Jan H. van Bemmel[1], Alexa T. McCray[2]

[1] Department of Medical Informatics, Erasmus University Rotterdam, Rotterdam, The Netherlands.
[2] National Library of Medicine, Bethesda, Maryland, USA.

Introduction

Synopsis of Scientific Contributions

Health and Clinical Management

It is undeniable that after a long period in which primarily administrative patient data were stored in computers, the time has come for patient records to be structured for electronic processing. Several papers in the Yearbook deal with this subject, starting with the article by Rector et al. in section 1 of the Yearbook, and ending with the paper by Van der Lei et al. in section 5. Very early in the history of Medical Informatics, first attempts were made in the USA to document patient data in computers by, e.g., Collen [1], Barnett [2], McDonald [3] and Warner [4], to mention a few; but in the last few years, since PCs have become cheaper and user interfaces more friendly, such attempts appear to be successful and have gained momentum.

In Europe, the Medical Informatics Group from Manchester, UK, has invested much time and effort in developing the basis for electronic medical records. The first article in the Yearbook describes the Pen & Pad method by Allan Rector et al. Successful implementations of full-fledged electronic patient records for primary care can now be seen in a great number of routine practices in the United Kingdom and The Netherlands. For many models of computer-based patient records the Problem-Oriented Medical Record (POR) developed by Larry Weed [5] forms the background, although for some implementations the POR appears to offer practical problems. It is fascinating to see that medical record systems are coming into real use, and that in the USA it is recognized that a new era in the use of computer-based patient data for a host of different applications has begun; this is apparent from the foundation of an Institute for Computer-based Patient Records, based on a study by the Institute of Medicine [6].

The articles by Jick *et al.* and Johnson *et al.* assess the value of computer-stored patient data, as compared with manual documentation, and the use of such data for population surveillance, respectively. It is reassuring to observe that the diagnosis based on computerized patient data is comparable (the agreement was 87%) with that based on conventional records. Jick *et al.* widen their conclusion by stating that *"computer records of the general practitioners who participated in this study are satisfactory for many clinical studies."* Several research projects have now been started to investigate the use of routinely collected patient data, e.g., for monitoring the population. The transition from surveillance to prevention is then only a gradual one. Johnson *et al.* conclude from their study that the use of such data is very promising, although data collection at the source should be less haphazard and better structured.

The use of computers for prevention and the generation of patient reminders is the subject of the article by Ornstein *et al.* The computer-generated reminders were compared with conventional, physician-generated reminders. It appeared that computer-generated reminders improved the adherence to preventive services in primary care for a variety of tests, from cholesterol measurements to Papanicolaou smears. The authors state that settings without computer-based medical records will be unable to institute such interventions. Citing a report from the US Preventive Services Task Force, the authors conclude that the recommendations in that report are ideally suited to computer-based screening and prompting systems.

Now that parts of patient records are collected by different medical specialists, both cross-sectionally and longitudinally, it is of importance that such different records can be linked, even when no complete and reliable patient identification is available. This is the subject of a paper by Wajda *et al.*, of which the first part by Roos *et al.* appeared in an earlier issue of *Methods of Information in Medicine*. Several approaches and probabilistic algorithms are described by the authors and success rates for record linkage are reported. Especially for epidemiological purposes, such as for AIDS studies, these models are of great use to increase the completeness of medical data. Successful matches are on the order of 80% or higher, dependent of course on the reliability of the source material.

Practice of the nursing profession varies in many countries. Areas in the world where nurses have a recognized and autonomous professional profile include, especially, the USA and Canada. For that reason it is not surprising that most work on the standardization of nursing data was performed on the North-American continent, where nurses have their own well-defined responsibility in collecting patient data, making a nursing diagnosis, and providing nursing care. For medical purposes, several projects were carried out in the past, both in Europe (e.g., by Roger *et al.* [7]) and the USA to define a Basic Medical Data Set. The article by Werley *et al.* describes a first attempt to define a Nursing Minimum Data Set (NMDS) for the documentation of the most essential nursing data. The authors foresee that such NMDSs will be used for nursing care, for research and education, and also for health policy decision making. It is correctly stated in this article that such records are also of great value for the evaluation of services. Now that the integration of systems in medicine and health care is no longer a dream, it is hoped that for the benefit of the patient, medical and nursing records can also complement each other - maybe, in the future we will be able to speak of true patient records instead of separate medical and nursing records.

A most important aspect of the use of computer-stored patient data is the prevention of adverse drug events (ADEs). This is reported in the contribution by Classen *et al.* in a study on such events in hospital patients. The authors have studied the prescription of drugs, documented in a hospital information system, in circumstances where the prescription of drugs was changed, increased or decreased, or stopped. They related these events with laboratory values. An impressive number of almost 37,000 patients was followed in this study in which over 1% ADEs were identified. The authors offer solid proof (they speak of a *markedly increased detection of ADEs*) that computerized monitoring of ADEs in hospitals, using computer-stored patient data, offers many advantages over voluntary reporting of such events.

The subject of the eighth article, by Tsevat *et al.*, differs the most from those of the others in this section, but is of no less importance. It forms an essential foundation for many preventive interventions such as the one reported by Ornstein *et al.* The study by Tsevat *et al.* discusses the impact of risk factors such as a high blood pressure or increased serum cholesterol, smoking, or high body weight, on the life expectancy of patients with coronary heart diseases (CHDs). The outcome of such heavily computer-supported studies has a large impact on health policies and costs of health care. Before health care planners introduce preventive measures in the population and before investigators develop computer-supported systems for prevention, it is wise to weigh carefully the results of studies such as the one by Tsevat *et al.* A surprising result is that even when - theoretically - CHDs were totally prevented, the increase in life expectancy would only be a mere three years for a 35-year-old adult. Single-risk prevention would result in life prolongation of 0.2 - 1.1 years. Researchers in medical informatics, together with their medical colleagues, would be wise to balance such results against the implementation of preventive modules in the systems they develop; this is, in similar words, also the conclusion of the authors.

Clinical Support Systems

This section deals with systems that have a direct impact on patient care. The first four articles (by Lerou *et al.*, Gelsema *et al.*, Popp *et al.*, and Dawes *et al.*) are allrelated in one way or the other to computer-supported intensive care monitoring. A subsequent article, by Kudenchuck *et al.* reports on the usefulness of computerized ECG interpretation for thrombolytic therapy, possibly preceding CCU monitoring. The following two publications are on the use of computers to assist people with physical handicaps, dysarthric speech and impaired vision, respectively. The last article by Okajima *et al.* is on the quality assessment of portal images for the preparation of radiation therapy.

In the past, and especially by biomedical engineering groups, much research was done on the development of models and systems to support the monitoring of the severely ill. The early publications by Ty Smith [8] or Sheppard [9], to mention only a few, already indicated the usefulness of computer-supported patient monitoring systems. The measured parameters in such systems were based on physiological models, also implemented on computers. Closed-loop peri- or post-surgical monitoring was primarily applied to increase the quality of care and to decrease the number of patient-days spent in a highly intensive care unit. The study by Lerou *et al.* builds on those early experiences and combines the use of a model approach with a monitoring system for closed-circuit inhalation anesthesia. It is a nice example of the combination of physiological principles, control engineering, and medical informatics.

The contribution by Popp *et al.* on a computer simulator of the circulation for knowledge acquisition in anesthesia, could also have been categorized under section 5, Knowledge-based Systems. It points out that several systems developed today also contain knowledge processing components. The purpose of the study reported here was to develop a decision-support system to generate intelligent alarms during anesthesia of patients undergoing cardio-surgery. Similar to the publication by Lerou *et al.*, this study is also an excellent illustration of the combined usage of multi-disciplinary knowledge in physiology, engineering and medical informatics. The simulation of part of the circulation was used as a tool for the development of a decision-support system, which is characterized by the authors as a "*deep knowledge source*".

A study which has direct impact on the interpretation of biochemical data for the monitoring of patients in intensive care units is reported by Gelsema *et al.* The main message of the article is that all major chemical parameters, such as pH, pCO_2, and base excess, should be used in a coherent, three-dimensional manner to predict the acid-base metabolism of severely ill patients. The authors correctly state that "normal" values for certain patient groups differ significantly from normal values derived from normal people so that, in other words, each specific patient group should have its own standard reference values. Now that large collections of biochemical data are available in hospital or laboratory computers, it is the proper time to derive such reference values for different groups of patients. This issue of using routinely collected data for research is also addressed in the article by Safran in section 6 of this Yearbook.

Evaluation of computer-supported monitoring is of utmost importance. In that respect, in the last few years much on-going debate, especially in the USA, has been based on the outcome of technology-assessment studies, with regard to the value of perinatal monitoring. Since the early systems of the mid-Sixties, computerized fetal monitoring is now in common use in most obstetric clinics and often expanded with computerized analysis of fetal heart rate (FHR) patterns. Since the beginning of the development of these FHR-based monitoring systems, the group of Geoffrey Dawes in Oxford has been active in this field. For that reason it is interesting to see how researchers with such a long tradition now look at the results they have obtained thus far. The most striking message from this publication is that presentation of the "raw" computerized FHR patterns is hazardous if it is not accompanied by rigorous, computer-supported interpretation. This should, in turn, then be based on statistical analysis of a large collection of FHR registrations. It is clear that such large studies are still to be done, in the same manner as was carried out for computerized ECG interpretation systems, reported by Willems *et al.* in section 4.

For several years now, many patients with impending infarction are given thrombolytic therapy. In such circumstances it is time that counts - the earlier the treatment is given, the better - and for that reason such therapy is given even when the disease status of the patient has not yet been properly assessed. For instance, in several countries this therapy is already given by paramedics when the ambulance arrives to bring the patient to the nearest hospital and coronary care unit. Such thrombolytic therapy, however, is expensive, and when given incorrectly there is a certain risk for the patient of side effects such as bleeding. For those reasons, it is important to assess by ECG analysis whether the patient shows ST-depressions. Kudenchuck *et al.* have analyzed the contribution of computer-interpreted ECG analysis for the proper administration of the therapy. The authors have shown in their study that computer-supported indication for thrombolytic therapy was feasible and highly specific. They recommend the evaluation of systems other than the one they assessed themselves - a study that could follow the same lines of assessment as reported in the article by Willems *et al.*

The next two publications in this section deal with the support of persons with a physical impairment; the one by Deller *et al.* on the support of people with dysarthric speech, and the other by Peli *et al.* on image enhancement for visually handicapped persons. Partly because computers are now very powerful and reasonably cheap, speech recognition has gained much more attention than in the past. For clinical applications, such as radiological reports, speech recognition is being used in experimental settings. It is most interesting that similar systems are being developed for communication with persons with severe speech and motor disabilities. Deller *et al.* developed a hidden Markov model and applied automated clipping of transitional speech acoustics, to enhance speech recognition. Speech is digitized by a 12-bits A-D converter operating at 10 kHz on low-pass (4.5 kHz) filtered speech. The paper describes a fascinating application of signal processing and pattern recognition. Recognition of certain vocabularies was on the order of 90% or higher, which appears to be a significant improvement. Undoubtedly, application of these methods is also useful for many other tasks in which speech is used, e.g., as input to computers.

The article on support for the visually impaired by Peli *et al.* uses similar techniques for images, but intended for the visually handicapped persons themselves: digital contrast enhancement by either adaptive filtering or adaptive thresholding techniques. This is realized by using computers in combination with video images. The method was evaluated for a variety of experimental set-ups. Patients had a statistically improved recognition rate of images, such as of human faces. Proof of this improvement was presented by the authors in the form of ROCs. It is encouraging that in this way advanced techniques in signal and image processing and pattern recognition are also available for persons with impairment in speech or visual communication.

The last publication in this section is devoted to a subject which is different from the others; the study by Okajima *et al.* has compared portal images with and without contrast enhancement, in some way comparable to the study by Peli *et al.*, but for an entirely different purpose, i.e., the preparation of patients for radiation therapy. In that sense, this article could also have been categorized under the next section on image and signal processing. In radiotherapy it is important to verify the exact location of radiation beams. In this study this is done by using the radiation beam itself, so that the field can be properly adapted to the area that should be radiated. It appears that the image quality of portal images - which is quite different from that of usual X-ray images - is of great importance for this beam adjustment and the correction of setup errors. ROC curves were used to evaluate the contribution of image enhancement. The authors conclude that proper preprocessing of images improves the subjective assessment of beam adjustments for radiation therapy. Adaptive histogram equalization appears to be one of the best models to obtain this goal.

Image and Signal Processing

One of the earliest areas of the application of laboratory computers is biomedical engineering. Signal analysis and image processing were originally done on analog or hybrid computers or special purpose processors, but at present such tasks are almost solely done on the powerful, reasonably priced workstations of today. Only 10 years ago, most image processing was done for two-dimensional (2-D) pictures, but nowadays 3-D reconstruction, analysis, and boundary detection is performed by many research teams. Clinical applications of 3-D reconstruction are available for diagnostic and surgical purposes. Signal analysis also is easily done on PCs, even in real time, as was also apparent from the article by Deller *et al.* in the preceding section and the one by Thorpe *et al.* in this section. The first three articles on Image and Signal Processing report on 3-D reconstruction of vascular structures, one by Garreau *et al.* derived from biplane angiographic projections, one by Vesely *et al.* from high-definition casts, and one by Tanioka et al. on 3-D reconstruction of the inner ear from MRI. On the condition that the resolution of the 2-D images is large enough, 3-D computer reconstruction offers the possibility to process the reconstructed images for any purpose. After applying 3-D boundary detection it is then equally possible to obtain quantitative parameters from the images.

The publication by Garreau *et al.* is an elegant example of a methodology for the reconstruction of coronary arteries from 2-D biplane X-ray images. The disadvantage of biplane angiography is, of course, that certain assumptions must be made about the 3-D structure of the ventricles, for which it is often usual to assume either a cylinder or an ellipsoid. A novel approach in this study is the use of prior knowledge (with 30 rules and 130 predicates, programmed in Prolog) for this reconstruction regarding the structure of the coronary tree. As the authors state, their approach is based on a deep qualitative model, a fascinating combination of knowledge processing and image analysis.

Vesely *et al.* use high-quality, i.e., with high contrast and resolution, sectional images for 3-D reconstruction. These images were photographed from the milled surfaces of the cardiac valves, at 0.2 mm distances and cast in epoxy. The spatial resolution of the images was on the order of 0.1 mm and all images were stored on video tape before further digital processing. The images were edited and segmented by contour detection to obtain the different structures. The preprocessed and reconstructed image data are then used for further quantitative analysis of the vascular structures.

High-resolution imaging is still a challenging area in both CT scanning and MR imaging. To detect malformations or anatomic deviations caused by disease, Tanioka *et al.* applied MRI to the temporal bone of the inner ear. Since such deviations are difficult to see in 2-D, they constructed 3-D images. Because it is difficult to distinguish between cerebrospinal fluid, white matter, and fatty tissue, two sets of imaging parameters were used to enhance the contrast between these substances. Scans were made in three directions for a field of 15 cm, with projections in the axial, coronal, and sagittal planes. In this way voxels as small as 0.58x0.58x1.00 mm^3 could be computed, which is impressively small. Images were offered to the ENT physicians in a quasi-cinematographic, stereoscopic mode.

A further example of image processing in cardiology is the study by Feng *et al.*, which combines image processing with fuzzy reasoning. The input data are 2-D echocardiographic images of the ventricles in which the endocardial and epicardial boundaries are to be detected. Ultrasound images, however, are blurred and noisy and part of the data is easily dropped out. Global knowledge required for training the algorithms for the proper detection of the boundaries is derived from experts and represented as fuzzy linguistic descriptions and relations. It is striking that the trained human eye (and brain) is able to detect rather easily the ventricular boundaries in moving echocardiographic images. Although color helps to see the direction of flows, structures are best seen in images with only gray values. It is very complex to develop straightforward image processing algorithms for boundary detection in such blurred images. For that reason it is encouraging to see that the combination of image processing and expert knowledge was successfully applied in the studies by Garreau and Feng *et al.*

Tahoces *et al.* start their article by saying that the majority of X-ray pictures consists of chest radiographs. Now that radiological departments are rapidly adopting digital techniques such as CT, MRI, DSA, PACS, etc., for image acquisition and archiving, it is no surprise that conventional radiographic images are also digitally processed. The first step in such processing is image enhancement. In the study reported here, standard chest and breast (mammograms) X-ray pictures were digitized as 512x512 8-bits images and digitally processed. The contrast and edges were enhanced by the combination of two smoothed images and the original picture, and a non-linear transformation. The pictures appear to be sharper for diagnostic interpretation; especially micro calcifications in mammograms are better visualized. However, no diagnostic assessment was done, e.g., by ROC analysis, as in Okajima's study in section 2. The processing takes a minimum amount of time so that it is well suitable for fully automatic routine implementation.

The study by MacMahon *et al.* has also investigated the digital processing of chest images. It is realistic to state that routine chest X-ray pictures require a 2-D resolution of at least 2048x2048 pixels of 10 bits each, if no zooming-in is applied. For that reason it makes sense, even in the present era of cheap mass storage devices - magnetic or optical -, to investigate the possibility of data compression and its effect on diagnostic outcome. The data compression algorithm that was used is based on so-called adaptive block cosine transform coding, and noncompressed images were diagnostically compared with a 25:1 and a 50:1 data compression. The diagnostic assessment was done by 12 radiologists. It appears that a pneumothorax is still best diagnosed by uncompressed images, although a 25:1 compression gave only slightly less diagnostic performance, given the high inter-observer variation (which is known already from many diagnostic assessment studies in radiology, such as the ones by Lodwick [10]). For the detection of interstitial infiltrates all images, compressed or not, offered comparable results. The article offers an interesting method of how to assess the diagnostic differences between various techniques, although the authors did not comment on how to improve the large inter-observer variation, for instance by using a Delphi technique, as was done in the study by Willems *et al.* in section 4.

Almost microscopically small defects are to be found in dental X-ray pictures that are - similar to small microcalcifications in mammograms - sometimes difficult to detect by the dentist. This especially applies to periodontal bone loss or periodontitis, which is preceded by an accumulation of plaque and inflammation of the gingiva. Early diagnosis, when the defects are only small, is important. Van der Stelt and Geraets developed a method for the early detection of such periodontal defects. Clinicians were asked to rank the severity of the lesions in radiographs and the computer detection of microlesions was compared with the diagnosis by human observers. Pictures were scanned and presented as 512x512 8-bits images. The image processing consists of filtering, segmentation, and bone density estimation. The outcome of the study was that the automatic detection compared very well with the diagnosis by the group results of seven experienced observers, although the authors did not clarify how the group result was obtained.

A surprising subject for image analysis is offered in the paper by Fairhurst *et al.*, who wanted to detect as objectively as possible neurological dysfunction in patients suffering from Alzheimer's disease or other dementing illnesses, or patients who suffered from a stroke (CVA). Progressive development of such diseases is also difficult to monitor without objective appraisal. The idea of the authors was to let these patients copy a geometrical shape by free-hand writing and then to compare the result with the original target shape. It is to be expected that patients with neurological disorders will have difficulties in completing such tasks and that the result will be a function of the severity of the disorder. If the patient uses a pen connected to a computer-controlled writing tablet, several parameters can then be followed, such as execution time, pen velocity and acceleration, hesitation in task execution, tremor, errors, degree of matching, and distortions in lines and corners. Of course, to be able to compare different groups of patients, a standard data set should be defined. The resulting illustrations as shown in the article are very convincing, although in forthcoming studies the authors will certainly extend their study by comparing patient groups in a more formal way. The authors foresee that their technique will also be used for screening purposes.

The last two articles in this section are on signal processing applications, the one by Khadra *et al.* discusses phonocardiograms, and the paper by Thorpe *et al.* cough sounds Phonocardiogram analysis by computers is one of the early applications of biosignal processing. A phono cardiogram is a typical non-stationary signal so that an overall spectral analysis would hide certain important frequency components. For that reason, data windowing is first applied before further analysis such as FFT. The authors have introduced a new method called the Wavelet Transform, which is essentially a decomposition of the signal into orthonormal components, the wavelets. FFT can be used to compute the contribution of each wavelet in the original signal. It is clear from the publication and its conclusions that the method awaits further research to assess its diagnostic value.

The analysis of cough sounds can be of value for patients with chronic respiratory diseases such as asthma. Thorpe *et al.* have developed a system for the analysis of such sounds, described in their paper, but also here the method needs further diagnostic assessment. The signals that are acquired at the mouth instead of at the chest, consist of an air flow rate and sound pressures, and the method can be combined with a usual spirometric test. Signals are displayed together with their spectrograms during patient examination. Running frequency spectra are displayed in a quasi 3-D mode for user inspection, but as yet no further signal analysis is done, although the authors foresee the use of signal segmentation and further quantification of parameters.

Decision-Support Systems

Just about 10 years ago, most papers on decision-support systems only described ideas and methods, without presenting results on extensive evaluation studies. From the articles in this and the following sections of this Yearbook it is, therefore, impressive to see that so many decision-support systems have now been thoroughly assessed and that several systems have been or are on the verge of being introduced in the clinical and health care setting. The central question in all evaluation studies is and will remain: *what is the Objective Reference or Gold Standard by which to evaluate the systems?* Such objective reference material should preferably be collected in formal studies or by panels of independent experts. Ideally, the Reference is material which is not biased by the study itself. However, especially in medicine, this cannot always be avoided: the mere fact that participants of a study are aware that they are taking part in a formal evaluation study hampers the collection of objective material. A certain measure of objectivity can be achieved by using opinions from a panel or data from independent sources, such as clinical evidence from post-mortem evaluation. Several papers in this and the following section are outstanding examples of how to conduct assessment studies.

The paper by Willems *et al.* is the report of an extensive international evaluation study (known as the CSE project, *Common Standards for Quantitative Electrocardiography*) in which virtually all computerized ECG interpretation systems in the world were diagnostically tested. The diagnostic classifications of about 10 ECG systems were compared with ECG-independent clinical evidence and with the group outcome of eight cardiologists, with the help of 1,220 well-documented ECGs, falling in seven main diagnostic categories. The best systems had either an overall agreement with the clinical evidence of about 70% or of about 80% if compared with the group of cardiologists. Several aspects of CSE may offer guidelines for similar assessment studies of other decision-support systems, such as how to collect a properly documented database; that an independent coordinating institution should be involved; how to arrive at a weighted group consensus of experts; and how to present evaluation results in a balanced and concise manner. The effect of how the knowledge was acquired in the ECG interpretation systems was clearly seen in the outcome of the CSE project: most programs were trained on cardiological expert knowledge, and some on ECG-independent criteria. The former group showed a better agreement with the group of cardiologists, whereas the latter had a higher total agreement if compared with the clinical evidence.

De Dombal is one of the pioneers in developing decision-support methods for internal medicine, and especially acute abdominal diseases. Over the years his team has published a host of innovative papers [11, 12] so that the editors were happy to include one of his recent articles in the Yearbook. In the past, studies by De Dombal *et al.* have indicated the importance of continuous training and support of physicians in complex areas such as the diagnosis of acute abdominal pain. In the present study, an impressive total of over 12,500 patients was involved and true patient outcomes, such as perforation rate or negative laparotomy, were taken as references for the evaluation of the support methods. Three different support methods, or combinations thereof, for inexperienced staff were mutually compared: structured data collection forms, real-time decision support, and computer-assisted teaching. It appeared that a combination of forms and teaching gave equally good results as forms plus direct decision support. The largest decrease in unnecessary interventions was seen in perforation rates among patients with appendicitis (falling from a base-line value of 27% to about 12% in this study). De Dombal could also show - important in this time of steadily increasing costs of health care - a decrease in the usage of surgical beds. An interesting observation is made at the end of the article: The use of the computer for teaching may be a good alternative *"for those who are apprehensive about the real time use of diagnostic computers in a clinical setting"*, which might become a superfluous statement when the time has come that computerized decision support will be an accepted adjunct to clinical care.

Ideally, an evaluation of decision-support systems is conducted by institutions that are independent of the original investigators who developed the systems (as was done in the CSE project). Such projects are expensive, all the more so when the interpretation systems and validation procedures are complex. In such circumstances it is wise to test components of the system separately or to carry out the evaluation in different stages. Papers by Bernelot Moens are examples of assessment studies in which the original investigators took no part either in the collection or in the evaluation of the specific decision-support system. However, in the present article Bernelot Moens and van der Korst compared their own Bayesian model for the diagnosis of rheumatic disorders with the diagnosis of expert rheumatologists, but they used a fairly large independent test database for evaluation. The results were expressed in ROC curves, in which it is striking to see that both the model and the experts stay on the safe side and report only a few percentages of false positives, indicating a high specificity.

The work of Spiegelhalter contains several studies on the value of structured questionnaires for diagnostic support [13]. In the article by Franklin *et al.* in which Spiegelhalter also participated, a questionnaire was developed for life-threatening situations in which speed of action and correct referral may save lives of neonates suffering from congenital heart diseases. Three situations were compared: conventional referral of children to specialist cardiac centers, referrals supported by a structured questionnaire, and a method where a diagnostic algorithm operated on the data from the questionnaire. The mere usage of the structured form gave an improvement in correct referrals, increasing from 34% to 48%; the implementation of the diagnostic algorithm would have improved this percentage to as high as 78%. The appropriateness of management during transit from the referring center to the top cardiac centers was also assessed. Here, the computer model would have given an improvement from the usual 80% to 91%. Correctly so, the authors see no reason for not introducing this method in real practice since it reduces both the morbidity and the mortality in neonates who are afflicted with congenital heart diseases.

In the article by Nykänen *et al.* reflections are made on how to evaluate decision-support systems in medicine. Criteria are given for evaluation studies, based on an analysis of the literature in this domain. Evaluation should cover, in the words of the authors, all stages of the development process of the support systems. An important remark is that evaluation is a process to be continued after the introduction of a decision-support system into practice. The process of knowledge acquisition should also be part of the assessment. The authors discuss, too, what parties should be involved in this evaluation: users, domain experts, and above all third, independent parties. It is reassuring to observe that several studies reported in the articles in this Yearbook have already adopted sound and objective rules for the evaluation of decision-support systems.

Middleton *et al.* reported on a reformulated version of the QMR knowledge base, restructured with the help of probabilistic inferencing. This version of the system was evaluated with 23 single-disease CME cases published in the literature. The probabilistic model they used is based on Bayes' rule for mutually exclusive diseases. The results of their model were comparable with those of QMR itself, although the probabilistic version was not very sensitive to uniform prior probabilities. The authors think that their approach is well suited for an extension with utility parameters. An interesting observation that the authors make and which might be applicable to other knowledge bases as well, is that they are convinced that their reformulation of an existing KB saved them an enormous amount of time over developing their own knowledge base.

A further paper on Bayesian reasoning is that of Gammerman and Thatcher. The authors did not assume independence of symptoms, though it lies at the root of Bayes' rule. The most discriminating symptoms are searched by a sequence of chi-squared tests. The results of the approach for a large collection of patients with acute abdominal pain (regretfully not the same as in the study by De Dombal *et al.*) were compared with a conventional Bayes' model and diagnoses made by physicians. Although their model did not work out as well as "simple" Bayes (overall accuracy 65% and 74%, respectively) and the preliminary diagnoses of physicians were 76% correct, the authors still see many advantages for their method. The reason that their method performed less well is said to be caused by the insufficient size of the database that was used, so that it lacked sufficient detail. Currently, the authors are also investigating logical regression.

The paper by Wu is based on his winning student paper at one of the last SCAMC meetings in Washington. It addresses a problem which is difficult to formalize by computer models: how to interpret multiple disorders. This problem is even more complex if the diagnosis of different diseases is based on overlapping collections of symptoms. The problem can be tackled by its decomposition into several subproblems which each form a coherent problem area. Only very few diagnostic problems, however, can easily be decomposed; in that case at least a partial decomposition can be tried. This is accomplished by assigning each existing symptom into clusters of existing decompositions. To that end, the paper states that four different operations can be followed: covering, restriction, addition, and extraction. To test the model, a specific disease (*prerenal azotemia*) was chosen with the help of the QMR knowledge base [14], requiring that the symptoms of this disease should also be relevant for as many as possible other diseases. In this case, each symptom had a relation from 2 to even 76 other diseases, in total resulting in 147 diseases. Test cases were randomly generated from the 14 symptoms, containing 7 to 12 symptoms each. Symptom decomposition appeared to operate quickly and efficiently when compared to candidate generation. Above all, diagnostic performance increased substantially so that the author concludes that problem decomposition is a step forward toward a new generation of decision-support systems.

Knowledge-based Systems

This section is not significantly different from the preceding one, although here the structuring of medical knowledge plays a more prominent role. The section starts with two papers, one by Wyatt and one by Van der Lei *et al.* that both reached the inner circles of clinicians by virtue of their publication in The Lancet.

Jeremy Wyatt discussed for a forum of practicing physicians the value of knowledge-based systems for medicine. This concise article offers a good overview of the present state of such systems, cites the most prominent principal investigators, and describes the ideal characteristics of a knowledge-based system. Wyatt tells his audience that no system as yet meets the ideal criteria of comprehensiveness, accuracy and verifiability, accessibility, and adaptability, but he expects that such systems are, given the advances in present information technology, around the corner. He states, very properly, that professional medical societies are increasingly taking responsibility for their members' competence. The establishment and maintenance of a medical knowledge base, however, would be a major undertaking, as is the evaluation of existing systems. If this would be adopted by these societies, he expects a major leap forward in the quality of medical care.

The article by Van der Lei *et al.* that addresses the quality of care and the evaluation of a knowledge-based system in a limited medical domain — the treatment of patients with hypertension — appeared in the issue following that in which Wyatt's paper was published. A novel approach in the use of a knowledge-based system was followed in this study. The medical data contained in a routine computer-based patient record system (Elias [15]), served as input to an expert system (Hypercritic [16]) that acted as a critiquing system. In this way, the physicians were not urged to enter patient data into a decision support system in a separate session. The question posed in this study was whether this combination of medical data and medical knowledge would generate similar critiques as domain experts would have done. From the study it appeared that - within its domain - Hypercritic could compete with eight expert reviewers and, which is equally important, that the computer-based patient record system contained a sufficient amount of data to be used as input for the expert system. This is an encouraging result, since it should be realized that if ever knowledge-based systems are accepted for routine clinical use, they should be fully integrated in medical practice.

One of the early systems introduced for routine decision support is DXplain, developed by the research group of Octo Barnett [17]. The paper by Feldman and Barnett describes the strategy for evaluating the accuracy of DXplain's differential diagnoses. Once it has obtained disease symptoms, DXplain generates a list of plausible diseases that best explain the case terms, instead of picking just one single diagnosis. In this evaluation study 46 test cases were obtained from a variety of sources. It appeared that the system agreed well with the consensus of experts, scoring for 27 out of the 46 cases within one standard deviation from the score of the experts.

The Section of Medical Informatics at Stanford University is one of the earliest departments that invested a large amount of time and effort in the development of knowledge-based models and systems. There is a steady stream of papers in this field originating from this group, headed by Ted Shortliffe [18, 19]. A few papers in the preceding and in this section are directly or indirectly connected to the research at Stanford, such as the paper by Herskovits and Cooper, that by Lehmann and Shortliffe, and the one by Kahn *et al.* The high-quality work in this department shows that it takes a large amount of time and effort before a solid foundation is laid for a scientific approach to designing knowledge-based systems.

In a way parallel to some other papers in the preceding section, the article by Herskovits and Cooper shows that during the last years successful attempts have been made to close the gap between heuristic and probabilistic methods. This paper attempted to represent such probabilistic relationships among the variables in expert systems by adopting Bayesian belief networks. In essence, a belief network is a graph of nodes and arcs, in which the nodes represent the variables and the arcs probabilistic associations. Since computation is complex, the authors have designed efficient algorithms based on precomputation or caching. Speedy algorithms are of importance in real-time monitoring situations or in circumstances where interaction between system and user is required.

Another example of a Bayesian approach is the paper by Lehmann and Shortliffe. The authors distinguish between probabilistic, methodological, and domain knowledge. Citing Clancey, the authors state that one of the reasons to construct statistical expert systems is that the medical community has completely misconstrued the nature of statistical models. The system they designed (called Thomas), was intended to assist physicians in interpreting the results of randomized clinical trials (RCTs). Physicians are often confronted with the outcome of such trials and may wonder what the usefulness is for direct patient care. Thomas is able to give advice and assistance with respect to a clinical decision to be made on the basis of RCTs, patient data, and prior probabilities on diseases. It will be interesting to see the evaluation of this system on real cases in future publications.

During this synopsis of the articles in the Yearbook thus far, several times the work on Internist or QMR [14] of the Section of Medical Informatics in Pittsburgh was mentioned, since QMR was either extended (Middleton *et al.*), or used as an input for further research (Wu). The present paper by Giuse *et al.* addresses the problem of how to obtain quantitative data for knowledge acquisition from the literature, in this way easing the work of knowledge acquisition by experts. This approach is also of great importance for knowledge maintenance, a research area that is not yet covered by most research groups. The method that is followed is that from diverse literature sources numerical figures are drawn and converted to a standard format. This pertains to ratios, percentages, and a pertinent reference for each entry. Non-numerical values such as qualitative texts were considered too imprecise to serve as inputs. These heuristics were then checked by a simple algorithm. The study is considered by the authors as very encouraging in that automated creation and maintenance of numerical data in medical knowledge bases is in principle possible. This is at the same time equally important for knowledge validation.

There are several extensive knowledge bases operational for medical purposes. An example is the QMR KB which contains about 600 diseases and 4,300 symptoms. Still no one can give solid proof that the knowledge contained in such KBs is complete. Heckerling *et al.* studied the effect of incomplete knowledge on diagnostic outcomes. To that end, they used the Iliad system [20] (containing 700 disease profiles) to generate diagnostic suggestions, and compared the results with the diagnoses of two internists who were requested to diagnose the cases before and after confrontation with Iliad's verdicts. Fifty cases were used to test the system of which 28 were contained in Iliad's KB. The physicians were able to improve their diagnoses significantly for the 28 cases for which Iliad contained knowledge, but not so for the other 22 cases. The authors conclude in their last sentence that "*when human consultants reach the limits of their knowledge they often express uncertainty, or gracefully defer to other experts; computer consultants ideally should do the same.*"

Kahn, Fagan and Tu studied the problem of temporal reasoning in medical expert systems that use data that change over time. In a way, also Van der Lei *et al.* were confronted with similar problems in critiquing hypertension treatment based on temporal data in computer-based patient records. Kahn *et al.* extended the TOD (Time-Oriented Databank) model by two versions, to use the data as inputs to an expert system such as Oncocin, an expert system for chemotherapeutic advice. Both extensions have proven to be able to encode complex temporal relationships.

The article of Patrick *et al.* is another example of a successful combination of image processing and knowledge processing. In a way it is comparable to the problem addressed by Tahoces *et al.*, discussed in section 3. Clinical findings and parameters from image processing are entered into a network of expert learning systems. This network was first trained and subsequently tested by 20 randomly selected cases from a set of 128 difficult cases with a ratio of 50/50 benign versus malignant. The main question to the system was to discriminate between malignant and benign calcifications in mammograms. In comparison with the clinical evidence, the system proved to be correct in 72% of the cases. The authors conclude correctly that their method requires a more extensive evaluation and a presentation of the results by ROCs (as was done in the studies by Deller *et al.*, Okajima *et al.*, Willems *et al.*, and Tahoces *et al.*). Given the high incidence of breast cancer, even a slight improvement of diagnostic accuracy would be of great help which should also be balanced against costs.

The contribution by Jones is an enthusiastic plea for the introduction of expert systems in nursing. Jones calls it an inevitable development. Essentially the paper is intended to be an introduction for non-experts and to explain the present possibilities and tools for the development of expert systems. However, knowledge in the nursing domain is often less quantifiable and less structured along formal lines. The author states that automated rule induction is a methodology which is best suited for knowledge-based systems in nursing.

Neural Networks

It has taken some time before neural networks found their applications for biomedical problems. The editors are happy to include here five papers that use a neural network approach for subjects that are directly or indirectly related to medicine. In the basic sciences and for rather restricted problem areas, neural networks already found a proper place some time ago. By incorporating these papers in the Yearbook, the editors hope that further research is stimulated for medical decision support by neural networks as well.

The paper by Baxt offers a comprehensive introduction into the methodology followed in neural networks. Although Baxt depicts a four-layer neural network, in most applications only three layers are used. As in all pattern recognition, the network is trained by a learning set, and thereupon tested by an independent test set. The purpose of the study was to assess the usefulness of a neural network for the identification of myocardial infarction in patients presenting with chest pain. A collection of 352 patients was used to train the network; testing was done with 331 consecutive patients. Input data consisted of elements from the patient history, from physical examination, and ST-T findings from the ECG. A group of cardiologists also diagnosed the same patients and had a sensitivity and specificity of about 78% and 85%, respectively. These figures were for the neural network about 97% and 96%, respectively. These results look very promising for the neural network, although the authors properly conclude that the method should be tested on a larger group of patients. Besides, it should be questioned whether these good results are due to the neural network or mainly to the choice of input parameters. If the latter is the case - which would be no surprise -, other decision models should be able to derive at least at equally good results.

The article by Furlong *et al.* also deals with myocardial infarction and is an attempt to predict infarction with the help of enzyme data. The neural network was properly trained and subsequently tested with the help of enzyme data sets of 53 patients: 24 with and 29 without infarction. The sensitivity of the system was 100% and its specificity 93%. Because of the few cases, this system should also be tested on larger material, but its results are certainly promising.

The next two papers deal with applications in biochemistry: identification of ^{1}H-NMR spectra of oligosaccharides and prediction of E.coli RNA polymerase promotor sequences. As stated in the article by Meyer *et al.*, neural networks can successfully be used in circumstances where the researcher deals with strongly overlapping feature distributions, a crucial problem in pattern recognition. In this study, it is the intention to use ^{1}H-NMR spectra to discriminate between complex carbohydrate structures. Features from spectra of oligosaccharides were used as inputs to the neural network with an input layer of over 2,000 nodes or neurons, a hidden layer of 10 and an output layer of 13 nodes. Spectral resolution was in this way 0.5 Hz per input neuron. The authors see as a clear advantage of neural networks over conventional pattern recognition, that no experts were required to define features in the spectra and define rules for discrimination.

Demeler and Zhou used a neural network for the determination of nucleic acid sequences by optimization of the parameters of a neural network of three layers. The network was trained with 80 known promotor sequences with different numbers of random sequences, and for testing a set of 30 promotor sequences and 1500 random sequences were used. The authors obtained very high accuracies: 100% for the promotor set and 98.4% for the random set. They found that the number of nodes in the hidden layer did not have a significant influence on the outcome of the classification. They have the impression that a higher number of random sequences in the training set would further increase the sensitivity of the network.

In the article by Gindi *et al.*, fluorescence spectra were used to classify tissue either as atherosclerotic or as normal to guide selective laser ablation of plaques. The study compared nearest-neighbors methods from the field of pattern recognition with the performance of a neural network. Two different training sets were used: for the aorta 50 normal and 50 abnormal specimens, and also 70 each for the coronary arteries. The neural network with 22 input, 8 hidden, and 1 output node had an overall performance for the aortic data of 96%; for the 1-nearest neighbor method this was 98%. These figures were for the coronary data 87% and 87%, respectively. The authors also compared other classification techniques with these two, such as multivariate linear regression, principal component analysis, and Bayes' decision rule. Most classification results were the same order of magnitude as the other ones, although the nearest-neighbors method and the neural network back-propagation technique performed best.

Biomedical Research

The last section of the Yearbook contains articles that cannot be sharply categorized under any of the previous sections. All papers in this Section deal in one way or the other with the general topic of Biomedical Research.

The article by Safran has a general message for a host of research projects: whether it is possible to use the large quantities of medical data stored in hospital and departmental computers as a mine for data analysis, "*a largely untapped resource for research*". In a way, this issue was also addressed in the articles of Gelsema *et al.* and Van der Lei *et al.* The system developed by Safran is called ClinQuery and supports the interactive retrieval of routinely collected data. He cites an impressive use of ClinQuery: over 900 clinicians have used the system more than 10,000 times since it was made available. Of these consultations, the system was used in 38% of the cases for clinical research, in 16% for patient care, and in the remainder for teaching, data exploration, and administration. Application areas varied widely: from cardiology to AIDS research, clinical chemistry, or radiology. If Safran's paper indicates a trend, all developers and users of hospital or departmental information systems should take care that the databases concerned are as reliable and complete as possible.

The contributions in this Yearbook by Jick *et al.*, Johnson *et al.*, Van der Lei *et al.*, Ornstein *et al.*, and Tsevat *et al.* already indicated the importance of routinely collected data for preventive usage. It is of great importance that all epidemiological or preventive studies investigate the recall bias when patients with certain risks are asked to report on certain accidents or disorders. Therefore, the editors were happy to include the paper by Lachenbruch *et al.* in the Yearbook. Lachenbruch investigated the effect of keeping records by patients themselves on such events; the group studied were elderly persons who suffered from more or less severe falls. The method that Lachenbruch used is called the Proportional Hazards Model. The authors think that their adjustment of this model can be generalized and can be useful for other areas in which time to an endpoint is of interest.

In the preceding section, neural networks were used to predict the RNA sequence from ^{1}H-NMR spectra of oligosaccharides. The present article by Gultyaev also deals with RNA sequencing, but now on the basis of using Monte Carlo simulation to study stepwise RNA folding by generating random structures. The article is a fascinating contribution of computers to molecular biological problems. It was made operational on a 8 MHz 8088 processor, using Turbo C, and was efficient enough to arrive at satisfying results in a short period of time. The medical informatics community should be aware that not only clinical applications deserve attention, but also applications of models and large databases in the basic sciences request the attention of our field; this was an additional reason to include this paper in the Yearbook.

A group that has built up a tradition in artificial intelligence is the one headed by Perry Miller. Miller is also aware of the fact that molecular biology can be a very interesting area for researchers in medical informatics. In that respect he shares this view of Don Lindberg of the National Library of Medicine, who founded at NLM a substantially large group that works closely with NIH's human genome project. The article by Miller *et al.* is one of the first publications of activities in molecular biology by medical informatics researchers. The main issue of this study is software development and performance. In their project, the authors applied parallel computing to the system Linkmap, a program which is in use for genetic linkage analysis. The authors think that their exercise is a representative illustration of how to convert existing conventional, sequential and iterative programs to a parallel program, which is useful now that parallel processors are steadily becoming cheaper.

The last paper in this section concerns the evolution of the National Library of Medicine in Bethesda, USA, from a rather small unit to the largest institution for the documentation of medical literature in the world. More and more, the NLM is becoming "intelligent", thanks to the challenging research projects taking place there. DeBakey gives in his paper an enthusiastic overview of the developments at the NLM in this electronic age. For the readers of the Yearbook it is important to see that the proliferation of the knowledge contained in the NLM is indeed worldwide. For that reason, the editors decided to put this article at the end of this series of high-quality papers in medical informatics.

REFERENCES

[1] Collen MF. General requirements for a medical information system. Comput Biomed Res 1970;3:393-406.

[2] Barnett GO. The application of computer-based medical-record systems in ambulatory practice. N Engl J Med 1984; 310: 1643-50.

[3] McDonald CJ. *Action-oriented Decisions in Ambulatory Medicine.* Chicago: Year Book Medical Publishers, 1981.

[4] Warner HR. A sequential Bayesian approach to history taking and diagnosis. Comput Biomed Res 1972; 5: 218-7.

[5] Weed L. *Medical Records, Medical Education and Patient Care.* Cleveland: Case Western Univ Press, 1969.

[6] Dick RS and Steen EB, eds. *The Computer-based Patient Record. An Essential Technology for Health Care.* Washington DC: National Academy Press, 1991.

[7] Roger FH. The Minimum Basic Data Set for Hospital Statistics in the EEC: Review of Availability and Comparability. In: *Technical Report EURO 7162.* Luxembourg: Office for official Publications of the EEC, 1981: 1-151.

[8] Smith NT, Quinn ML, Flick J, *et al.* Automatic control in anesthesia: a comparison in performance between the anesthetist and the machine. Anaesth Analg 1984; 63: 715-22.

[9] Sheppard LC. The computer in the care of critically ill patients. Proc IEEE 1979; 67: 1300-9.

[10] Lodwick GS, Haun CL, Smith WE, Keller RF and Robertson E. Computer diagnosis of primary bone tumors. Radiology 1963;80:273-5.

[11] De Dombal FT. Computer aided diagnosis of abdominal pain. Brit Med J 1972; 2: 9-13.

[12] De Dombal FT, *et al.* A computer assisted system for learning clinical diagnosis. Lancet 1969: 145-50.

[13] Spiegelhalter DJ, Knill-Jones RP. Statistical and knowledge based approaches to clinical decision support systems, with an application in gastro-enterology. J Royal Statist Soc 1984: 1.

[14] Miller RA, Pople HE, Myers JD. INTERNIST-1: An experimental computer-based diagnostic consultant for general internal medicine. N Engl J Med 1982; 307: 468-76.

[15] Westerhof H, Boon WM, Cromme PV, van Bemmel JH. ELIAS: Support of the Dutch general practitioner. In: Reichertz PL, Engelbrecht R, Piccolo U, eds. *Present Status of Computer Support in Ambulatory Care.* Berlin: Springer Verlag, 1986: 1-10.

[16] Van der Lei J, Musen MA. A model for critiquing based on automated medical records. Comput Biomed Res 1991; 24:344-78.

[17] Barnett GO, Cimino JJ, Hupp JA, Hoffer EP. DXplain. An evolving diagnostic decision-support system. JAMA 1987;258:67-74.

[18] Shortliffe EH, Buchanan BG, Feigenbaum EA. Knowledge engineering for medical decision making: A review of computer-based clinical decision aids. Proc IEEE 1979; 67: 1207-24.

[19] Clancey WJ, Shortliffe EH, eds. *Readings in Medical Artificial Intelligence.* Reading: Addison Wesley, 1984.

[20] Warner Jr HR. Iliad: moving medical decision making into new frontiers. Methods Inform Med 1989;28:370-2.

Addresses of the authors:

Jan H. van Bemmel Ph.D.,
Department of Medical Informatics,
Faculty of Medicine and Health Sciences,
Erasmus University,
P.O. Box 1738,
3000 DR Rotterdam,
The Netherlands

Alexa T. McCray Ph.D.,
National Library of Medicine,
8600 Rockville Pike,
Bethesda Md 20894,
USA