Auf dem Markt existiert daher eine mittlerweile kaum noch überschaubare Fülle von sog. expert system shells. Dies sind i. A. nichts anderes als Programmierumgebungen, die dem Entwickler von Expertensystemen einen Regeleditor sowie eingebaute Dialogkomponenten, Inferenzmaschinen und Erklärungskomponenten zur Verfügung stellen. Nebenbei bemerkt illustrieren zumindest die meisten mir bekannten Shells zwei alte Prinzipien des DV-Markts: »Kaufe nie niedrige Versionsnummern« und »Preise werden von der Mode diktiert«. Von Kinderkrankheiten einmal abgesehen, erlauben diese Systeme aber jedenfalls eine relativ bequeme Erstellung regelbasierter Programme.

Von der Zahl (sicher > 100) der verfügbaren shells ausgehend müßte man auf eine ungeheure Zahl von angebotenen Expertensystemen schließen. Tatsächlich wurde in einer kürzlich veröffentlichten Studie festgestellt, daß weltweit nur neun

Expertensysteme kommerziell vertrieben und tatsächlich erfolgreich zur Lösung praktischer Probleme eingesetzt werden. Diese Probleme sind i. a. Reparatur- und Konfigurationsprobleme (d. h. also Expertenwissen auf der Ebene des Technikers und Sachbearbeiters).

Wie sehen die Verhältnisse auf dem Gebiet der Statistik aus? In der Princeton-Conference über AI und Statistik existierte nur ein einziges lauffähiges Expertensystem (REX), das die Autoren selbst als Experimentalsystem bezeichnen. In Aachen wurde als lauffähiges System nur das Programm PANOS von Wittkowski bezeichnet, das ebenfalls nicht kommerziell verfügbar ist. In Rom gab es trotz der Fülle von Vorträgen über Expertensysteme kein einziges wirklich laufendes. Die Standardfrage, was von dem jeweils vorgestellten Wunschkatalog denn wirklich implementiert sei, wagte man am Ende kaum mehr zu stellen. Mir ist noch die Bemerkung eines der chairmen im Ohr, ob es nicht sinnvoller wäre, die angewandte Energie in die bekanntlich realisierbare Aufgabe der Ausbildung menschlicher Experten zu investieren.

Ich werde im folgenden versuchen, für das Mißverhältnis von angebotenen shells und realisierten Expertensystemen eine Erklärung zu geben. Diese kann m. E. nicht mit der relativen Neuheit des Gebiets argumentieren (shells existieren seit mind. 10 Jahren), sondern muß aus inneren Eigenschaften der intendierten Problemlösungen und der verwendeten tools abgeleitet werden.

### 3. Über Inferenzkalküle

Jedes Gehäuse für Expertensysteme bietet auch eine sog. Inferenzmaschine an. Dies ist nichts anderes als ein Interpreter für Computersprachen eines bestimmten Typs. Allgemein handelt es sich bei diesen Sprachen um sogenannte Produktions- oder Übergangssysteme. Ein Übergangssystem auf einer Menge  $S$  der sog. Zustände ist durch endlich viele Regeln  $R_1, R_2, \dots, R_m$  gegeben. Eine Regel  $R$  ist dabei einfach eine Relation auf  $S$ , d. h. eine Menge von geordneten Paaren  $(s, t)$ , worin  $s$  als Prämisse und  $t$  als Konklusion bezeichnet wird.  $R$  heißt anwendbar auf einen Zustand  $u$  aus  $S$ , wenn  $R$  ein Paar  $(u, x)$  enthält. Ein Übergangssystem  $U$  kann man sich als gerichteten Graphen auf  $S$  mit  $m$  Kantenmengen  $R_1, \dots, R_m$  vorstellen – der Graph enthält den mit  $i$  beschrifteten Pfeil  $(u, x)$  genau dann, wenn  $(u, x)$  in  $R_i$  ist. Es heißt  $x$  aus  $u$  ableitbar, wenn ein gerichteter Kantenzug von  $x$  nach  $u$  führt (irgendwie beschriftet). Der Interpreter kann auf zwei verschiedene Weisen auf diesem Übergangssystem operieren, die man als forward bzw. backward chaining bezeichnet.

Forward chaining ist die Lösung des Wortproblems mit gegebenem Anfangszustand  $u$ , d. h. es sind alle aus  $u$  ableitbaren Zustände  $x$  anzugeben. Backward chaining ist die Lösung des Wortproblems mit gegebenem Endzustand  $x$ , d. h. es sind alle  $u$  anzugeben, aus denen  $x$  ableitbar ist. Man kann das System also als Synthese- oder als Analyseautomat betreiben. Backward Chaining in  $U$  ist nichts anderes als forward chaining im inversen System  $U'$ , in dem alle Pfeile umgekehrt werden.

Mit dieser Beschreibung ist allerdings noch sehr wenig ausgesagt, da ein Übergangssystem nur ein abstraktes Modell darstellt, in dem jeder beliebige Algorithmus beschrieben werden kann. In realen Expertensystemen findet man Übergangssysteme sehr unterschiedlicher Mächtigkeit vor. Die Spannweite erstreckt sich von sehr primitiven Systemen (im folgenden als Hüllsysteme bezeichnet) zu Systemen vom PROLOG-Typ.

Ein Hüllsystem basiert auf einer endlichen Menge  $M$  von Elementaraussagen, die nicht weiter interpretiert werden. Die Zustandsmenge  $S$  ist die Menge aller Teilmengen von  $M$ . Liegt zu einem Zeitpunkt ein Zustand  $s$  vor, so soll dies heißen, daß das System genau die Aussagen in  $s$  als »wahr« erkannt hat. Regeln werden etwa in der Form »a,b,.. if p,q,« angegeben. Formal ist eine Regel  $R$  eine Abbildung von  $S$  nach  $S$ , die durch zwei Teilmengen  $g, h$  von  $M$  definiert wird: es gilt  $R(s)$  gleich  $s$ , wenn  $s$  nicht alle Elemente von  $g$  enthält, sonst ist  $R(s)$  die Vereinigung von  $s$  und  $h$ . Im Beispiel »a if p,q« ist  $g$  gleich  $(p, q)$  und  $h$  gleich  $(a)$ , aus z.B.  $(p)$  folgt nur  $(p)$ , aus  $(p, q, b)$  folgt  $(p, q, a, b)$ . Jede solche Regel ist ein Hülloperator auf  $S$ , es gilt also (1)  $R$  ist aufsteigend, d. h.  $s$  ist in  $R(s)$  enthalten, (2)  $R$  ist monoton, d. h. ist  $s$  in  $t$  enthalten, so auch  $R(s)$  in  $R(t)$ , (3)  $R$  ist idempotent, d. h.  $R(R(s)) = R(s)$ .

Betrachten wir ein solches System im forward chaining-Betrieb. Hier sei etwa das statistische Problem des Benutzers durch eine Menge  $u$  von als wahr erkannten Aussagen beschrieben. Gesucht sind alle Aussagen, die daraus mittels der Regeln  $R_1, \dots, R_m$  ableitbar sind. Stufenweise kann man das wie folgt lösen: man bilde die Vereinigung aller Mengen  $R_i(u)$  für  $i$  von 1 bis  $m$ . Dies ist eine Abbildung  $H$ , die aufsteigend und monoton, jedoch nicht notwendig idempotent ist. Jetzt iteriere man  $H$  – da man in einer endlichen Menge nur endlich lange aufsteigen kann, ist man nach endlich vielen Schritten fertig und landet bei einem Element  $x = H^*(u)$ , welches offenbar das Wortproblem für den Anfangszustand  $u$  löst. Es ist bemerkenswert, daß man  $x$  auch polynomial schnell berechnen kann: dazu durchlaufe man die Regeln zyklisch in irgendeiner festen Reihenfolge  $1, 2, \dots, m, 1, 2, \dots$ . Man suche die nächste Regel, die eine echte Obermenge des jeweils aktuellen Zustandes  $s$  liefern kann. Eine solche nichttriviale Regel hat zwei Eigenschaften, sie muß auf  $s$  anwendbar sein ( $g$  Teilmenge von  $s$ ) und sie muß vorher noch nie angewendet worden sein (denn nach einmaliger Anwendung ist  $h$  bereits eine Teilmenge von  $s$ ). Diese kann durch lineares Suchen unter den noch nicht angewendeten Regeln gefunden werden. Das Verfahren stoppt, wenn in einem Zyklus  $1, 2, \dots, m$  keine solche Regel gefunden wurde. Andernfalls wird in jedem Zyklus mindestens eine Regel trivial, so daß höchstens  $0.5n(n+1)$  Regelanwendungen erfolgen. Es existiert ein schöner, einfacher Beweis von Herrn Röhmel, daß dieses Verfahren unabhängig von der gewählten Anordnung stets das gesuchte  $H^*(u)$  liefert. So einfach die Anwendung eines Hüllsystems auch ist, gibt es dabei jedoch das prinzipielle Problem der Interpretation von  $M$ . So kann  $M$  ja mit einer Aussage  $p$  auch die Aussage  $\text{non-}p$  enthalten. Damit ist es möglich, daß das Regelsystem inkonsistent wird, d. h. aus einem konsistenten  $u$ , welches nie zugleich  $p$  und  $\text{non-}p$  enthält, ein inkonsistentes  $x$  gewinnt. Ableitungen in derartigen logisch interpretierten Aussagenmengen werden in sog. Deduktionssystemen vorgenommen. Dies sind Übergangssysteme, deren Zustände logische Ausdrücke sind. Eine Inferenzmaschine muß zumindest in der Lage sein, das Entscheidungsproblem des Übergangssystems zu lösen, d. h. für einen gegebenen Ausdruck  $s$  (etwa entsprechend einem gesamten PROLOG-Programm als Konjunktion seiner Fakten und Regeln) zu ermitteln, ob ein zweiter gegebener Ausdruck  $t$  (etwa ein Zielfaktum) aus  $s$  ableitbar ist.

### 4. Komplexitätsprobleme von Inferenzkalkülen

Eine Reihe von tiefliegenden Untersuchungen, die zum Teil bereits aus den dreißiger Jahren dieses Jahrhunderts stammen,

zeigen nun prinzipielle Beschränkungen jeder beliebigen Inferenzmaschine. Kompakt lassen sich diese Beschränkungen durch Angabe der Komplexitätsklasse des Entscheidungsproblems für einen gegebenen Kalkül formulieren. Die Resultate sollen an dieser Stelle nur zitiert und in Hinsicht auf Expertensysteme interpretiert werden. Für die Definitionen und Beweise wird auf die angegebene Literatur bzw. auf eine in Vorbereitung befindliche Ausarbeitung des Verfassers verwiesen – STREITBERG (1987).

Hüllsysteme können das Entscheidungsproblem polynomial schnell lösen. Dieses günstige Komplexitätsresultat wird allerdings durch völligen Verzicht auf jede Semantik erkauft. Erlaubt ein Expertensystem zumindest logische Verknüpfungen (wie Negation, Konjunktion, Disjunktion, Implikation) zwischen den atomaren Aussagen, so entsteht das Entscheidungsproblem der Aussagenlogik. Bekanntlich existieren Kalküle zur Lösung des Entscheidungsproblems für aussagenlogische Ausdrücke, etwa der einfache Resolutionskalkül von ROBINSON (1965). Nach einem grundlegenden Satz von COOK (1972) ist jedoch in jedem Kalkül für die Aussagenlogik das Entscheidungsproblem NP-vollständig. Vereinfacht bedeutet dies, daß in einem Computer mit unbeschränkter Parallelität polynomiale Algorithmen für das Problem realisierbar sind. Es ist dabei unbekannt, ob hier auf die Voraussetzung der Parallelität verzichtet werden kann. NP-Vollständigkeit heißt, daß mit einer polynomialen Lösung für das Problem der Aussagenlogik, auch jedes andere NP-Problem polynomial schnell gelöst werden könnte. Dies geschieht wie folgt: Man übersetzt das vorgelegte NP-Problem (etwa das Travelling-Salesman Problem) polynomial schnell in ein äquivalentes aussagenlogisches Entscheidungsproblem, löst dieses polynomial schnell (ob dieser Schritt möglich ist, ist natürlich unbekannt) und führt die Rückübersetzung wieder polynomial schnell durch. Die praktische Folgerung aus dem Satz von COOK ist, daß ein Expertensystem mit aussagenlogischer Wissensbasis nur für relativ kleine Wissensbasen realisierbar ist, wobei die maximale Größe der Wissensbasis bei linearer Erhöhung der Rechenkapazität nur unwesentlich wächst.

Die Sprachmittel der Aussagenlogik erlauben wichtige Inferenzen nicht, z. B. kann aus »x ist normalverteilt. x-quer ist eine lineare Funktion von x. Lineare Funktionen normalverteilter Größen sind normalverteilt« nicht auf »x-quer ist normalverteilt« geschlossen werden. Dies setzt zumindest die Mächtigkeit der Prädikatenlogik 1. Stufe (PL1) voraus. Nach einem Satz von GÖDEL (1930) existieren Kalküle, d. h. Übergangssysteme für PL1. Das Entscheidungsproblem für PL1 ist damit rekursiv aufzählbar. Dies bedeutet vereinfacht, daß die Menge der Aussagen t, die aus einem gegebenen Ausdruck s folgen, der Reihe nach aufgezählt werden können. Ein bekannter Kalkül ist der Resolutions-Unifikations-Algorithmus. Leider ist PL1 jedoch nicht rekursiv, d. h. die Menge der aus s nicht folgerbaren Aussagen ist nicht mehr rekursiv aufzählbar. Damit ist das Entscheidungsproblem in PL1 nicht mehr entscheidbar, d. h. es existiert kein Algorithmus, der nach endlich langer Zeit bestimmen kann, ob t aus s folgt oder nicht. Dies ist die Aussage des Satzes von CHURCH (1936) und TURING (1937). Die Konsequenz für Expertensysteme liegt auf der Hand: Wenn die Folgerungen des Systems korrekt und vollständig sein sollen, muß der Ausdrucksreichtum von PL1 reduziert werden.

Das bekannteste derartige System ist PROLOG. Für dieses System gilt der Satz von PLAISTED (1984): Das Entscheidungsproblem der Bernays-Schönfinkel Klasse in determinierten Krom- und Horn-Formeln ist PBAND-vollständig. Ohne auf Details einzugehen, bedeutet dies, daß die Implementierung

eines PROLOG-Interpreters, der korrekte und vollständige Ableitungen vollzieht, zwar machbar, aber essentiell härter als alle NP-Probleme ist. Es muß angemerkt werden, daß konkrete PROLOG-Systeme sowohl unvollständig als auch unkorrekt sind. Die Unvollständigkeit drückt sich im Auftreten von unendlichen Zielreduktionsschleifen aus. Ein klassisches Beispiel ist das folgende PROLOG-Programm:

liebt(X,Y) if liebt(Y,X). liebt(paul,paula). liebt(paula,paul)? Je nach den Idiosynkrasien des Interpreters führt dieses Programm oder leichte Variationen davon zum Absturz wegen Stack-Overflow. Die Unkorrektheit ist am klarsten mit Hilfe der Negation zu veranschaulichen:

lustig(X) if not nüchtern(X).nüchtern(peter).lustig(peter)? liefert die erstaunliche Folgerung, daß Peter nicht lustig ist. Dahinter steckt die closed-world Annahme von Prolog, d. h. der klassische logische Fehlschluß des argumentum ad ignorantiam: was ich nicht beweisen kann, ist auch nicht wahr.

M.E. ist die Einschränkung von PL1 auf entscheidbare Teilmengen nicht der richtige Weg, um das Wissen des Statistikers im Rechner abzubilden. Bereits die volle Prädikatenlogik 1. Stufe hat gravierende Ausdrucksschwächen. So kann in PL1 etwa die Menge der natürlichen Zahlen nicht bis auf Isomorphie beschrieben werden, d. h. zu jeder Axiomatisierung der natürlichen Zahlen existieren nichtisomorphe Modelle. Eine amüsante Ausdrucksschwäche beschreibt das Lemma von Ehrenfeucht, nach dem die endlichen Ordnungen geradzahlig Mächtigkeit nicht in PL1 charakterisierbar sind. Die Tatsache, daß der Median aus einer Orderstatistik mit geradzahlig vielen Elementen anders zu berechnen ist als in einer mit ungeradzahlig vielen Elementen, wäre also in einer Wissensbasis ohne attachierte Prozeduren nicht formulierbar. In der Tat operiert der Statistiker andauernd mit Regeln höherer Stufigkeit. In den Regeln »wenn x eine Zufallsvariable ist, so hat x eine Verteilung« oder »wenn x eine Verteilung aus der Klasse der Exponentialverteilungen besitzt, so wende das zu dieser Verteilung gehörende Maximumlikelihoodverfahren an.« treten Existenz und Allquantoren nicht nur über Individuenvariable, sondern über Prädikatenvariable auf. Es handelt sich also zumindest um Ausdrücke der Prädikatenlogik 2. Stufe (PL2) – es ist notwendig, über Eigenschaften von Eigenschaften reden zu können. Sofern dies konzidiert wird, muß jede Umsetzung von statistischem Expertenwissen in Wissensbasen scheitern. Dies ergibt sich aus einem Satz von GÖDEL (1931), nach dem bereits die Menge der in PL2 gültigen Formeln nicht mehr rekursiv aufzählbar ist.

Warum existieren also keine statistischen Expertensysteme? Sofern die Wissensbasis überhaupt noch entscheidbar sein soll, müssen die Ausdrucksmöglichkeiten so drastisch beschränkt werden, daß statistisches Expertenwissen de facto nicht mehr darstellbar ist. Wer sich jemals an der Wissensrepräsentation auch nur eines Teilgebiets der Statistik versucht hat, wird den Korsettcharakter der zulässigen Sprachelemente bestätigen können. Versucht man andererseits die Sprachmächtigkeit so zu erhöhen, daß auch nur die einfachsten Schlußweisen des statistischen Experten formuliert werden können, stößt man an die prinzipiellen Schranken der Entscheidbarkeitstheorie. Es existieren dann keine zulässigen Inferenzmaschinen mehr.

## 5. Konklusion

Die erste Welle der KI in den 50er und 60er Jahren scheiterte an der Nichteinlösung ihrer Versprechungen für die Praxis. Das »crucial experiment« war damals die Ankündigung von

Systemen zur automatischen maschinellen Übersetzung von Texten – ein Problem, das heute genauso ungelöst ist wie damals.

In der zweiten KI-Welle, deren Euphoriemaximum wir in Europa momentan erleben (in den USA scheint mir die Euphorie bereits abzuebben), werden ähnlich grandiose Versprechungen gemacht. Das Zauberwort heißt heute Expertensystem. Ich erlaube mir, die Vermutung zu formulieren, daß auch in der laufenden KI-Welle der Kaiser keine neuen Kleider bekommen wird.

Dennoch steckt hinter der Arbeit an Expertensystemen ein durchaus realer Bedarf. Statistische Expertensysteme werden m. E. so stark diskutiert und gefördert, weil die existierende statistische Software zumindest zwei prinzipielle Defizite aufweist. Beide Defizite lassen sich aufheben, wenn auch wohl eher mit natürlicher denn mit künstlicher Intelligenz.

Das erste Defizit liegt in der Statistik selbst. Betrachtet man die geplanten Einsatzbereiche von Expertensystemen, so lassen sich Systeme im Bereich der Konfirmatorik und Versuchsplanung und Systeme im Bereich der explorativen Analyse unterscheiden. In beiden Bereichen steht im Vordergrund das Problem der Auswahl einer statistischen Strategie, einmal problem- und designgesteuert, das andere Mal datengesteuert. Während die statistische Theorie eine Vielzahl von rationalen Kriterien zur Beurteilung und Auswahl von Einzelverfahren zur Verfügung stellt (Unverzerrtheit, ARE, Beschränktheit der Influenzfunktion etc. etc.), existiert keine Theorie für die Auswahl ganzer Strategien, d. h. Abfolgen, Verzweigungen und Iterationen von Einzelverfahren für eine konkret gegebene Studie. Die ersten Ansätze zeigen jedoch, daß das Problem einer mathematischen Behandlung zugänglich ist. Ich verweise nur auf zwei herausragende Beispiele: Die Präzisierung des Niveauproblems multipler Tests durch SONNEMANN u. a. und die Untersuchung der Konsistenz von Modellselektionsverfahren durch HANNAN u. a. Das Strategiewahlproblem wird m. E. nicht durch die Anhäufung von immer mehr diffusen Regeln in Wissensbasen, sondern durch die kreative Arbeit des Mathematikers am Schreibtisch entschieden werden.

Zum zweiten haben die existierenden statistischen Programmpakete erhebliche Mängel, was ihre Benutzeroberflächen betrifft. Der Bereich des bereits heute möglichen wird auch nicht annähernd ausgenutzt, da auch in den PC-Versionen noch die Menüführungstechniken und Kommandooberflächen der alten Großrechnersysteme dominieren. Warum

existiert etwa kein graphisches System mit Ikonen für die wichtigsten statistischen Verfahren, etwa gerichtete Graphen für Systeme vom LISREL-Typ, deren Knoten vom Benutzer mit seinen Variablen gefüllt werden können? Warum werden aus den existierenden Kommandooberflächen nicht vollgültige funktionale Programmiersprachen, deren Grundfunktionen (»Prozeduren von SPSS«) auf definierten Datenobjekten operieren und als Resultat ebenso definierte Datenobjekte liefern, die dann wieder den Input anderer Funktionen darstellen können. Ansätze für derartige Systeme existieren bereits (S oder ISP) und lassen sich in verfügbaren Programmierumgebungen wie APL2 leicht erweitern und verallgemeinern. Auch hier ist weniger ein Expertensystem als die Kreativität des Informatikers gefordert.

## Literatur

- CHURCH, A. (1936): A note on the Entscheidungsproblem. *Journal of Symbolic Logic* **1**, p. 40, 100.
- COOK, S. A. (1972): The complexity of theorem-proving procedures. *ACM Symposium on Theory of Computing* 151–158.
- GALE, W. A. (1986) ed: *Artificial Intelligence and Statistics*. Addison-Wesley 1986.
- GÖDEL, K. (1930): Die Vollständigkeit der Axiome des logischen Funktionskalküls. *Monatshefte für Math. und Physik* **37**, p. 349–360.
- GÖDEL, K. (1931): Über formal unentscheidbare Sätze der Principia Mathematica und verwandter Systeme. I. *Monatshefte für Math. und Physik* **38**, p. 173–198.
- HAUX, R. (1986) ed: *Expert Systems in Statistics*. Gustav Fischer 1986.
- PLAISTED, D. A. (1984): Complete problems in the first order predicate calculus. *Journal of Computer and System Sciences* **29**, p. 8–35.
- ROBINSON, J. (1965): A machine-oriented logic based on the resolution principle. *Journal of the ACM*, **12**, p. 23–41.
- STREITBERG, B. (1987). On the non-existence of expert systems. In preparation.
- TURING, A. M. (1937): On computable numbers, with an application to the Entscheidungsproblem. *Proc. London Math. Soc.* (2) **42**, p. 230–265. A correction *ibid.* **43**, p. 544–546.

Eingegangen am: 15. April 1987.

Anschrift des Verfassers: Prof. Dr. Bernd Streitberg, Universität Hamburg, Institut für Statistik und Ökonometrie, Von-Melle-Park 5, D-2000 Hamburg.

HARMON, P., und DING, D.

**Expertensysteme in der Praxis**

Perspektiven, Werkzeuge, Erfahrungen  
1986, 315 S., DM 68,-  
R. Oldenbourg Verlag, München - Wien

Die vorliegende Übersetzung aus dem Englischen gibt einen umfassenden Überblick über Anwendungen von Techniken der Künstlichen Intelligenz. Verständlich und ohne zuviel technischen Ballast sind Konzepte dargestellt, die Expertensystemen zugrunde liegen. Ebenso wird auf bereits im Einsatz befindliche Systeme eingegangen.

Durch die vielen Hinweise, die z.T. in der Art einer Checkliste gegeben werden, ist dieses Buch insbesondere als Einstieg in dieses zukunftssträchtige Arbeitsgebiet geeignet. Ge.

WOLF, TH., und UNKELBACH, H. D.

**Informationsmanagement in Chemie und Pharmazie**

Seine Bedeutung in techn.-wiss. Bereichen  
1986, 244 S., DM 34,-  
B. G. Teubner, Stuttgart

Die Entwicklung der PCs hat vielfach zu „Insellösungen“ in den Betrieben und Instituten geführt. Es muß den Autoren hoch angerechnet werden, daß sie schon heute auf die Notwendigkeit von gesamtheitlichen Lösungen hinweisen. Es ist zu hoffen, daß diese ausgezeichnete Darstellung zu einem bewußten Umdenken führt, um die im Augenblick schon vorhandenen technischen Möglichkeiten auch möglichst bald voll zu nutzen Ge.

PFLUG, G. CH. (Hrsg.)

**Neuere Verfahren der nichtparametrischen Statistik**

Med. Informatik und Statistik, Bd. 60  
1985, 129 S., DM 29,50  
Springer-Verlag, Berlin - Heidelberg - New York - Tokyo

Dieser Band enthält sieben Beiträge zum Themenkreis „Nichtparametrische Statistik“, die im Rahmen des Biometrischen Seminars 1985 der Region Österreich-Schweiz der Biometrischen Gesellschaft in Graz vorgetragen wurden. Es handelt sich dabei neben einem Übersichtsreferat um Referate über: Adaptive verteilungsfreie Tests, Multiple Vergleiche mittels Rangtests, Adaptive Verfahren für Wachstumsdaten, Analyse von Verlaufskurven und Diskrimination und Klassifikation von Verlaufskurven.

BAUER P., SCHREIBER, V. und WOHLZOGEN, F. X.

**Sequentielle statistische Verfahren**

1886, 162 S., DM 48,-  
G. Fischer Verlag, Stuttgart-New York

Die Buchreihe „Biometrie“ behandelt Teilgebiete der angewandten Statistik in abgeschlossenen Einzeldarstellungen. Es ist dies ein lobenswerter Ansatz, da so in vertretbarer Zeit und mit einem verarbeitbaren Umfang diese Bände herauskommen.

Speziell für sequentielle Verfahren fehlte bislang eine entsprechende deutschsprachige Darstellung. In dem vorliegenden Band werden behandelt: sequentielle Prüfverfahren bei Häufigkeitsdaten, sequentielle Prüfverfahren bei Meßdaten, sequentielle Schätzverfahren und sequentielle Versuchsplanung. Dabei folgen jeweils einer allgemeinen verständlichen Einführung durchgerechnete und meist auch graphisch dargestellte Beispiele, die durchweg der Praxis entstammen.

Da speziell sequentielle Verfahren oft zu einer beträchtlichen Reduktion des Versuchsumfanges führen, sollte dieses Buch eigentlich in die Hand eines jeden Biometrikers gehören. Ge

GRABNER, G. (Hrsg.)

**WAMIS**

**Wiener Allgemeines Medizinisches Informations-System**

Med. Informatik und Statistik Bd. 59  
1985, 367 S., DM 58,-  
Springer-Verlag, Berlin-Heidelberg-New York-Tokyo

Vor über 10 Jahren, im Jahre 1975, wurde den Universitätskliniken und den Instituten der Medizinischen Fakultät in Wien ein integriertes Krankenhausinformationssystem übergeben. In diesem System wurden bisher Daten und Informationen von mehr als 700 000 Patienten mit über 1,2 Millionen Interventionen im direkten Zugriff gespeichert. Über die Möglichkeiten des Wiener Allgemeinen Medizinischen Informations-Systems WAMIS, das mit einigen seiner Module auch der unmittelbaren Krankenbetreuung dient und das sich im Alltag mehr als bewährt hat, wird in der vorliegenden Publikation ausführlich berichtet. Die Veröffentlichung mag mit dazu dienen, daß dieses System auch an anderen Stellen angewandt wird. Ge.

WEBER, E.

**Grundriß der biologischen Statistik**

Anwendungen der mathematischen Statistik in Forschung, Lehre und Praxis  
9., durchgesehene Auflage  
1986, 652 S., DM 62,-  
Gustav Fischer Verlag, Stuttgart

Es gibt immer nur wenige Lehrbücher, die sich mit der Zeit auch als Standardwerk durchsetzen. Die „Erna Weber“ gehört aber im Bereich der Biometrie ganz sicher dazu. Dies zeigt, daß hier Verfahren und Methoden so klar dargestellt werden, daß insbesondere die Anwender und Studenten immer wieder zu dieser umfassenden Einführung greifen, die durch die Neuauflagen auch immer wieder ergänzt wurde. Möge diese ausgezeichnete Darstellung uns noch möglichst lange zur Verfügung stehen. Ge.

APPELT, H. und STRAUSS (Hrsg.)

**Ergebnisse einzelfallstatistischer Untersuchungen in Psychosomatik und klinischer Psychologie**

1985, 195 S., DM 48,-  
Springer-Verlag, Berlin-Heidelberg-New York-Tokyo

Einzelfallanalysen haben in der Psychologie eine lange Tradition. Neuere Anwendungsbeispiele stehen im Mittelpunkt des vorliegenden Buches. Dabei wird in den Beiträgen des ersten Abschnittes fast die ganze Palette uni- und multivariater Zeitreihenanalysen angewandt.

Im zweiten Teil wurden einige methodische Arbeiten aufgenommen, die spezifische Fragen behandeln und alternative Ansätze zeigen und gegenüberstellen. Ge.

BUSCHLINGER, E.

**Software-Entwicklung mit UNIX**

1985, 277 S., DM 38,-  
B. G. Teubner, Stuttgart

Die Entwicklung von Betriebssystemen läßt sich zwar nur schwer vorhersehen, doch im Augenblick scheint es so, daß UNIX in der nächsten Zeit eine herausragende Rolle einnehmen wird. Daher ist es zu begrüßen, daß in einer verständlichen und klaren Darstellung UNIX erläutert und insbesondere die Softwareentwicklung mit UNIX, d. h. mit den verfügbaren Softwarewerkzeugen behandelt wird. Die Lektüre dieses Buches kann für den „Einstieg“ bestens empfohlen werden. Ge.

SCHNEIDER, H.-J. (Hrsg.)

**Lexikon der Informatik und Datenverarbeitung**

2. Aufl., 1986, 750 S., DM 98,-

R. Oldenbourg Verlag, München - Wien

Ein Kennzeichen für die Güte und für die Notwendigkeit für dieses Buch ist die Tatsache, daß nach knapp drei Jahren jetzt eine verbesserte und erweiterte 2. Auflage vorliegt.

Stellvertretend sei ein Stichwort herausgegriffen:

»Information: Kenntnis über bestimmte Sachverhalte und Vorgänge in einem Teil der wahrgenommenen Realität«.

Mit mehr als 6500 Einträgen ist hier auf überschaubarem Platz Information zusammengestellt, die in vielen Fällen zur Begriffserklärung in diesem noch so jungen Gebiet beitragen kann.

Ein Standardbuch für alle Tätigen im Bereich der Informationsverarbeitung. Ge.

HAUX, R. (Ed.)

**Expertsystems in statistics**

Selected papers from a workshop organized by the working group »Computational Statistics« of the German Region of the International Biometric Society

1986, 194 S., DM 48,-

Gustav Fischer Verlag, Stuttgart-New York

Statistische Expertensysteme gewinnen immer mehr an Bedeutung, wobei im Augenblick die möglichen Vorteile oft recht kontrovers diskutiert werden. In den ausgewählten Beiträgen des im Dezember 1985 durchgeführten Workshops wurde versucht, mehr grundsätzliche Probleme zu behandeln sowie über erste Versuche zu berichten. Damit liegt - wenn auch in englischer Sprache - eine gute Zusammenfassung über den augenblicklichen Stand dieses zukunftsträchtigen Gebietes vor. Ge.

SCHUBÖ, W. und UEHLINGER, H.-M.

**SPSSx/Handbuch der Programmversion 2.2**

Autorisierte deutsche Bearbeitung des SPSSx User's Guide

1986, 659 S., DM 58,-

Gustav Fischer Verlag, Stuttgart-New York

Die Verbreitung von SPSS als Datenmanipulations- und -auswertungsprogramm geht nicht zuletzt darauf zurück, daß schon früh auch in deutscher Sprache ein brauchbares Handbuch verfügbar war. Von daher ist es nur logisch, daß für die permanent vorgenommenen Änderungen (neue Versionen) auch immer wieder Neuauflagen der Handbücher erscheinen.

Die jetzt vorliegende zweite Auflage des SPSSx-Handbuches berücksichtigt die Änderungen und Erweiterungen, die die Version 2.1 und 2.2 gegenüber Version 2.0 gebracht haben. Dazu gehören neben vielen kleineren Verbesserungen vor allem SAS- und OSIRIS-Schnittstellen sowie Prozeduren für Ähnlichkeitskoeffizienten und multidimensionale Skalierung. Ge.

ESSL, A.

**Statistische Methoden in der Tierproduktion**

1987, 316 S., DM 58,-

Österreichischer Agrarverlag, Wien

Die Anwendung biometrischer Methoden nahm in den verschiedenen Bereichen der Nutztierwissenschaften in den letzten Jahrzehnten vor allem wegen der rasanten Entwicklung auf dem Sektor der elektronischen Datenverarbeitung ständig zu. Damit ergab sich aber auch die Notwendigkeit, die Lehrbücher den geänderten Anforderungen anzupassen. So stehen heute nicht mehr so stark Rechenverfahren im Vordergrund, da diese als Algorithmen in den Statistikprogrammen enthalten sind. Auch in dem vorliegenden Buch versucht der Autor wieder das richtige »statistische« Denken in den Mittelpunkt zu stellen. Durch die klare Darstellungsart ist dies auch sehr gut gelungen. Die gut gewählten Beispiele unterstützen diese Linie sehr. - Insgesamt liegt hier ein wirklich empfehlenswertes Statistiklehrbuch vor. Ge.

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

**EDV-Recht**

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

Herausgegeben im Auftrag der Interparlamentarischen Arbeitsgemeinschaft.

Ergänzbare Ausgabe, einschl. 51. Lieferung, 450 Seiten und 3 Ausschlagtafeln, DIN A5, DM 164,-, zuzüglich 3 Spezialordner je DM 11,80. Ergänzungen folgen von Fall zu Fall.

Erich Schmidt Verlag, Berlin-Bielefeld-München

Der Textteil dieses umfassenden Grundlagenwerkes enthält die einschlägigen Rechtsvorschriften, Dokumente, 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 umfaßt Kommentare zum Bundesdatenschutzgesetz sowie zu den Vertragsbedingungen für EDV-Anlagen und -Geräte.

Durch diese Erläuterungen wird die komplizierte Materie des EDV-Rechts auch Nichtjuristen verständlich gemacht. Ge.

ARNOLD, F. (Hrsg.)

**ISDN: Viele Kommunikationsdienste in einem System**

1987, 335 S., DM 80,-

Verlagsges. R. Müller, Köln

ISDN (Integrated Services Digital Network) ist das zukünftige dienstintegrierende digitale Nachrichtennetz. Wenn man bedenkt, daß im Augenblick schon Pilotanwendungen laufen, dann ist es erfreulich, daß mit der vorliegenden Veröffentlichung eine Zusammenstellung über die Philosophie von ISDN, die zu erwartenden Konsequenzen für andere Dienste, die Möglichkeiten und Probleme aus nationaler und internationaler Sicht verfügbar ist. Da ISDN in absehbarer Zeit als Kommunikationsnetz im Einsatz sein wird, ist diesem Buch zum Verständnis dieser neuen Technik eine möglichst weite Verbreitung zu wünschen. Ge.

RAPPELSBERGER, P., PFUNDNER, P. u. GELL, G. (Hrsg.)

**Medizin-Technik - Medizinische Informatik '86**

1986, 434 S., DM 62,-

R. Oldenbourg Verlag, Wien-München

Der vorliegende Band enthält die Beiträge der 11. Jahrestagung der Österreichischen Gesellschaft für Biomedizinische Technik sowie des parallel veranstalteten Symposiums »Medizinische Informatik '86«.

In den Referaten wurde ein Überblick über den augenblicklichen Stand der Entwicklungen gegeben, wobei im Bereich der Med. Informatik insbesondere Expertensysteme, die medizinische Dokumentation sowie Krebsregister angesprochen wurden. Insgesamt wurde aber bei fast allen Referaten auf die Anwendung von Computern eingegangen. Ge.

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

**Bundesdatenschutzgesetz (BDSG)**

Ergänzbarer Kommentar nebst einschlägigen Rechtsvorschriften

Ergänzbare Ausgabe, einschl. 21. Lieferung

1286 Seiten, DIN A5, DM 86,-, zuzüglich Ordner

DM 11,80. Ergänzungen folgen von Fall zu Fall

Erich Schmidt Verlag, Berlin-Bielefeld-München

Das Bundesdatenschutzgesetz (BDSG) gewinnt in der Praxis und in der Rechtsprechung immer mehr an Bedeutung. Die vorliegenden Lieferungen 20 und 21 werten die jüngste Rechtsprechung, insbesondere auch die des Bundesgerichtshofes, auf. Dieser Kommentar sollte in keinem Unternehmen und in keiner Verwaltung fehlen und gehört in die Hand eines jeden Datenschutzbeauftragten. Ge.