

## DATABASES FOR CARDIOLOGY

# Developments in Cardiovascular Medicine

VOLUME 115

*For a list of titles see end of volume.*

# DATABASES FOR CARDIOLOGY

*edited by*

**GEERT T. MEESTER**

*Interuniversity Institute of Cardiology,  
Utrecht, The Netherlands*

*and*

**FRANCESCO PINCIROLI**

*Department of Bioengineering, Politecnico di Milano,  
Milan, Italy*



**SPRINGER-SCIENCE+BUSINESS MEDIA, B.V.**

## Library of Congress Cataloging-in-Publication Data

Databases for cardiology / edited by Geert T. Meester and Francesco Pinciroli.

p. cm. -- (Developments in cardiovascular medicine ; v. 115)

Includes bibliographical references.

Includes index.

ISBN 978-94-010-5653-3 ISBN 978-94-011-3720-1 (eBook)

DOI 10.1007/978-94-011-3720-1

1. Cardiology--Data bases. I. Meester, G. T. (Geert T.)

II. Pinciroli, Francesco. III. Series.

[DNLM: 1. Cardiology. 2. Information Systems. W1 DE997VME v. 115 / WG 26.5 D232]

RC682.D35 1991

616.1'2'0285574--dc20

DNLM/DLC

for Library of Congress

90-5009

ISBN 978-94-010-5653-3

---

*Printed on acid-free paper*

All rights reserved

© 1991 by Springer Science+Business Media Dordrecht

Originally published by Kluwer Academic Publishers in 1991

Softcover reprint of the hardcover 1st edition 1991

No part of the material protected by this copyright notice may be reproduced or utilized in any form or by any means, electronic or mechanical, including photocopying, recording, or by any information storage and retrieval system, without written permission from the copyright owners.

# Table of contents

Acknowledgements	ix
Foreword	xi
<i>Geert T. Meester and Francesco Pincioli</i>	
Contributing Authors	xv
<b>Part One: Basic approaches</b>	<b>1</b>
1. Introduction to basic concepts on methods and techniques for Databases	3
<i>Richard F. Walters</i>	
2. Observational databases: a clinical perspective	41
<i>Suzanne B. Knoebel</i>	
<b>Part Two: Tools and services</b>	<b>51</b>
1. Advanced instruments and methods for the development of databases applied to cardiology	53
<i>Francesco Pincioli and Carlo Combi</i>	
2. A query language for medical statistical analysis	77
<i>R. Brennecke, H.J. Rupprecht, R. Erbel and J. Meyer</i>	
3. Enhancing relational database management systems by applying artificial intelligence techniques	87
<i>Willem R.M. Dassen, Rob G.A. Mulleneers and William P.S. van Braam</i>	
4. Electronic information in cardiology: review of external databases and services	97
<i>Luciana Libutti</i>	
5. Computer-aided clinical problem solving as an educational paradigm for teaching preclinical cardiac pathophysiology	113
<i>Robert A. Greenes, Bryan P. Bergeron, Mark S. Dichter and John T. Fallon</i>	
6. Continuing education of physicians and nurses in the DBMS area	141
<i>G. Invernizzi, C. Combi and G. Pozzi</i>	
<b>Part Three: Subject-oriented databases</b>	<b>153</b>
1. The ARTEMIS data and knowledge base for hypertension	155
<i>Partrice Degoulet, Marion Lavril, Pierre-François Plouin, Gilles Chatellier and Joël Ménard</i>	

2. Databases for prevention, pacemaker and postoperative treatment: the charité experience in cardiology	167
<i>K.H. Günther, R. Bohm, V. Dilba, R.R. Moebes, St. Petzschmann, P. Piorkowski and J. Witte</i>	
3. Development of and experience with the coronary angiography and PTCA information systems at Leuven University	179
<i>Jos L. Willems, Stef Renkens, Marcel De Wel and Jan Piessens</i>	
4. A database for the follow-up of heart transplant patients	201
<i>Ph. Doublet, P. Rubel, G. Dureau and P. Vuillard</i>	
5. The K.U. Leuven Coronary Surgery Data Base: a clinical research data base	223
<i>P. Sergeant and E. Lesaffre</i>	
6. Clinical results with computer support of the decisions (in the cardiosurgical intensive care unit)	239
<i>Vladimir A. Lishchouk</i>	
7. A pediatric cardiology diagnostic coding system and database	259
<i>R.W. Brower, E. Harinck and A.C. Gittenberger-de Groot</i>	
8. Databases and decision system for diagnosis of congenital heart disease	273
<i>Ivan Malčić and Darko Ivančević</i>	
9. Practical PC-based data management in paediatric cardiology	289
<i>K.J. Sprenger and M.M.A. de Moor</i>	
<b>Part Four: Departmental applications</b>	<b>305</b>
1. Patient documentation for the ultrasound laboratory	307
<i>Rüdiger Brennecke, Frank Schön and Raimund Erbel</i>	
2. A research-oriented database management system for Holter data	315
<i>Pierfranco Ravizza</i>	
3. A database management system of coronary care unit data	335
<i>Marco Tubaro, Carlo Iacoangeli, Enrico Natale, Roberto Ricci and Filippo Milazzotto</i>	
4. Information management for decision making by critical care physicians and nurses	343
<i>Thomas W. Rosenal, Margaret White and Kathryn J. Hannah</i>	
5. Practical data management in the Cardiology Department of a City Hospital	357
<i>A. Pizzuti</i>	
6. A PC-based implementation of a multi-service software for cardiology	367
<i>F. Pinciroli, C. Combi, G. Pozzi and R. Rossi</i>	
7. An out-patient clinical data base management system	381
<i>G. Pinelli, G. Carini, G.L. Manini, S. Urbinati, M. Ruffini and L. Guazzotti</i>	

8. Databases for network systems	393
<i>L. Goncalves and R. Domingos</i>	
9. CADANS: the nervous system for cardiology	401
<i>Geert T. Meester, O.E.H. Sartorius, D. Fanggidaej,</i>	
<i>M.G.J.M. Gerritsen, J.A.T.G. van Kester and</i>	
<i>F.M. van Rappard-van der Veen</i>	
Index	415

## Acknowledgements

Bayer AG and Bayer Italia SpA supported in full the second Como Workshop of 'Databases for Cardiology', held in January 1990 at the Centro di Cultura Scientifica 'Alessandro Volta' at Villa Olmo in Como, Italy. The Workshop was promoted by the editors. It was specifically intended for the compilation of this book. Mrs. Chiara De Santis served as general secretary to both the workshop and the book.

**FRANCESCO PINCIROLI and GEERT T. MEESTER**  
*Editors and Promoters of the Second Como Workshop*



## Foreword

A database is in principle just a large collection of related or separate data, systematically stored in a computer. It should be possible for the data to be easily entered into the database-structure and afterwards also easily read, corrected and processed.

The later analysis of data from such a database is greatly enhanced by the availability of special query languages and statistical analysis programs, not only for serial items but also for large combinations of data.

Query languages, such as SQL (Structured Query Language) developed especially for these purposes, make databases easily accessible, also to researchers who may not be very well versed in computer programming.

The cardiological/medical clinician and researcher of today is of necessity confronted more and more with computer-based data storage.

Interest is of course focused primarily on the clinical use of such databases more than on the technical design itself, except for some very specific, personalized applications.

For the latter approach, there are at present many software packages commercially available, especially designed for use in the personal computer environment.

This book is comprised out of a number of contributions by various authors with differing backgrounds and from many different countries.

The editors, being a cardiologist and an information scientist, have strived to achieve an equilibrium between these two fields. The chapters in this book form a cross-section of the many approaches to database design and implementation in the area of cardiology.

The material included herein will assist the reader to make an informed choice when facing the necessity to design or utilise a patient-oriented database. It should also provide the opportunity to familiarise him/herself with the intricacies of computerised databases, as seen from the viewpoint of the practicing cardiologist or researcher in the cardiovascular domain.

Successful data-analysis is an essential condition to database design. Therefore, several chapters are dedicated to the interrogation of existing databases. A survey of query languages would deviate too far into the technical

aspects, other texts supply abundant information regarding the choice and use of these tools.

During the assembly of the text the editors became aware of the lack of an annotated bibliography in this field. Furthermore, a software catalogue, however useful, could not be constructed within the scope of this book, because it would be almost immediately outdated. For this, the reader is directed to the Cardiovascular Software Directory compiled by the American College of Cardiology [1], the Annual Medical Software Buyers' Guide in the Journal of M.D. Computing [2] and with a broader spectrum with regard to area of application, the Directory of Portable Databases [3].

In addition, the editors stressed the need to include databases used to develop artificial intelligence and confirm testing of expert systems and diagnostic programs and to organize databases for testing purposes, taking into account the experience of the ECG.

One of the authors in this volume, Dr. R. Brennecke (Mainz, Germany) is planning to maintain a catalogue of cardiological databases. Copies are available on request. Readers are recommended to supply additional information on databases not yet covered in this catalogue. As far as the situation in Italy is concerned, workshops for Cardiological Databases were held in 1987 and 1988 and the related proceedings were published [4].

The contents of this book show four main sections, covering basic approaches, tools and services, subject-oriented databases and finally department applications.

The first chapter, written by R. F. Walters, introduces concepts and methods for the design and use of databases. This chapter serves as an overall introduction to the topics in this book. The following chapter, by Suzanne B. Knoebel, approaches the cardiological database from a clinical point of view. How to extract the optimum information from an available database in a natural course of events. F. Pincioli and C. Combi cover advanced instrumentation in their chapter, including hypermedia and the MEDIX standard for the exchange of medical data.

Next, R. Brennecke *et al.* present a special query language enabling large sets of interdependent queries. Suggestions were also made to hire if necessary an information scientist educated in databases, who could help the cardiologist in keeping data consistency, back up and data dictionaries always updated. In the following chapter, Dassen *et al.* cover the application of artificial intelligence techniques to enhance a relational Database Management System. An increasing number of external databases are available to the cardiologist. Libutti surveys the field and encourages the use of these large sets of data. This topic is continued and approached from another angle in the discussion of problem-solving as an educational tool in teaching cardiac pathophysiology (R.A. Greenes *et al.*).

In addition, Invernizzi *et al.* discuss the educational aspects of the use of database systems.

Moving to the field of patient-oriented applications, the following chapters

describe several different clinical databases and their uses. Firstly, the ARTEMIS database for hypertension (P. Degoulet *et al.*) and then the East German 'Charite' experience, described by K.H. Günther *et al.* J.L. Willems *et al.*, present the coronary and PTCA information systems in use at Leuven University.

Ph. Doublet *et al.* explain the use of French videotex system as applied in the follow-up of heart transplant patients.

A large and intricate clinical research database for coronary surgery is summarised by P. Sergeant *et al.* (Leuven). It has a follow-up of 3500 patient-years, with more than 1000 variables for each record. Decision support in the intensive care unit is the topic of Lischouk *et al.* Standardising a diagnostic coding system for paediatric cardiology with a database as used in the Netherlands, forms the chapter by R.W. Brower *et al.* This is followed by a decision system to aid in the diagnosis of congenital heart disease as an attempt to rationalisation, is discussed by I. Malcic *et al.*

The data management in paediatric cardiology based on a microcomputer approach is covered by Sprenger *et al.* The echocardiography laboratory (R. Brennecke), the Holter tape laboratory (Ravizza *et al.*), the coronary care unit (Tubaro *et al.* and Th. W. Rosenal *et al.*), the out-patient clinic (G. Pinelli *et al.*) are functional units that can greatly benefit from automated database management, particularly in an integrated combination (Pizzuti). A combination of different uses in a single software package as described by F. Pinciroli *et al.*, is often practical. Finally, databases in a network environment is covered by L. Goncalves *et al.* and G.T. Meester *et al.*

The editors also wish to use this opportunity to make a few recommendations both for the design and the usage of medical databases.

With the inevitable progress in telecommunications, the need for data-transportation between databases within one hospital, one country and perhaps even between countries will certainly increase. This can only be effectuated when data-definitions are not only specified, but preferably equal from one site to another. Terms such as 'unstable angina' and 'ejection fraction' should be defined clearly. International standardisation is therefore a must. The major professional organisations in the cardiological field should promote and sponsor these activities, probably most effectively through working committees and consensus-conferences. An initiative of the American College of Cardiology [5] led in October 1989 to the formation of a Database Committee (ACCORD: ACC Outcome and Risk-Stratification Database) under the leadership of Suzanne B. Knoebel, M.D. The goal of this new committee is to develop a national computer database on selected cardiovascular procedures.

Not only in Italy and the Netherlands but also in many other countries, activities in the medical database field are taking place. Many of the chapters in this book form together a major indication of the growing attention to this area.

As far as future initiatives are concerned, the suggestion was made to organize a very specific International Summer School, focusing on the use of

large databases, to be Europe based but not restricted to Europe.

Another initiative could be the creation of a user group, with newsletters to exchange information and experiences.

Finally, it was suggested to hold the Como conference on a regular basis, to discuss the different topics in the light of the latest breakthroughs.

The editors are aware of the fact that this book is not and cannot be a manual for the construction of medical databases. Its aim is to provide the reader with an extensive overview of the medical cardiological database field, with examples of design, construction and data-analysis, while not omitting possible problems and pitfalls.

G.T. MEESTER

F. PINCIROLI

## References

1. Cardiovascular Software Directory (1988) Compiled by the American College of Cardiology Computer Applications Committee.
2. The Fifth Annual Medical Software Buyers' Guide (1988) M. D. Computing 5: No. 6.
3. Directory of Portable Databases (1990) Published by Cuadra/Elsevier, New York, Vol. 1, No. 1.
4. Milazzotto T, Martiny W, Matta F, Pinciroli F, (eds) 1987 and 1989 Banche di Dati Integrate in Cardiologia. Published by C.S.T. Torino, Italy.
5. Cardiology (1990) Official Newsletter of the American College of Cardiology 19: No. 2.

## Contributing Authors

R. Brennecke, II. Medical Clinic, Mainz University, Langenbeckstr. 1, D-6500 Mainz, FRG. ch II.2 H.J. Rupprecht, Raimund Erbel and J. Meyer. ch IV. 1. Frank Schön and Raimund Erbel

R.W. Brower, Medical Computing Consultants, P.O. Box 35010, 3005 DA Rotterdam. E. Harinck and A.C. Gittenberger-de Groot

Willem R.M. Dassen, Department of Cardiology, University of Limburg, P.O. Box 616, 6200 MD Maastricht. Rob G.A. Mulleneers and William P.S. van Braam

Patrice Degoulet, Medical Informatics Department Broussais University Hospital, 96 Rue Didot, F-75014 Paris, France. Marion Lavril, Pierre-François Plouin, Gilles Chatellier and Joël Ménard

Ph. Doublet, INSA, Department of Informatics, F-69621 Lyon Cedex 03, France. P. Rubel, G. Dureau and P. Vuillard

Lino M.M. Goncalves, University Hospital of Coimbra, Av. Bissaya Barrete, 3049 Coimbra Codex, Portugal. R. Domingos

Robert A. Greenes, Dept. of Radiology, Brigham and Women's Hospital, 75 Francis Street, Boston, MA 02115, U.S.A. Bryan P. Bergeron, Mark S. Dichter and John T. Fallon

K.H. Günther, Department of Internal Medicine, Humboldt University Hospital, Charité, Schumannstr. 20/21, Berlin, DDR 1040. R. Bohm, V. Dilba, RR Moebes, St. Petzschmann, P. Piorkowski and J. Witte

Giorgio Invernizzi, Cardiology Department, Ospedali Riuniti, L-go Barozzi 1, I-24100 Bergamo, Italy. C. Combi and G. Pozzi

Suzanne B. Knoebel, Krannert Institute of Cardiology, 1001 West 10th Street,  
Indianapolis, IN 46202, U.S.A.

Luciana Libutti, Institute of Studies on Research and Scientific  
Documentation, National Research Council, Via Cesare de Lollis 12, I-00185  
Rome, Italy

Vladimir A. Lishchouk, Bakulev Institute of Cardiovascular, Surgery,  
Academy of Medical Sciences, Moscow 117031, U.S.S.R.

Geert T. Meester, Interuniversity Institute of Cardiology ICIN, P.O. Box  
19258, 3501 DG Utrecht. O.E.H. Sartorius, D. Faggidaej, M.G.J.M.  
Gerritsen, J.A.T.G. van Kester and F.M. van Rappard-van der Veen

Francesco Pinciroli, Department of Bioengineering, Politecnico of Milan,  
Piazza Leonardo da Vinci 32, I-20133 Milan, Italy. ch II.1. Carlo Combi, ch  
IV.6. Carlo Combi, G. Pozzi and R. Rossi

G. Pinelli, Cardiology Department, Bellaria Hospital, I- Bologna, Italy. G.  
Carini, G.L. Manini, S. Urbinati, M. Ruffini and L. Guazzotti

Alfredo Pizzuti, Department of Cardiology, SS. Annunziata Hospital, I-12038  
Savigliano, Italy

Pierfranco Ravizza, Department of Cardiology, General Hospital, Via  
Ghislanzoni, I-22053 Lecco, Italy

Thomas W. Rosenal, Department of Intensive Care, The Calgary General  
Hospital, 841 Centre Avenue E., Calgary, Alberta, Canada T2E 0A1.  
Margaret White and Kathryn J. Hannah

P. Sergeant, Department of Cardiac Surgery, Gasthuisberg University  
Hospital, Catholic University of Louvain, B-300 Louvain, Belgium. E.  
Lesaffre

K.J. Sprenger, Paediatric Cardiac Unit, Dept. of Paediatrics & Child Health,  
Red Cross War Memorial Children's, Hospital, Rondebosch 7700, Cape  
Town, South Africa. MMA De Moor

Marco Tubaro, Via Val Pellice 53, I-00141 Rome, Italy. Carlo Jacoangeli,  
Enrico Natale, Roberto Ricci and Filippo Milazzotto

Ivan Malčić, Department of Paediatric Cardiology, University Hospital  
REBRO, Yu-41000 Zagreb, Yugoslavia. Darko Ivančević

**Richard F. Walters, Division of Computer Science &, Department of Family Practice, University of California, Davis, CA 95616, U.S.A.**

**Jos J. Willems, Division of Medical Informatics, University Hospital Gasthuisberg, Herestr. 49, B-3000 Louvain, Belgium. Stef Renkens, Marcel De Wel and Jan Piessens**

## **PART ONE**

### **Basic approaches**



# Introduction to basic concepts on methods and techniques for databases

RICHARD F. WALTERS

*Division of computer science and Department of Family Practice University of California, Davis, CA 95616, USA*

## 1. Introduction

Every scientific discipline deals with information, which is derived from raw data obtained from experiments, patients, surveys, or other sources. Today's world has grown so complex, the amounts of information so vast that computers are increasingly necessary to allow us to manage the sheer volume of data relating to our research, clinical practice, and daily lives.

Managing information is itself a scientific discipline. Techniques for classifying, grouping, storing, retrieving and analyzing information can be generalized beyond any specific discipline. These techniques form the body of one aspect of Computer Science, namely, Database Management Systems. Although computer scientists used to regard this field as unworthy of scientific analysis, progress in formal database theory and practical solutions to information management has convinced many that information management is a problem worthy of computer scientists.

In the meantime, specialists in almost any body of knowledge continue to accumulate knowledge about their field as they accumulate large volumes of data that must be interpreted, classified, stored, analyzed, and displayed. As knowledge in cardiology continues to advance, so too does our understanding of information management. It is no longer possible for a single individual to be a specialist in both fields (cardiology and information science). However, a cardiologist who understands the basic concepts of information management will be able to communicate with information specialists in order to derive maximum benefit from both disciplines.

The purpose of this chapter is to introduce specialists in a medical discipline (cardiology) to concepts of managing information. If it succeeds, the chapter will explain some of the most important concepts that lie behind data management, and it should also indicate how the computer can be used to help with problems of information handling.

As with any scientific discipline, a knowledge of the tools is essential. This chapter will not discuss computer hardware, operating systems, or programming languages in any detail. Readers should have some familiarity

with the rudiments of these areas, however, since technical terms dealing with these subjects will be introduced while discussing database concepts. Readers with minimal background in personal computers will probably find it necessary to learn some fundamentals in order to derive maximum benefit from this text.

The chapter begins with an overview of database systems, then discusses techniques of encoding data. It then moves into ways in which data can be classified and stored, presents several models of database systems, and concludes with guidelines on creating databases and a brief review of some available database packages.

## **2. Components of database management systems**

Consider for a moment the patient charts in a doctor's office. How did they come to be the way they are, and why? A simple answer might be that the charts were designed to meet certain clinical, financial and administrative needs. The charts consist of forms that provide spaces for specific information as well as open spaces for more flexible information. The size of the chart is determined by (or determines) the storage cabinets in which a group of charts will be stored, and the manner in which those charts are filed is determined by the expected uses of the information they contain. (For example, 'inactive' and deceased patients are usually stored separately from patients seen regularly in a practice.) The charts are used by many people – nurses, physicians, laboratory technicians, bookkeepers, secretaries – each of whom has different needs that may dictate the ways in which information is stored and retrieved. Sometimes these needs conflict, and sometimes a single chart may be needed at the same time by more than one individual. Sometimes a chart is misplaced. Some information in a chart may be out of date; some may be inaccurate. Sometimes, essential information is difficult to find because of the manner in which the chart is designed. It is difficult to design a manual system that resolves all the conflicting demands for its use. Compromises must be made, and some users will be less satisfied as a result.

This example illustrates the design problems inherent in management of a paper record system. When the same type of records are computerized, there are additional factors to deal with. To begin with, machine-readable data is generally not directly readable by humans. Instead of going into a chart storage area and searching visually for desired information, it is necessary to instruct the computer to do the searching. For this reason alone (as well as others), it is necessary to be very careful in designing a computerized data system. No computer scientist can do a decent job of designing a clinical database system alone (many have tried unsuccessfully). By the same token, few health scientists can do much better until they are armed with sufficient knowledge about the process of computerizing clinical data. This chapter will outline some rudiments of such a design process.

Computer packages designed specifically to meet the needs of data manage-

ment are referred to as *Database Management Systems*. There are many, many such packages available for personal computers, and many others running on larger systems. Knowing which one best meets individual needs requires full-time investigation. No one has that kind of time, so it is necessary to compromise, searching for generalities, relying on advice of others, and eventually making a decision on the basis of the (incomplete) information available. In this chapter, we will try to provide some clues as to the ways in which the packages are put together, so that relating them to our own needs might be a more feasible task.

In general, Database Management Systems (DBMSs) must allow certain activities to take place. These activities include:

- Design of the overall system;
- Input of data into the system in accordance with design specifications;
- Editing of data (for update and error correction);
- Searching the data for one-time queries ('browsing');
- Generation of reports on a scheduled basis;
- Analysis of the data for various purposes;
- Maintenance of the overall system (backup files, transferring certain information to other locations, preventive maintenance of hardware, upgrades of software and hardware).

Although other functions may be added to these basic components, all DBMSs have these functions available in one form or another. The design component is the most difficult of all, requiring knowledge not only of the information to be stored, but the other components in the DBMS itself. In order to design a system properly, one must know not only the nature of the data to be encoded, but also some underlying principles of database systems and specifics of the database management package that will be used. For this reason, we will defer discussion of design to the latter part of this chapter.

With this introduction, we are ready to analyze individual components in a little more detail. Let's begin by looking at how information can be stored in a computer.

### 3. Computerized data

When clinical information is entered into a computer, it must be organized in ways analogous to the organization of manual patient data systems. There must be *files* designed to hold certain types of data: patient record files, billing data, Xray files, and so on. The text-oriented files will consist of different types of forms which, when computerized, will become *records*. Paper forms usually have boxes in which specific items are to be entered; in computers these boxes are called *fields*, and the items to be entered into the fields are called *data elements*. Items appearing on paper forms usually consist of letters, numbers, and symbols. To computerize this type of data, it is necessary to convert each one of these characters into *machine readable form*. In this section, we explore

the ways in which the process of computerizing data takes place.

### 3.1. *Machine readable data*

Computers use code systems, built of binary digits (bits) grouped together into clusters (bytes) to represent numbers, letters, and other graphic characters common to data collection. The standard byte contains 8 bits, which means that a total of 256 different characters could be uniquely described. The most common code system used in today's personal computers is the American Standard Code for Information Interchange (ASCII), which uses 7 bits (reserving the 8th for future use to define 127 characters, consisting of 32 control characters, numerics, upper- and lowercase alphabets (without accent marks) and punctuation. Having only 127 characters makes it difficult to accommodate special characters such as those used in many European languages: ü in German, ê, ø in Scandinavian languages, etc. These characters often appear in databases (especially in the names of patients), but they are not easy to deal with using only ASCII characters.

The International Standards Organization (ISO) has defined several 8-bit character sets, each capable of dealing with subgroups of European and other languages. At present, personal computers do not generally recognize ISO character sets, but this situation may change in the future.

When a key on a keyboard is depressed, a code is sent telling the computer which key was typed. The action taken may be to display the character on a screen or to take some other action (such as returning the cursor to the start of a new line), depending on how the computer interprets that code. Function keys on personal computer keyboards may send several code characters with one keystroke, allowing the computer to take more varied actions. Function keys can also be redefined to emit other codes, and these features are sometimes used in database packages to provide specific features unique to that package (such as storing patient data in a particular way within a database).

As data systems become more sophisticated, it is probable that new character codes besides ASCII will become available on personal computers. It is therefore important to recognize that even the basic building blocks of creating machine readable data are subject to change.

*3.1.1. Data types.* In the preceding section, we referred to several different *types* of characters: numeric, alphabetic, other graphic (such as parentheses, space, or equals sign) and control characters. We can extend the idea of types into the ways in which we classify data that is to be stored in a computer.

Consider a patient chart that includes demographic information as well as vital signs, laboratory results, a digitized x-ray, and printed copy of a vector cardiogram. Even a simple chart such as this gives us an opportunity to examine the concept of data types. The demographic information usually consists of items such as the following:

- patient name

- identification number (hospital and/or national)
- sex
- address (apartment number, street, city, state, postal code)
- telephone number(s)
- insurance provider
- employer
- etc.

How would we categorize these data elements? To begin with, patient name will consist of alphabetic characters, sometimes including some punctuation marks (-, ', etc.). A decision must be made as to the length of this field, and whether surname is to be stored separately from given name(s). Address, insurance provider, and employer likewise consist of alphabetic fields, although most addresses also include numeric data, hence are called alphanumeric fields. These distinctions are important because the database package can be instructed to check data as it is entered to be certain that it falls within the specified data type description. Since accuracy of data is perhaps the single most important element in database construction, it is essential to define each data element as completely as possible.

Date of birth is clearly a different data type, not alphabetic, but also not numeric in the usual sense. All database packages allow for a 'date' type, and most also allow 'time' to be specified. Error checks can be built around 'reasonable' dates for given fields: a date of birth cannot be a future date, but it should not be more than 100 years ago (except in rare cases).

There are several fields in this list consisting of numbers (telephone number, identification numbers). However, these numbers are not used for calculations in the usual sense. Telephone numbers are not averaged, and they are usually entered with a dash (-) between groups of numbers. A hospital identification number may have a 'check digit,' calculated based on the previous numbers in such a way as to catch most input transcription errors, but it has no other numeric significance. These fields are sometimes referred to as 'string numeric fields,' indicating that they contain numbers but are not strictly used for calculation.

Sex is a data field that (in humans) may be either male or female, but no other value is accepted. This data type is a part of an 'enumerated set,' and it is necessary to define all possibilities so that the database package can check to make sure that a valid data element has been entered.

The demographic database illustrates several data types. Let us turn next to some clinical information (using only a small sample to illustrate some other types). Typical values that might be in a patient chart would include:

- height
- weight
- body surface area
- attending physician number
- chief diagnosis
- other diagnoses

- laboratory data
- Xray
- vector cardiogram

The first two data elements are numeric, with well defined ranges of reasonable values, depending on patient age, sex, etc. Numeric values may be integer (height in cm.), or they may be decimal (weight in kg). There are other special numeric forms: monetary values (salary, billing data), and very large numbers (not often seen in patient charts) are two examples. Numeric values can be used in calculations: for example, body surface area might be calculated from height, weight and other parameters. Hence we have a new data type: 'calculated,' in which the value is not entered directly, but instead it is calculated from other values.

The attending physician number and ICDA codes for diagnoses are also numbers, but these numbers represent yet another data type called 'pointers.' They are actually codes, used to look up information from other files. For instance, the physician number field 'points' to a file in which physician data is stored, sorted by the identifying physician number. It makes no sense to store in each patient file all information about physicians in a practice, and the use of a code reduces the size of the field in the patient file. Although ICDA codes may be primarily numeric (some include letters), they are also codes, used to look up disease names in a separate dictionary.

Laboratory data will often be numeric, although other enumerated sets are used in many cases. However, the 'data types' of vector cardiogram and Xray are much more difficult to define. If a vector cardiogram is printed and stored in a chart, it may be necessary to identify it by code number, date or some other unique characteristic. If the Xray is stored as a film, it is usually filed in a separate place from the patient records, and those files are organized in a different manner than the patient chart. If an Xray is digitized, a separate computer file is used to store this information (in fact, it is often on a separate computer). Hence, there is need to leave options for several types of graphic input. Since most database packages do not provide for graphic data, we will simply classify them as 'graphic' for the purposes of this chapter.

From this overview of a simple database, we see that a limited set of data types can define almost all data elements in typical databases. These include:

- alphabetic string
- alphanumeric string
- date/time
- enumerated sets
- string numeric
- numeric (integer, decimal, monetary, large)
- calculated numeric (or calculated date or time intervals)
- codes (pointers to other dictionaries or files)
- graphic

This list covers all data types found in common database systems. It is the responsibility of a database designer to determine which type is appropriate for

each data element, and what are reasonable values or string lengths for each field.

### 3.2. Other data attributes

Although the preceding list includes every data type found in today's database packages it does not cover all information required to describe data elements in a database system. The remaining attributes of each data element help the database user to maintain a system in which details are as complete, accurate, and secure as possible.

Certain data elements are *required*. For example, a patient record must include the identification number, name, and several other data elements. On the other hand, other data elements may be *optional*: there may be no employer or insurance provider; diagnostic information may not yet be available; only some laboratory tests may have been performed, etc. The decision as to which fields are mandatory is a management choice that must balance the need to provide complete information against the difficulty of obtaining certain data elements (e.g., with emergency room patients).

Some information may be confidential, and as such it must be classified as *restricted* to certain classes of users. The options available for restricting access to certain data elements vary with different database packages. In clinical charts, however, it is essential to define properly access privileges for many data elements.

The partial list of clinical data illustrates another characteristic associated with certain data types. Many items in a clinical record change over time: patient weight, laboratory results, vital signs – in fact, most clinical results have *multiple* values, each representing a different time at which the measurement was taken. Providing for multiple fields is one of the most difficult problems in defining databases, as we will see later.

These examples illustrate the general types of auxiliary data type attributes that a database designer must consider. Each database package offers a variety of such options, and it takes experience to recognize which of these attributes are vital for given applications.

## 4. File design

Having selected the data types that will be used for each data element, the database designer must next make decisions as to how to group data elements together in files. To do so, the designer must understand what sorts of files are available, and know the advantages of each file type. Most database packages provide very little guidance in this area, but an awareness of different options is useful in efficient database design. In this section, we will discuss file types, and then deal briefly with benefits of certain file types for certain applications.

#### 4.1. File types

A *file* consists of groups of *records*, each of them structured in the same general manner, with several *fields* appearing in a logical sequence. Each field is defined in terms of its *data* type, which describes the character codes permissible in the field.

There are, however, many different categories of files. For our purposes, we can divide file types into two broad categories: *fixed-length fields* and *variable-length fields*. A fixed-length field is one in which each field is predefined to consist of a specific number of characters. A postal code in the USA would be defined to consist of exactly five numeric digits. Fixed fields are useful in organizing records, since the database package knows that a given field will begin at a specific character position in a record. For data elements such as postal codes, fixed length fields make sense. In other cases, such as addresses, names, and similar text-oriented data, defining a fixed length for a field results in waste space in a computer file. If the average length of a last name is, say, seven characters, but the maximum length anticipated is 18 characters, then, on average, 11 characters of record space will be wasted in that field alone.

This example illustrates one of the typical design decisions inherent in defining a database. The designer has to determine what is a reasonable length for a field, recognizing that increasing the size results in greater waste space, whereas decreasing the size may prevent entering complete information for that data element. Most database design decisions present similar tradeoffs, as we will see throughout this text.

A file that consists of fixed length in each record will automatically have a *fixed length record*. This means that such a file can be designed to take advantage of hardware features to achieve maximum efficiency. Knowledge of the hardware and software issues associated with this type of efficiency design is beyond the scope of this text; an experienced systems programmer would probably be the best person to assist in such a design process. However, knowing that such efficiencies can be achieved is something that will help a clinician to discuss the design problem with such experts.

**4.1.1. Variable-length record file types: MUMPS and Pile.** Variable-length fields generate records that are also variable in length, although there is usually a limit prescribed for the maximum length that such a record might attain. In variable-length record files, there has to be a way to determine when one field stops and the next begins. There are two common ways to identify individual fields in a variable-length record. The first is to insert a special character, called a *delimiter* between each field. For example, a patient demographic record might look like this:

Pincioli/F./983-26-2789/20OCT1940/Piazza L. da Vinci, 32/Milano/20133/  
Italy/



In such a record, the *sequence* of data elements is specified (last name, first name, identification number, date of birth, street address, city, postal code, country), but the length of each field is variable, and the start of a new field is determined by the presence of a '/'. Another record might contain the following data:

Range/Jon/783-99-1387//189 Stratton Rd/Hampshire//UK/

In this case, the date of birth and postal code are not known, but a '/' must be present to place other fields in their proper sequence. Some computer languages (MUMPS is the best example) make use of records of this type to achieve much more efficient storage of databases. However, processing records of this type will require more machine time. Once again, the tradeoffs of design require a knowledge of the real (not hearsay or imagined) benefits of each design choice. In this case, for example, it is commonly stated that processing such records is 'too slow' to permit their use in clinical databases, but numerous studies have shown that, using MUMPS, this is not the case when compared with other well-known database packages.

The second way in which a variable length field can be defined is by inserting a *count* of the number of characters in the field to follow. For example, the last example above might be redefined as follows:

5.Range4.Jon11.783-99-1387015.189 Stratton Rd9.Hampshire0.2.UK

In this example, the value '5.' means the following field is 5 characters long, and the value '0.' indicates an empty or blank field. (There are ways to encode numbers that make it possible for computers to distinguish a single value, even one that is more than one digit long, from following numbers. We will not concern ourselves with this form of internal representation, using instead a decimal point to indicate the field length count.)

Some variable length records consist of an arbitrary sequence of fields. In such a case, it is necessary to define the type of field as its contents. For example, a portion of a patient clinical record might be designed as follows:

NAME = 'Walters'/ID = '99976'/DX1 = '524.1'/DX2 = '324.0'/STATUS = 'hospitalized' /...

This type of file is referred to in technical terms as a *Pile*. It is characterized by having variable length, variable sequence fields, and each new entry is appended to the end of the existing file. Records of this type are useful in recording every transaction that has taken place in modifying a database. For example,

REC.#	LENGTH	DATE / TIME	
1	072	032290/140315	FILE="ADMISSION"/REC="3190"/CODE="ADD"/DATA="MICHAEL JONES"
2	102	032290/140610	FILE="LAB_EXAM"/REC="2947"/CODE="UPDATE"/FIELD="RED_CORP" /VALUE="4700E3"/EX_DATE="900322"
3	040	032290/140705	FILE="SERVICE"/CODE="PRINT"
.....			
	051	032590/151530	FILE="ADMISSION"/REC="3190"/CODE="DEL"

Figure 1. Example of a Transaction Log in 'Pile' format. Each record has its own information about length, date and time, plus names and parameters of operations performed stored in its fields.

For most other database purposes, however, it is more advantageous either to use a MUMPS type of variable length record, or to design fields with fixed-length fields and records. Most current database packages use fixed-length fields for the data elements in a database. Therefore, it is useful to know the different types of fixed length record files that can be used.

4.1.2. *Fixed-length file types: sequential file.* A typical database file will consist of multiple records, each having the same types of information in fixed-length format. For example, Figure 2 illustrates a hypothetical hospital ward census file.

ROOM	BED	PAT. ID#	NAME	PHYSICIAN	ADMISSION
E31	1	59817	MCFLY	MEESTER	2/26/90
E31	2	78710	SMITH	FRIEDMAN	3/10/90
E32	1	60850	PITTS	FRIEDMAN	1/31/90
E32	2	68391	REDS	MICHAELS	2/15/90
E33	1	50910	CARRUTHERS	SPENCER	3/23/90
E33	2	67801	COLLINS	MEESTER	2/2/90
E34	1	71784	FOX	MEESTER	3/3/90
E35	1	74780	GRIFFITH	MICHAELS	3/22/90
...	...	...	...	...	...

Figure 2. Example of a 'Sequential File'. Records are fixed length sorted in order of Room # and Pat. ID #.

There are two fundamental features that characterize this file. First, it is fixed-length, both in terms of the field and the length of each record. The file is maintained in sorted order, so if a new bed is added to a room, the file must be rewritten with the added bed inserted. Second, the file is sorted according to one or more *keys*. In this case, the primary key is the room number, and a secondary key is the bed number in each room.

Sequential files are very frequently found in databases. They are useful for presenting information from a single *viewpoint*, in this case the room and bed numbers. They require no separate index files that must be maintained, and they

are simple to process. Using computer merge methods for updating, it is relatively simple to create a sorted list of update information, then read the two files, updating outdated records as they are encountered.

On the other hand, there are some drawbacks to Sequential files. The first shortcoming relates to the single viewpoint referenced above. This file is useful for producing a daily bed census, but what if the information is needed for other purposes? Suppose for instance, that a receptionist at the hospital entrance needs to tell a visitor the room number of a specific patient? This type of query represents a second viewpoint, one for which the file as illustrated is not suited. In order to locate a specific patient, the receptionist would have to search the entire file, since the names are not in alphabetical order.

The second problem is a more technical one, relating to computer execution speeds. Computers can access any record in such a file directly, provided that it is stored on disk (not on tape). However, each disk access represents a significant delay in computer operations. Data stored in a computer's main memory is accessed much faster, and database packages take advantage of this fact when they seek to increase execution speed. To illustrate, suppose it takes about a tenth of a second to access a specific record in such a file from disk, whereas it can be accessed almost instantly from main memory. Next, let us assume that we have a hospital with 500 beds. The fastest search technique for a sequential file is called a *binary search*, in which the computer divides the file in half, examines the middle record to see if the desired name is in the first or second half, then repeats the operation for the appropriate half-file, and continues until the record is found. If there are 500 beds, the computer would have to retrieve a maximum of 9 records in order to find a specific name. At a tenth of a second for each retrieval, 0.9 seconds are needed to perform the search. While this delay may not seem too damaging, it is only a portion of the processing time needed for the receptionist to give the visitor the desired information. If other techniques can be found that obtain the answer more quickly, then they would probably be preferable.

In this example, we have introduced another criterion for measuring the performance of a database file, namely, the retrieval speed. It is a general rule of database design that update efficiency and retrieval efficiency are conflicting criteria; to improve the first usually slows down the second. It also points out the increasing need for rapid responses to interactive queries (in the old days of batch processing large computers, requests were usually answered in hours or even days, and interactive queries were out of the question.)

As we can see from this example, there are a number of different factors associated with the decisions affecting choices of files for specific databases. Later in this chapter we will summarize these criteria, comparing different file types in the light of each measure.

*4.1.3. Fixed-field files: index sequential.* The next file type is probably the most common in database packages running on large mainframe computers. This file type is called an *Index Sequential* file, because a secondary index file has been

created to allow for more rapid retrieval in interactive mode. Figure 3 illustrates the concept.

BLOCK.REC	Pat. ID#	NAME	OVERFLOW PTR.	INDEX	
4.1	47911	MICHAEL JONES	NULL	KEY	PTR
4.2	48141	HERBERT DOW	NULL	47911	4
4.3	50910	STEVE CARRUTHERS	NULL	59817	5
4.4	56412	JOHN GOULD	41.3	71784	6
4.5	58189	JOAN CARPENTER	NULL		
5.1	59817	GEORGE MCFLY	41.5		
5.2	60850	JOSEPH PITTS	41.1		
5.3	68391	RODNEY REDS	NULL		
5.4	69800	CHARLES COOMBS	41.2		
5.5	70497	MICHAELA MURPHY	NULL		
6.1	71784	ELIZABETH FOX	42.5		
6.2	78710	MARC SMITH	NULL		
6.3	83811	ANN CONNORS	NULL		
:					
:					
:					
41.1	61810	LUKE FRANCIS	42.1		
41.2	70189	CRISTOPHER STONE	5.5		
41.3	57811	SARAH FIELD	4.5	OVERFLOW 1	
41.4	65281	LAURA DOUGLAS	42.3		
41.5	60001	ISAAC GOLDBLUM	5.2		
42.1	63711	BARBARA SHORT	NULL		
42.2	64983	GORDON LEEDS	41.4		
42.3	67801	PETER COLLINS	5.3	OVERFLOW 2	
42.4	74780	ALEXANDRA GRIFFITH	6.2		
42.5	72810	SUSAN MASON	42.4		
:					
:					

Figure 3. Example of an 'Index Sequential File'. Records are grouped into physical blocks, sorted by Patient ID #. A separate file indexes the first record of each block. Overflow areas are available for inserting new records in sequence, with pointers indicating the correct insertion sequence.

Several features distinguish this file type from the sequential file described earlier. The main file is sorted according to a primary key (in this case patient ID number), but the entire file is divided into *blocks* that contain several records. Several blocks are filled, but interspersed at regular intervals are empty blocks. In addition, each record has a new field added at the end, also empty at the outset.

The index file is a separate file, designed to be small enough to fit in main memory. (In the case of very large files, there may be a higher level index file referencing the primary index, so that this higher level file can reside in memory.) The index file contains the first name in each block, and a pointer to the block number in the file.

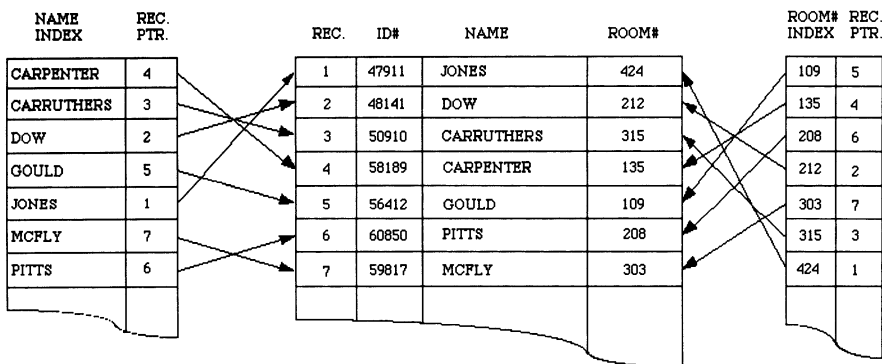
Blocks are designed so that an entire block may be accessed by a single disk read (most operating systems retrieve a block of data rather than a single record each time a disk read is executed). In such a system, therefore, the search for a

patient name would involve first, examination of the index file (in main memory) to determine which block contains the desired patient record, then a *single* disk access in order to obtain that record. Compared to the 0.9 seconds for the sequential file access, this access would require only 0.1 seconds.

Another feature of this file structure improves the update performance in an interactive system. Suppose a new patient is registered. The name probably belongs between two names in a block that is already full. Instead of rewriting the entire file, the patient record is placed in the nearby empty block, and a *pointer* to that record is inserted in the record just previous to the new patient, while the new patient record's pointer field identifies the location of the patient record immediately following. In this manner, a number of updates to the file can be accomplished interactively with only minimal retrieval slowdown in cases where the desired patient name is in an overflow block.

Because the index sequential file offers reasonable efficiency in both retrieval and update, it is commonly used, especially in older database systems. There remains one major drawback, however, which is that index sequential files are designed for only one viewpoint. Searching for all patients on a specific ward, for example, would require a search of the entire database.

**4.1.4. Fixed-field files: indexed file.** In order to serve the interactive needs of multiple viewpoints, database researchers have invented a different type of file called the *Indexed File*. Figure 4 illustrates the structure of one such file.



*Figure 4.* Example of an 'Indexed File'. Indexes are created according to Patient Name and Room #. Masterfile is not kept in any particular order.

This file type differs from previously described files in two important respects. First, the records are *not* sorted according to any key. Instead, new records are added to the end of the file as they occur. Second, there may be several indexes to the file, not just one. In our example, we have two indexes, one for patient name, the other for room number, so as to provide rapid answers to questions from two different viewpoints. Other index files could be added as other viewpoints are identified. In addition, the main file contains no pointers,

thereby reducing the record size slightly.

Index files can provide answers to many questions simply by looking at the index files, without actually accessing the main file. For example, if one wanted to know how many patients of Dr. X are in Ward Y, it would be necessary to create a physician index, then, using that index and the ward index, one could obtain the answer without reading the main file. This type of data analysis is common, and cardiology databases could in many cases benefit from application of such a file structure.

In fact, this type of file is extremely common in today's PC-based database packages. Packages such as dBASE III and DATAEASE maintain indexed files as their normal option.

Although indexed files offer very rapid retrieval from multiple viewpoints, they do present problems in updating. It is easy enough to add a new record at the end of the main file, but each index file must be updated by inserting a new entry in the appropriate place for that file. Some database packages separate the indexing process for normal data entry, so that a user may enter a number of new records to a file, then do a single update of the indexes, a process that may require several minutes in typical files. This approach speeds up the data entry process, although interactive query based on indexes cannot access the new records until indexes have been updated.

*4.1.5. Fixed-field files: direct access.* Indexing can be done in several different ways. One method is to use a *formula* (computer people call it an algorithm) to calculate an address directly, without storing a separate index file. When such a method is used the file is said to be a *Direct* file, and the algorithm used to calculate the address is called a *Hashing Algorithm*. These files use the block concept described in the section on Index Sequential, but, instead of filling most blocks and leaving intermediate blocks empty, an appropriate number of blocks are first created, and then the algorithm is used to place records in an appropriate block. An example is given in Figure 5.

This is an example of the use of the *Soundex* coding system. It is frequently used in hospital registration systems, to access patient data on the basis of the patient name. An admissions clerk may have difficulty spelling the patient name correctly, or the name may not be correctly spelled in the register. Soundex simplifies the spelling problem by calculating a code value using the following algorithm:

1. The first letter of the last name is the first Code character.
2. A three- or four-digit number is calculated from the remaining letters of the last name to form a code, using the following rules:
  - a. The remaining letters are processed in order. Numbers assigned are
    - 1 if B,F,P, or V
    - 2 if C,G,K,Q,S,X, or Z
    - 3 if D or T
    - 4 if L
    - 5 if M or N

	NAME	DOB	ID#		NAME	DOB	ID#
M320	MIDGE	4/06/61	87901	\$152	SPENCER	8/21/62	25432
	MITCH	7/21/55	69012		SPENSER	2/26/63	18590
	.....				SPANGLER	10/05/59	45789
					.....		
M324	MITCHELL	11/28/40	23977		.....		
	.....						
M325	MADISON	3/09/51	14598	\$315	STEVENSON	11/04/55	38745
	MATHIESON	12/27/66	65431		STEPHENSON	1/01/64	88218
	.....				.....		
M625	MORRISON	5/21/58	24644	\$363	STUART	11/06/55	44987
	MURCHISON	10/03/66	76534		STEWART	5/12/61	62888
	.....				.....		

NAME	SOUNDEX CODE	NAME	SOUNDEX CODE
MADISON	M325	SPANGLER	\$152
MATHIESON	M325	SPENCER	\$152
MIDGE	M320	SPENSER	\$152
MITCH	M320	STEVENSON	\$315
MITCHELL	M324	STEPHENSON	\$315
MORRISON	M625	STUART	\$363
MURCHISON	M625	STEWART	\$363
.....		.....	

Figure 5. Example of a 'Hashed File' using Soundex Code. Names sounding alike tend to cluster together.

6 if R

no value is assigned if the letter is H,W,Y, or a vowel

- b. If two letters with the same values appear together, the repeated values are deleted
- c. If less than the 3 (or four) digits are calculated, the number is padded with zeros.

The resulting file works remarkably well to split western names into groups that are distributed in the file in a manner that is closely related, but not specific to the spelling. Soundex look-up programs generally give the user the most likely block calculated from the value nearest to the one generated by the name types by the user. If that block does not contain the correct name (verified by such additional entries as date of birth, address, sex, and hospital number), adjacent blocks are retrieved until the correct name is found or it is evident that the name is not in the file. Hospital registers of a million or more names have been encoded in this manner, with very satisfactory interactive response times.

Direct files are sometimes called *Hashed Files*, because the codes result in distributing the values in a manner that appears inconsistent with normal spelling or with sequential numeric values. They are good for Soundex files, but they are rarely used in other settings. One problem with direct files is that they only present the user with a single viewpoint (in the case of Soundex, the patient name). Other viewpoints can only be referenced by reading the entire file until the desired record is found. In addition, the hashing algorithm must result in an acceptable distribution of the files within the blocks allocated. Otherwise, some blocks would be empty, others full or overflowing. Consider the situation when a large number of patients with Chinese family names are coded using the Soundex hashing algorithm. Common family names such as Li, Lee, Low, Lai and Lou all calculate to precisely the same Soundex value (L000 or L0000). Since many Chinese names are monosyllabic and tend to have few consonants, the Soundex system is not satisfactory for such a patient population.

In other words, the direct file is only as good as its hashing algorithm, which is data-dependent. It may work for some databases, but not for others. Sometimes, as a file grows, it is necessary to revise the algorithm, and this may require a careful study of the past, present, and predicted nature of the key on which the hashing algorithm is to operate.

#### *4.2. Special files for indexing: balanced trees (B-trees)*

The Hashing technique described above requires knowledge of the database. There is, however, another indexing technique that operates quite well independent of the nature of the data keys. This technique is referred to as a *Balanced tree*, or *B-tree*. It is a system that guarantees maximum efficiency in searching for keys by maintaining a balance in the file. In its basic form, the B-Tree consists of a *Tree* of key values, and *Leaves* in which the remaining data are stored. Figure 6 illustrates the general design of this type of a structure.

Note that the top entry in the tree portion of this file is not the first key, but rather one close to the middle of the values present. Searching a B-tree is more efficient than the binary search described earlier, since a number of different choices, not just two, are available from each block in the tree. In fact, it is quite common to have as many as 100 keys in a single block. In this case, it would require no more than three accesses of B-tree blocks in order to locate a record in a file containing up to a million entries!



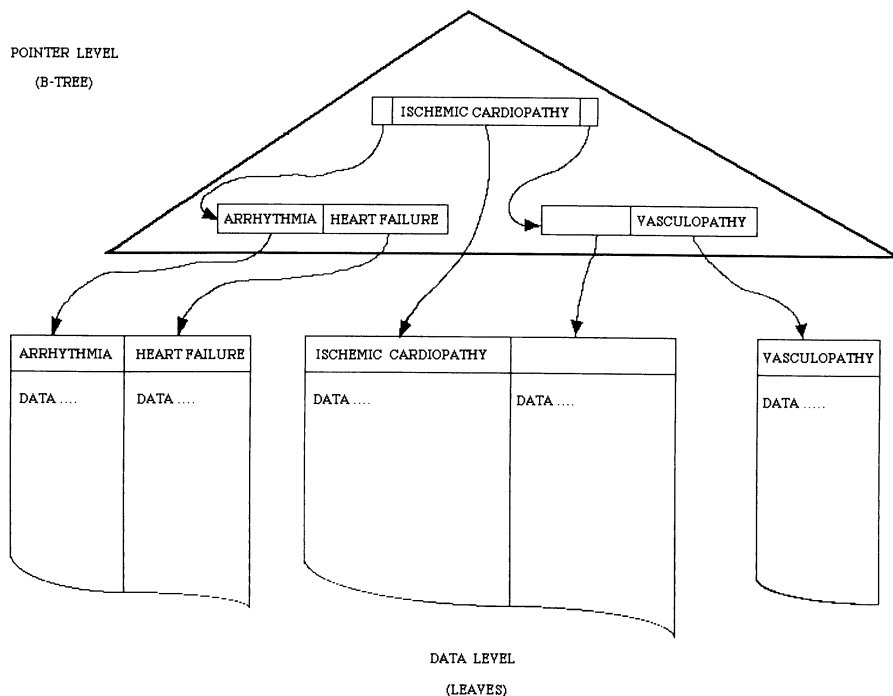


Figure 6. General structure of a 'B-Tree'

Although B-trees are often used as single viewpoint entries to files stored in the leaves of the same file, it is also possible to store *pointers* to another file in the tree, omitting the leaves entirely. This is the technique used to achieve multiple viewpoints in some database systems. Figure 7 illustrates this concept. In this example, two independent B-trees are constructed, each representing a single viewpoint. The B-trees point to a separate file, which is a non-sequential, fixed-field file containing the complete data for each patient.

The structure shown in this figure is precisely the structure of an Indexed File described earlier, with a B-tree used for indexing (instead of a sequential file). This system can give excellent query response times, and for this reason it is used in a great many database packages (dBASE, Dataease, and others). The design is most efficient for lookup processing, but it can also be reasonably efficient for inserting or deleting records, since the file does not have to be completely restructured for such updates.

B-trees may use fixed-length or variable-length keys. Although most commercial systems rely on fixed-length keys, the MUMPS language makes very efficient use of variable-length keys to increase greatly the versatility of its design.

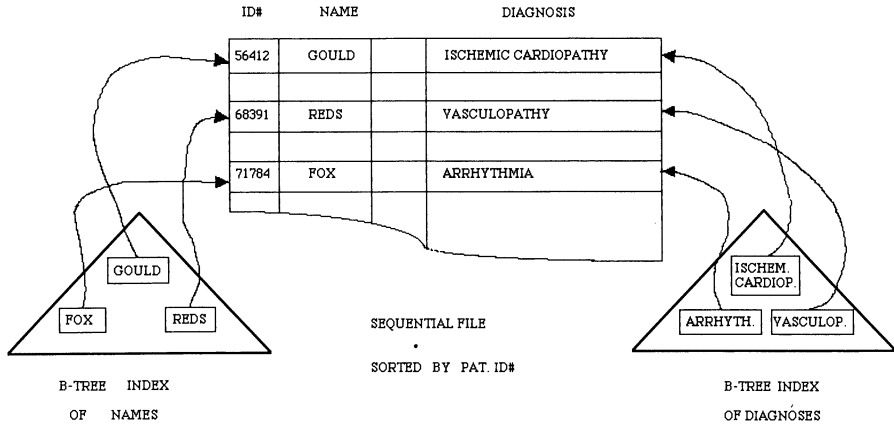


Figure 7. Example of the use of 'B-Tree Files' to index a Sequential File.

### 4.3. Summary of file structures

The process of designing a database requires a knowledge of the data that will go into the database, the ways in which the data will be used, and the options available for defining the database structure. Many different operations go into the use of databases, and each may present unique problems to the database designer. Table 1 summarizes the most common computer operations performed on a database, and gives a relative ranking of the strengths and weaknesses of the different file structures described earlier.

Table 1. Performance characteristics of file types

Measure	Pile	Sequential	Index seq.	Indexed	Hashed	B-tree
Storage efficiency	Poor	Best	Good	Medium	Medium	Medium
Get record on key	Poor	Medium	Good	Best	Good	Good
Get next on key	Poor	Best	Best	Good	Poor	Medium
Get record multikey	Poor	Medium	Medium	Best	Poor	Medium
Add new record	Good	Poor	Medium	Good	Good	Medium
Modify record	Poor	Good	Good	Good	Good	Good
Delete record	Medium	Poor	Medium	Medium	Good	Medium
Read whole file	Medium	Good	Good	Good	Poor	Medium
Reorganize	Not Available	Medium	Medium	Medium	Poor	Medium

The operations listed on the lefthand side of the table are described below.

**Storage Efficiency:** This measure takes into account the amount of space used by the database in comparison with other files discussed.

**Get record on key:** This is an indication of how well (comparatively) the file type performs in accessing a record based on the primary key (if any).

*Get next record on key:* In many applications, it is desirable to read records in order of the primary key, one after another. File types perform differently in this respect.

*Get record on multikey:* This measure describes a search based on more than one viewpoint. There is a wide variation of performance in this category.

*Add new record:* Updating a file may involve adding a record, and performance of different file types are quite varied in this respect.

*Modify record:* Another form of update involves changing the value of one or more fields in a record. This attribute is common, and performance tends to be good in most of the file types (except Pile).

*Delete record:* Removing a record from a file may cause a number of side effects, such as updating index files. This measure summarizes file type performance in deletion of records. Note that deletion of a large segment of records is a different matter not considered in this summary.

*Read whole file:* Surprisingly, there is a difference in reading the entire file for certain file types.

*Reorganize file:* This category relates to indexed files (including the implicitly organized Sequential File). This factor is especially important for dynamic files which are frequently modified and updated.

Careful analysis of this table will show that there are no ‘winners’ and losers. Rather, each file has strengths and weaknesses that can be exploited for particular applications. Often, the end user of database package will not even know what types of files are being used (many actual files may represent hybrid variations combining different features of the files described above). However, sometimes the design of files proves to be an important factor in selecting a database package, and it is usually possible to determine which file types are used.

## 5. Database models

When a database is constructed, it will usually consist of several different files, using variations of the types described above. The individual files must be linked together in order to form a complete database system. There are several ways in which such links can be accomplished. The combined process of selecting file types and associated file relationships depends on the *database model* that is used. In this section we introduce the principal models currently in use in today’s database packages, referencing database packages using each.

### 5.1. User viewpoints

In the previous section, we frequently commented that an important aspect of some database systems is that they will be accessed by several different types of users. It is much easier to design a database application if only one user is

involved, but rarely do real systems conform with this approach. Consider, for example, a patient record file that includes the diagnoses of each patient. From the patient's point of view, a file structured to give all information organized by patient ID is ideal. A physician, on the other hand, will want to retrieve only those patients for whom he or she has responsibility. A specialist may want to study all cases of a particular disease. An administrator may need information on all unpaid patient accounts. Designing a database system to meet any one of these needs is simple. Designing it to meet all of those viewpoints requires making compromises that cannot be optimal for any single viewpoint without penalizing some other viewpoint. Satisfying all viewpoints equally may result in a system that is too slow to meet the needs of any of the users.

Some database models emphasize a single viewpoint, others try to provide multiple viewpoints, sometimes paying a performance penalty as a result. However, as we will see, there are other tradeoffs that must be taken into account if one is to find the best database model for a given application.

As this chapter is being written, research on defining new models is taking place. Knowing the state of database model concepts at this time may serve as a valuable guide in evaluating the benefits of new models as they appear.

## 5.2. *The hierarchical model*

From a classification point of view, we tend to regard many knowledge domains as hierarchical. Consider the normal notation used for organizing lectures, papers, etc. The headings I, II, III, and so on are followed by subheadings A, B, C, D, which in turn are subdivided into categories 1, 2, 3, subcategories a, b, c, and so on to smaller and smaller subdivisions of the information to be organized. To take another example, consider the classification of living things: Kingdoms, Phyla, and so on.

Inherent in each of these classifications is the ability to have an *unequal* distribution of categories at each level. There may be five items (A,B,C,D, and E) under major heading II, whereas III has only A and B. In the Animal kingdom, there are only 16 species of penguins, but hundreds of species of many insect genera.

A second fundamental fact about hierarchical classifications is their adherence to a single *viewpoint*. If one wished to classify all flying creatures, one would have to examine the entire hierarchy of plants and animals to determine which groups are able to glide or fly of their own volition. One might develop a different classification, based on means of flight, in which the normal biological classification was replaced by a more specialized one representing a different viewpoint. In fact, the MeSH headings used by the U.S. National Library of Medicine for the Index Medicus classification contains not just one but many different classifications. Anatomic, pathologic, diagnostic, and other viewpoints are independently defined, and a single reference may be indexed according to several of these hierarchical classifications. SNOMED is another multiple classification that is essentially hierarchical in nature, providing ways

to define anatomical, etiological, functional, and morphological features of an item of medical interest, be it a biopsy slide, patient diagnosis, or other.

The hierarchical database model begins from the same basic approach: it defines data in accordance with a single viewpoint defined by the user, and all records are stored hierarchically in that form. The patient record naturally lends itself to hierarchical classification. In the Problem-Oriented Medical Record (POMR), progress notes are organized as follows:

```

Patient
. Problem
.. Encounter Date
... Subjective findings
... Objective Findings
... Assessment
... Plan

```

Other subdivisions may be added, but the classic 'SOAP' format for this type of progress notes derives from the first letter of each category under the entries for each patient encounter.

Hierarchical models have been in existence ever since computers were first used for database support. A very few large mainframe hierarchical database packages such as the *IMS* system marketed by IBM probably account for the majority of files on such systems (or they did until recently). However, in recent years, the advent of newer models has led people to consider hierarchical systems as obsolete. For reasons given below, this assessment may be premature.

There are two shortcomings to 'classic' hierarchical models as implemented in systems such as IBM's *IMS* package. First, they require the definition of one viewpoint as the principal mechanism for classifying items entered in a database. Although means are provided for alternate indexes to give some support for separate viewpoints, the internal structure is such that these packages tend to perform very slowly when alternate viewpoints are used.

Second, the hierarchical model is one that cannot be formally defined in terms of its ability to perform standard database functions. Database packages were usually evaluated by the use of 'benchmarks,' in which a typical set of data were entered into the target package, and a standard set of retrievals were run to see how different packages performed.

Although the hierarchical model is currently held in some disrepute, one package has proven surprisingly successful in competing against other models in a number of hospital information system bidding situations. That package is *Fileman*, a database package that was designed by the U.S. Veterans Administration for use in its 172 hospitals. *Fileman* has also been adopted by other HIS systems, and it owes its current success here in part to the flexibility of its options and the availability of the package on a wide variety of hardware, ranging from personal computers to large mainframes. Public domain versions of this package (written in the MUMPS language) are available from several sources.

In summary, the Hierarchical Model favors a single viewpoint, requiring a user to understand in some depth the manner in which data are stored in a given database in order to retrieve desired information. For batch type computer operations, they are well suited to centralized systems maintained by database professionals. Often, however, these implementations are slow in interactive mode, and as a result the Hierarchical Model has been largely overlooked in designing modern interactive database packages.

### 5.3. *The network model*

In the 1960s, database managers, seeking to find solutions to the multiple viewpoint problem, defined a model that was given the name *Network Model*.

Although there are still a number of network-type database packages around, they are indeed becoming obsolete, and no new work is being done based on this model. We will therefore omit discussion of the rather complicated manner in which the network model is defined and implemented.

### 5.4. *The relational model*

The history of the *Relational Model* goes back to 1970, when an IBM scientist (E.F. Codd) published a paper in which he defined a new approach to database systems which was characterized by being a) viewpoint-independent; and b) capable of being defined (and evaluated) in precise mathematical terms. The paper stimulated computer scientists (who had previously shunned the database field) to take a fresh look at database storage and retrieval, giving birth to a new subdiscipline in computer science.

The easiest way to begin thinking about the relational model is to consider it as a series of tables, each of them called a *relation*. The tables contain fixed length fields and fixed-length records. Each record (row) in a relation is called a *tuple*. Each field (column) is called an *attribute*. Each tuple is defined in terms of a *primary key*, and there can be no duplicate tuples in a relation. The sequence of tuples in a relation is not defined; in other words, there is no sequence according to a key. When the relation is first defined, the sequence of attributes is also user-definable, but once the relation has been defined, each tuple must have attributes in the same order. Sometimes, a primary key may consist of more than one attribute, as for instance a patient ID and encounter date may serve as the primary key to define an encounter summary relation.

In order to simplify the process of maintaining relations, data elements should be grouped together in such a way as to minimize redundancy. To formalize this process, a set of rules referred to as *normalization* was defined. These rules may be summarized as follows:

1. There can be no duplicate tuples in a relation.
2. There can be no tuples with the primary key missing.
3. The attributes in a given relation must relate to ('depend on') the primary key.

4. No attribute in a single tuple may have more than one value.
  5. Nonkey attributes of a relation must be independent of each other.
- These rules can be illustrated with an example as shown in Figure 8.

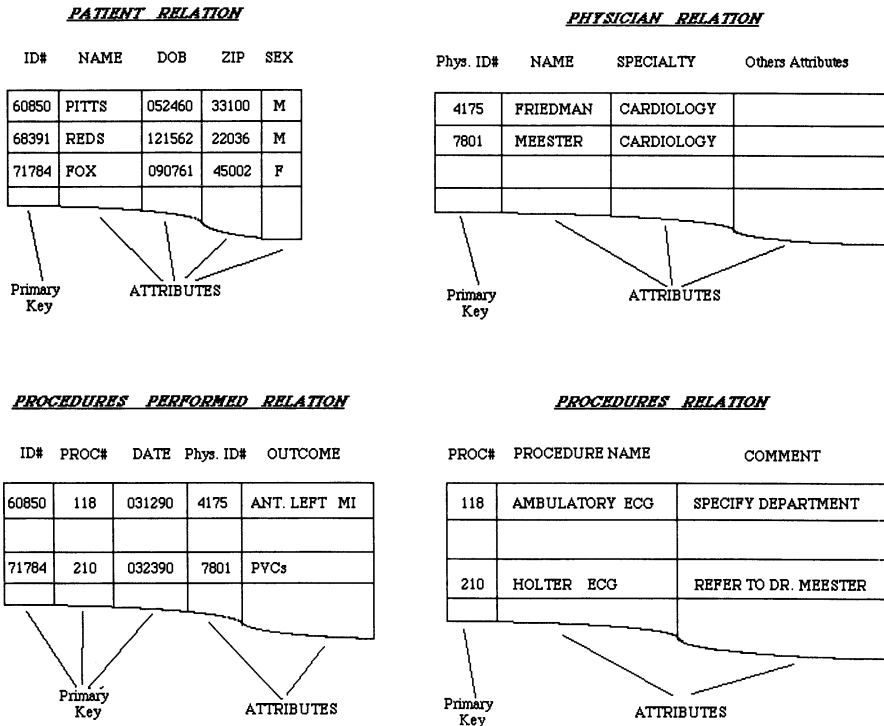


Figure 8. Example of Relational Model built with some of previous used data.

To understand the normalization process used to create these separate relations, let us consider for a moment how these relations are linked to each other. Note that the primary key for one relation may not be a primary key in another relation. For example, physician number is not a primary key in the procedures relation. Also, observe that the primary key may consist of more than one attribute, as in the procedures-performed relation.

Comparing these relations with the normalization rules shows that none of the rules are violated. Rules 13 are self-explanatory. To understand Rule 4, we must realize that, if a physician performs two procedures on the same patient on the same day, the procedure number would change, and a new tuple would be required to enter the necessary information.

The final rule can be explained by considering an alternative design. What would be the effect of including physician name and/or procedure name in addition to physician number in the Procedures-Performed relation? The chief difficulty would come in correcting mistakes (which *always* occur in data systems). If it turns out that an error has been made in the spelling of a

procedure or physician name, the spelling error might or might not be repeated in the Procedures-Performed relation. It would be much easier to correct the spelling error once, in the Procedures relation (or the physician name in the Physicians relation), and avoid the possibility of conflicting or inaccurate values in more than one location.

This example illustrates the general way in which files must be defined in the relational model. This type of model is viewpoint-independent. Since there is no requirement that tuples be ordered in any sequence, new tuples can be added at the end of the file as they occur. On the other hand, unless there are indexes reflecting individual viewpoints, the relation would have to be read beginning to end in order to find each pertinent value. For this reason, most implementations of the relational model permit users to specify which attributes will be used to index the file.

The second major feature that is facilitated by the relational model has to do with the operations that can be performed on a database created using this approach. The simplest and most common operations are those that can be performed on a single relation. The first is *restriction*, involving searching for all tuples in a relation that meet certain criteria. For instance, one could select from Procedures-Performed all tuples describing the procedures done by a specified physician between two dates. An algebraic way of stating this retrieval request would be

```
SELECT from PROCEDURES-PERFORMED where Phys-ID = "29784" and
DATE >31 Dec 1988 and DATE <1 Jan 1990
```

which would create a list of all procedures performed by the named physician during 1989.

A second operation on a single relation is a *projection*, involving the listing of only certain attributes of a relation. For example, one might list only patient ID, date of birth, and sex for a study in which patient names should be held confidential.

The most powerful operations on relations are those that involve two or more relations at a time. Suppose we wanted to create a list with patient name, surgeon name, and the name of each procedure performed, using the relations given above. To do so, we would have to *join* the relations, obtaining the patient names from the Patient relation, the physician names from the Physician relation, and the procedure names from the Procedure relation, and defining the associations between them based on the Procedures-Performed relation. This join would usually be combined with projections (eliminating other information about the patient, physician and procedure). Linking relations is done by finding two attributes, one in each relation, with the same values and combining the appropriate information from each. This type of join, often called a *natural join* is probably the most common operation on relational files.

Another type of query on a database might involve comparing two relations



in a different manner. Suppose we had two Patient-relations, one derived from a hospital census, the other representing all patients with the same primary physician. By performing an *intersection* between these two relations, one could determine which patients of that physician are hospitalized on some specific date. The opposite type of search (finding all patients of that physician who are *not* hospitalized) is called a *difference* operation.

These operations, illustrated graphically in Figure 9, describe in simple terms the ways in which one can query relations in this type of model.

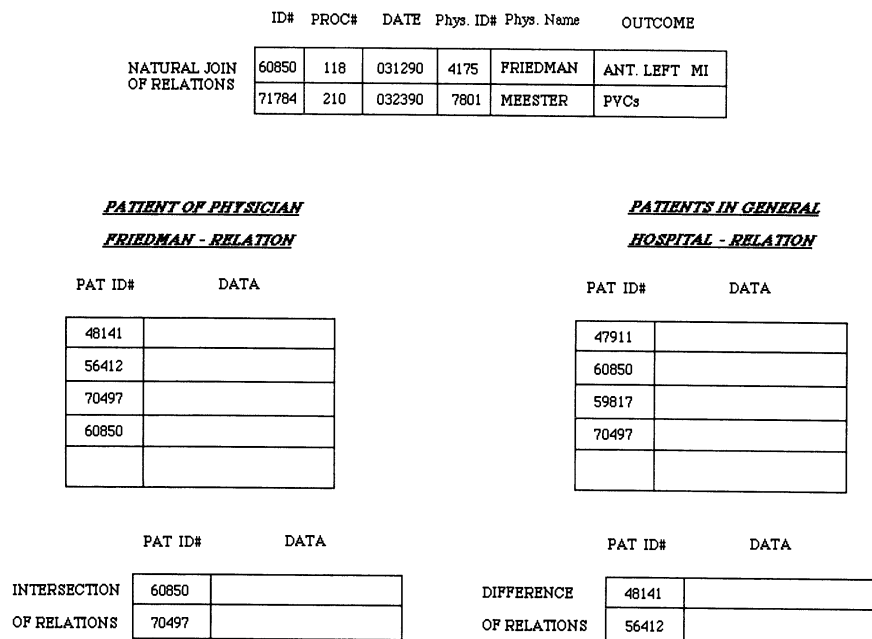


Figure 9. Example of 'Natural Join' of relations from Figure 8, of an 'Intersection' of first and second relations shown above and of a 'Difference' of first and second relations above.

Unlike the hierarchical model, it is possible to define in clear, unambiguous algebraic terms all operations that should be capable of being performed on relations. This is one of the great strengths of this model (another being the flexibility of viewpoint). Databases built on this model are also usually easier for a layman to understand. Often, unraveling the complexity of a hierarchical structure requires specific expertise in the language used to define that model, whereas it is possible for untrained users to become familiar.

There are several different types of languages that have been defined to facilitate manipulation of the relational model. An emerging standard is named *SQL*, for Structured Query Language. An International Standards Organization (ISO) committee is working on an updated definition of this language (provisionally called *SQL2* in its most recent draft form), and, since most of the important vendors of relational model systems are participants in

the ISO standardization discussions, it seems likely that this standard will be adopted within the next few years, greatly simplifying the process of exchanging data between different relational model vendor packages.

There are, however, a few drawbacks to the relational model. The basic design is not necessarily efficient in terms of execution speed. Having no user viewpoint predominate may mean that all user viewpoints will receive equally slow responses. Unless an implementor is careful, the manner in which queries are executed may have a great impact on overall performance. These problems are the subject of current research in databases, but it is fair to say that current relational systems tend to be somewhat slow in many types of operations.

A second major drawback lies in prohibiting a given tuple from having duplicate attribute values. Stated another way, this requirement prevents hierarchical definition of a database. Since, as we have seen, a great deal of patient information is inherently hierarchical, this is a rather serious limitation, one which is also being researched at the present. If it were possible to extend the relational model to contain hierarchical relationships (such extensions have been proposed at the research level), then it seems likely that clinical information could be much more easily incorporated into these models. Several research groups are now exploring this type of extension to the relational model.

### 5.5. *The entity-relationship model*

Another problem with the relational model is its incompleteness in describing our knowledge of information. Consider the following observations:

Incidence of non-specific morbidity has been directly correlated with stress levels in several epidemiological research studies.

High levels of cholesterol increase incidence of congestive heart failure by a significant percentage.

Patient X is the father of Patient Y, and the husband of patient Z.

How can we codify this information accurately, using a relational model? For instance, is the morbidity associated with stress the same kind of morbidity associated with increased cholesterol? Is morbidity a 'characteristic' of cholesterol? How high is 'high cholesterol'? Is fatherhood a characteristic of Patient X (i.e., does it apply to all people with whom X is associated)? To answer these questions, we need to define a new kind of model to store the important information contained in these observations.

A little over ten years ago, another researcher (Peter Chen) developed an extension to the relational model which he named the *Entity-Relationship Model*. In his view, knowledge can be associated with things (objects, people, even ideas), but it can also be associated with *relationships* between entities. One might, for instance have one set of entities consisting of known *diseases*, (rabies, congestive heart failure, etc.), and another set of entities that we will

call *signs and symptoms*. In the disease relation one would enter its pathology, etiology, anatomic site(s), functional impairment, etc. Signs and symptoms could be classified by the manner in which they are detected or measured, their intensity, duration, and so on. Then, one could have a third type of file, called a *relationship*, which links diseases with signs and symptoms and ascribes values to the relationship. The primary key in this file would be a known disease and a known sign or symptom and the other attributes would consist of those factors associated with the link between disease and cause: levels, severity, and other conditions affecting the linkage between finding and disease. We would, in effect be developing a 'knowledge coupler,' in which we identify attributes associated with relationships as distinct from attributes inherent in entities themselves. In our family tree example, we would have a *relationship* relation (how these terms do cause added confusion), in which the primary key consists of two patients, and the association between them (X is father-if Y) is entered as an attribute of the relationship. Using this approach, we don't have to say that patient X is universally a father or a son, but only with respect to the appropriate individual to whom that link applies.

Figure 10 shows the informative nature of an Entity-Relationship Diagram—a means of describing more completely terms the information that can be

<u>ENTITY RELATION</u> <u>"DISEASES"</u>				<u>ENTITY RELATION</u> <u>"SIGNS &amp; SYMPTOMS"</u>	
NAME	PATHOL.	ETIOL.	ANAT. SITE	DESCRIPTION	COMMENTS
ISCHEM. CARDIOP.				CHEST PAIN	
CORONAROPATHY				RESP. DIFFIC.	
VASCULOPATHY				HEAVINESS	
				COLD EXTREM.	
				INSENSITIVITY	
				LOW PERIPH. PRESS.	

<u>RELATIONSHIP RELATION</u>					
ENTITY ATTRIBUTES		RELATIONSHIP ATTRIBUTES			
		DETECT	SEVERITY	DURATION	ANAT. SITE
ISCHEM. CARDIOP.	CHEST PAIN	PATIENT	HIGH	MINUTES	CHEST, LEFT ARM
ISCHEM. CARDIOP.	RESP. DIFFIC.	PATIENT	MEDIUM	MINUTES	
VASCULOPATHY	COLD EXTREM.	PATIENT	VARIABLE	CONTIN.	FEET, HANDS
VASCULOPATHY	INSENSITIVITY	PAT, PHYS	VARIABLE	EPISODIC	FEET, LOW LEGS
VASCULOPATHY	LOW PERIPH. PRESS.	PHYSIC.	VARIABLE	CONTIN.	LOWER LEGS
CORONAROPATHY	CHEST PAIN	PATIENT	HIGH	SUDDEN	STERNUM

Figure 10a. Example of 'Entity-Relationship' Tables, illustrating links between Diseases and Symptoms.

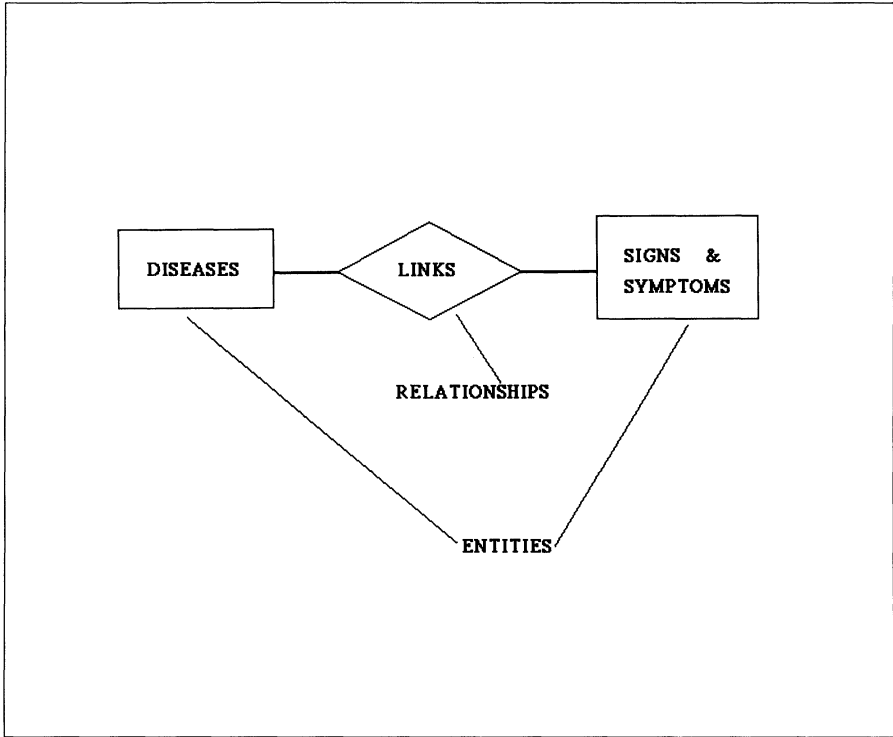


Figure 10b. Diagrammatic representation of Entities and Relationships using data from Figure 10.a.

derived from many real-world observations.

It is surprising that this model has not been more widely accepted by database vendors. There is one commercial implementation available (ZIM) which will be described in the final section of this chapter.

## 6. Designing a database

So far, we have presented some of the most important tools that go into the construction of database systems today. The field is evolving, and new concepts as well as new hardware will affect the way in which we can design and implement clinical databases, but some features will remain important no matter how advanced the technology becomes.

It would be a mistake to assume that a reader of this introduction is ready to design an effective clinical database system, select an appropriate DBMS package, and oversee implementation and maintenance of the system once it becomes operational. However, the goal of this chapter has been to indicate some of the major points that *someone*, or more likely, several people, will need to know in order to design an effective application. The remainder of this text

gives numerous examples of specific problems associated with clinical databases. This section attempts to summarize a few important guidelines that should be followed by *clinicians* who participate in the design process.

### 6.1. *Object-oriented models*

Research in computer languages during the last two decades has developed a completely new type of programming language, called an *object-Oriented* language, which treats program components as objects that communicate by sending messages to each other in a non-sequential manner. For example, a 'printer object' might receive a message to output a given file, and a 'square root object' might be sent a message to calculate the square root of a number. This approach has been extended to the database world, and current research in object-oriented databases seems likely to result in commercial products within the next few years. As no commercial packages are currently available, it is sufficient to note that they may soon appear, so that the reader will at least have some idea of what is meant by this term.

### 6.2. *The role of the physician in database design*

Computer Scientists and database professionals do not and cannot fully understand clinical data and clinical information. Many such individuals, especially the more glib sales types, will profess to know 'all about' how to set up a clinical database. 'Just tell us your needs and show us your data, and we'll fix it up for you' is a phrase that should convince any clinician to look elsewhere for help.

With the central points presented above, a clinician can and should *participate* in the team effort necessary to define an effective database application. As a senior member of such a team, his or her insight is irreplaceable in the design process. Armed with details of design presented earlier, the clinician should be able to participate more fully. An effective way to prepare for such team effort is to list the viewpoints that are likely to be needed in the particular application. If it is a patient database, the viewpoints will include other members of the health care team (nurses, radiologists, pathologists, laboratory technicians, and epidemiologists), and perhaps administrators, bookkeepers, and others. To these viewpoints should be added one more: that of the information scientist, since you will need to bring such an individual into the design process early on. Identifying these viewpoints will be an important way to avoid designing a system that fails to meet the unique needs of one of these individuals.

Having identified the viewpoints, the next crucial step is to find an articulate representative of each of these viewpoints. *Do not try to represent others' viewpoints yourself.* There are many reasons for this cautionary statement. Not only is it hard for you to place yourself in their shoes and try to guess their needs and preferences. Of equal importance is the fact that their participation will

make them feel a part of the design process, and therefore their acceptance of the final product is much more likely to be enthusiastic.

By the same token, *do not allow someone else (i.e., a non-clinician) to present your viewpoints*. If you consider the database design process ‘unimportant’ or something that can be designed by others, it will never meet your needs. It is just as important for the clinician to participate in the design process as it is for each other viewpoint representative. Many physicians have given up hope for computer systems in their practice because they failed to participate actively in the application design.

Include a computer professional early in your design discussions. The kind who can serve your needs best is one who knows how to listen first, then offer some professional guidance on alternatives, but not suggest solutions until all participants can agree that the problem(s) have been clearly stated.

Despite the best planning process possible, *the first design will be incomplete*. No matter how articulate, how careful, the team members and computer consultant may be, no one can fully appreciate the effect of implementing a database system will have on the operation it is to serve. Count on this being true, and design accordingly, leaving abundant opportunities to revise the assumptions, add new viewpoints, or add to the data elements being stored.

The next step involves coordinating the different perceived needs of the viewpoint representatives. This process is a ‘committee’ venture, and it needs someone to guide the committee’s deliberations without restricting creative suggestions. That person should, if possible, be an impartial outsider, *not the clinician*, whose individual biases may otherwise dominate the discussions. Some of the useful exercises for the group to address are:

- deciding on data elements to be entered into a system, beginning with discussion and then asking each participant to prepare a list for a subsequent meeting;
- defining data types, error checks, possible repeating fields, source of input, authority for modification, responsibility for maintenance, and access authority;
- preparing an initial list of reports desired by each user representative: nature of report, frequency, items required, output format (again, it is well to begin by outlining the problem, then requesting users to bring written suggestions to a subsequent meeting; (*Note: be assured that this list will be incomplete, since new ideas will be generated by the availability of the database itself. Assure others that the list is initial, not final.*))
- assigning responsibility for tasks associated with the selection and design of the package, delegating research into questions of hardware, software packages, vendor availability, and other items.
- deciding what current manual practices in data handling might be adjusted or improved prior to implementing a computer-based system (this is an important step, since manual methods may themselves need improvement);
- defining backup procedures to be followed *after* a system is installed, so that users will be aware of the need to provide for possible systems down times

(these procedures should be reviewed after the system is installed, and there should be 'fire drills' periodically to ensure that staff are aware of the procedures as well as the need for contingency planning).

This list is partial, of course, but it suggests an extensive planning process prior to acquisition of a system. An important component of user acceptance is user participation in design. The more individuals see that their needs have been heard and, insofar as possible, incorporated, the more likely they are to support, rather than sabotage, a system when it arrives.

## **7. A brief review of current Database packages: relational models**

There are a great many packages available for many different sizes of computers. These packages are mostly patterned after the relational model, but *none* of them completely satisfies the requirements of the formal relational model. There are also two packages described below that represent alternatives to relational model systems. The intent in this description is not to urge one system over another, but to look at a few of the design considerations behind several packages. Researching new database packages can be a full-time occupation, and results are guaranteed to become obsolete long before a book can be published. Hence, this section should be considered as a guideline for investigation.

The descriptions below contain some strictly personal observations and biases. Please accept them as the opinions of the author, not the result of painstaking and careful evaluations. However, look for some of the factors mentioned as you investigate options yourself in the complex marketplace of changing database packages.

### *7.1. Database packages designed originally for personal computers*

A large number of database packages appeared during the 1980s intended for use on personal computers. They are for the most part relational models. The most famous of these is dBASE, which was the first to appear and is probably the most widely used at this time. This package has evolved from dBASE II to dBASE III, dBASE III +, and now dBASE IV (with prospects of dBASE V in the wings). It is a relational model system with its own query 'language,' but it can be used by novices without having to learn the language. There are a great many dBASE consultants, programmers who will, for a fee, create extensions to dBASE using its programming language. These consultants are naturally biased in favor of a system that will offer them a source of income. Don't be fooled by them into thinking that this is the best, or the only system available. Despite some good points, dBASE has a great many shortcomings, including the inadequacy of its 'language' and inefficient execution when large databases are installed under this package. It has its place on PCs, but it is not really designed to be used in a large database application (e.g., one with files

exceeding 10 megabytes). Studies using dBASE on large files have shown that some retrievals require long periods (half hour or more) of machine time, and that the same retrievals can be done faster on other systems.

In general, database packages designed for personal computers assumed file characteristics of disk systems in the range of 5–20 megabytes, and as much larger systems are now available on personal computers, it is uncertain whether they can compete with systems that were originally designed for larger databases.

Another database package that was designed for personal computers illustrates a slightly different approach. DATAEASE is also a relational model, and it was also designed initially for personal computers (though it now runs on mini-computers as well). However, it has one feature that distinguishes it from dBASE: it provides security of access to files. When a new database is defined, it can be accessed only by user-defined ID and password. Files are protected from user tampering by being stored in such a way that they cannot be edited by programs other than those in DATAEASE (without a major programming effort to get around this security provision). In contrast, dBASE files are available for simple modification, and no security passwords are required. The supervisor of a DATAEASE database can define user menus, restricting different users to different operations on the database (e.g., specific reports and data entry limitations). The system is designed for naive users, with options for expert users also available (dBASE has much improved naive user assistance in its more recent releases).

These two packages characterize some interesting features of personal computer database packages. As you look at others, consider their ease of use for novices, their flexibility for changing database definitions, the speed of input, update, searching and report generation (especially with large files), and the security provisions available to safeguard data and assure its confidentiality.

Another factor to be considered is simultaneous access by multiple users. Most personal computer-based database packages were designed for single users. The typical clinical information system will require multiple users from the start. The question that must be explored is how the package handles multiple users. Be sure to interrogate the vendors thoroughly on this matter (their answers are changing as their systems gradually introduce more multiuser capabilities). Find out, for instance, what happens when two users try to access the same patient's files. In some cases, this might lead to destroyed data!

## *7.2. Database packages for the Macintosh personal computers*

The Macintosh family of personal computers is different in many fundamental respects from the IBM PC family of systems. The operating systems are completely different in design philosophy, and the user interface is also radically different. It is possible to make IBM PC systems perform in ways similar to the customary user interface of Macintosh systems, but the reverse is not necessarily the case.



As a result of this difference, commercial database packages running on Macintosh computers tend to be quite different from those on PCs. Although the vendors of dBASE have fairly recently introduced a version of that package on the Macintosh, not many other PC vendors have done so. The names of Macintosh database packages are therefore different, and they would require separate evaluation if one is to select a Macintosh for clinical database investigation.

In addition to dBASE, some of the other database packages on Macintosh systems include 4th Dimension, FoxBase, and File Maker. No attempt is made to present individual features of these systems, but one package, 4th Dimension, has an interesting feature that bears on some of the discussion in preceding sections. Although it is a relational model database package, 4th Dimension allows users to specify *subfiles* and manipulate those. This feature makes it possible to incorporate a hierarchical approach to file design that makes this package well suited to certain medical applications. For example, it might be possible to create a subfile of laboratory values for specific tests, entering the date, time, and result of each test without having to repeat the basic information about the nature of the test itself. This and other features have made 4th Dimension a very popular system among Macintosh database packages currently available.

Another Macintosh program, Hypercard, is often referred to as a database package, but it does not fit the definition of database systems as used in this text. Hypercard is really an index card manipulation system, allowing users to define index cards of various types and then to link them to each other in a number of different ways. Retrieval of common groups and other standard database operations are not available without using the Hypercard programming language to write specific routines to solve such problems. (Hypercard is also not yet equivalent to the new category of file systems referred to as *hypertext*, a file organization and manipulation system with even greater sophistication, but once again not truly a database system.)

As vendors move into the Macintosh world, more database packages will become available for that family of computers. It is unlikely, however, that real compatibilities between PC and Mac systems will develop, so clinicians will have to make a decision early in their planning process as to whether to go in the direction of PC systems or to develop data using Macintosh computers. At present, the preponderance of database vendor development has been in the PC domain, but it is possible that important changes could occur in the next few years.

### 7.3. *Minicomputer database packages*

A number of packages were designed for the minicomputers of the 1980s (the distinction between micro-, and macro-computers has all but disappeared in the 1990s, since today's micros outperform yesterday's mainframes). Among them, INGRES and ORACLE are perhaps two of the best known relational

model implementations, with SYBASE, another strong competitor. INGRES was designed for a minicomputer known as the PDP11, running an operating system called UNIX. UNIX is a general purpose operating system that happens to have a number of features that make database manipulation more difficult, and as a result, INGRES had some performance difficulties on those systems. Its performance has improved, but performance is definitely something to be very conscious of when a database package runs under UNIX.

INGRES and ORACLE have been ported both to microcomputers and to larger mainframes. On small systems, they tend to use up a large amount of resources (ORACLE requires nearly 10 megabytes of disk storage for the system alone, before data applications are added), so it is unreasonable to think of using a personal computer for any other application if one of these packages is running. On mainframes, the implementations perform reasonably well.

Unlike the personal computer-based systems, these packages have extensive safeguards for database definition, input, update, and retrieval. It is worth studying how these packages describe their provisions for data reliability, integrity, and security, and then to go back to the personal computer systems to see what is missing (it is likely to be a lot)! There is, however, a performance penalty in adding the safeguards.

One other point to look into in evaluating these systems (besides performance) is their communication between systems. In tomorrow's world, it is almost certain that database applications will involve personal computers, mini-computers and even mainframes in a network of systems, each having responsibility for a portion of the total database. Vendor implementation of such distributed system capabilities is changing very rapidly, but one should be very careful to make sure that the needed features are available. For instance, knowing that a package runs on several sizes of systems does not guarantee that the same database package can be distributed transparently across all components of a user's hardware. Some transparency may be available today, other features may be 'under development,' a euphemism that should be heartily distrusted.

A second point relating to distributed systems is the ability of one vendor's packages to interface with data from other vendors. All vendors are coming to realize the importance of this feature, but to date few have really developed products that have implemented many such features. SYBASE is one important exception, worth investigating for that reason alone.

#### *7.4. Mainframe database packages*

Database packages of the 1960s and 1970s were almost all designed for mainframe computers running in batch mode, with overnight turnaround and minimal attention to interactive features. Some of the more recent systems have introduced a greater degree of interactive capability, but the heritage has been hard to overcome. A well-known mainframe package designed around the relational model is DB2, IBM's system which runs on their mainframes and was

designed in part by Dr. Codd, who first proposed the relational model. DB2 is an evolving system ('system R' is a research version that is often used to test new features prior to their incorporation in the commercial product), and it would be hard to give it a fair description within the limits of this chapter. However, it is fair to characterize DB2 as being available primarily on mainframes, with little likelihood that it will appear on other vendor's systems or even on IBM-marketed mid-sized or personal computers. For that reason, we will not consider it further in this text. In general, porting mainframe systems to smaller computers often turns out to be very difficult, with high penalties in performance and systems requirements.

## **8. Alternatives to the relational model: a brief review**

Not all commercial database packages are based on the relational model. Two alternatives are described in this section to provide options that may be augmented by many other packages in the future.

### *8.1. Zim: an entity-relationship model package*

A Canadian telephone subsidiary of the Bell system designed and implemented a version of the Entity-Relationship (ER) model for its own internal use. The system was so successful that it was developed into a commercial product named ZIM, marketed by Zanth Systems, Inc. This is perhaps the only true ER model system running today. It was originally designed for PCs, but it has been ported to minicomputers, and it is beginning to gain a modest level of acceptance in competition with pure relational models.

ZIM allows a more complete representation of real-world information, and it is particularly successful in handling relationship attributes (the strong point of the ER model). As such, it is worth investigating as an alternative to the relational models described above. Students working with ZIM find it easy to use, but there is not much information on its performance in comparison with other packages. More study is definitely needed to evaluate the strengths and weaknesses of this package.

### *8.2. File manager: a hierarchical package*

File Manager was designed originally to serve the administrative needs of the Veterans Administration Hospitals in the US. It was intended to be a public domain product, available on all systems capable of running the MUMPS language (which means all sizes of computers from micros to mainframes), and generalized to serve database needs of a wide variety of applications. It has competed successfully in comparison with other commercial packages, winning some important contracts over much better known vendors.

Fileman, as it is known, allows users to define *either* a hierarchical or a flat

file (quasi-relational) structure. This is an important consideration, since many clinical databases are hierarchical in nature, and this system does permit such databases to be defined in a natural manner. In its public domain versions (there is now a commercially supported version), it has deficiencies in input features and some awkwardness in user interface, but it can be used by novices without previous training, and it can also be refined by expert users to provide efficient database solutions.

While it would be improper to claim that Fileman can offer everything that some of the better support database packages have, it is certainly one alternative that should be evaluated.

## 9. Interface to statistical packages

Statistical analysis of data is an important component of most clinical studies. Physicians need to know the statistical significance of clinical trials, the correlation between epidemiological observations and morbidity, and many other more sophisticated measures of statistical significance.

Most database packages provide very simple forms of statistics, usually including such measures as *average*, *minimum/maximum values*, and sometimes *standard deviation*. Few commercial packages go so far as to provide regression analysis, chi-square or T-test or other common statistical measures, and none offer the more complex forms of correlation required to analyze some types of clinical data.

In order to analyze database information using more comprehensive statistics, it is necessary to transfer the data derived from a database to one of several standard statistical packages. Data transfer is becoming easier as more database package vendors agree on transfer formats. There are now several formats for copying a data subset from a database package into an ASCII file that can be read by other packages. Since statistical packages almost always expect to work on data derived from such sources, they provide for interface to the most common database output formats.

There are a number of statistical packages available on personal computers. Among the better known packages are BMDP (Biomedical Data Package), SPSS (Statistical Package for Social Science), and Systat. Comparison of these packages is beyond the scope of this text, but readers should be aware that a number of such packages exist, and that most of them offer means for accepting data from standard database packages. For example, SPSS has standard interface provisions for accepting data from dBASE, Lotus 123, and Symphony.

For the purposes of this text, it is sufficient to say that the need to transfer data from database packages to statistical packages has been solved in a number of ways by different vendors of each type of package. Clinical researchers need therefore to know what statistical packages offer the best combination of technical features, interface supports, and user friendliness in order to select the right package for their analytical needs.

## **10. Summary**

This introductory chapter has presented some of the key characteristics of database systems, beginning with primitive components and moving through database models to implementation considerations. In the remaining chapters, the principles touched on in this section will be reinforced with multiple examples. As a reader gains familiarity with specific clinical database examples, he or she may benefit from reviewing portions of this chapter in order to gain new insights to the concepts that might have been overlooked on first reading.

# Observational databases: a clinical perspective

SUZANNE B. KNOEBEL

*From the Krannert Institute of Cardiology, Department of Medicine, Indiana University School of Medicine, and the Roudebush Veterans Administration Medical Center, Indianapolis, Indiana\**

## Abstract

Observational databases by providing data on patient outcomes from diverse therapies and by severity of illness offer the potential for: 'living textbooks', quality care assessment, risk-stratification and subgrouping, clinical predictions, and health care policy modeling. Such databases are expanding exponentially primarily in response to the needs of cost-containment and quality assurance forcing increasingly sophisticated data collection and analytic techniques.

The limitations of observational databases reside primarily in the quality, consistency and comparability of data collected from multiple sources; the multifactorial nature of disease processes in individual patients and patient subgroups making statistically valid comparisons difficult; a perceived potentially negative impact on clinical practice; and, the problems that health care professionals encounter in conscientiously collecting the data. There are solutions to all of these limitations. The potential rewards from observational databases justify the efforts that will be required to make them feasible, reliable, and acceptable to the profession and health care planners alike.

Clinical and policy decisions need to be based on conclusions drawn from real life data. Observational databases can be designed to collect and analyze such data to give results comparable with and, in some cases, of greater utility and relevance than those derived from classical 'scientific' research projects as well as the randomized, controlled clinical trial. This paper provides an overview of the observational database, its beginnings, its problems, some solutions to the problems and projected utilities.

To know the natural progress of diseases is to know more than half of medicine... Knowing the physiognomy of the disease when allowed to run its own course, you can, without risk of error, estimate the value of the different medications which have been employed. You will discover which remedies

\* Supported in part by the Herman C. Krannert Fund; by Grants HL 06308 and HL 07182 from the National Heart, Lung and Blood Institute of the National Institutes of Health, Bethesda, Maryland; and the American Heart Association, Indiana Affiliate, Inc.

have done no harm and which have notably curtailed the duration of the disease; and thus for the future you will have a standard by which to measure the value of the medicines which you can see employed to counteract the malady in question.

Trousseau, 1801-1867 [1]

Medicine is the art of understanding diseases and of curing or relieving them when possible. Under this acceptation our sciences would, at least, be exonerated from reproach and would stand on a basis capable of supporting a reasonable and durable system for the amelioration of human maladies.

Dr. Jacob Bigelow, 1852 [2]

Doctors Trousseau and Bigelow undoubtedly would have agreed that if there were a system to keep complete and accurate records of the clinical experiences of a large number of physicians and the patient outcomes that resulted from the diverse therapies they employed and, furthermore, if those records could be analyzed meaningfully so as to demonstrate the benefit of specific interventions, the effort involved in creating and maintaining such a system would be well rewarded. In the early 1970s Rosati *et al.* [3], Wallace and Rosati [4] gave a contemporary plea for data collection. They gave powerful and creative arguments for the role that computers could play in the clinical practice of cardiovascular medicine and the advancement of cardiovascular knowledge. These clinician scientists pointed out that physicians function through pattern recognition. Physicians recall patients with clinical profiles similar to a particular patient under consideration, how those patients responded to specific therapies, and they compare both current and past patients with knowledge of the natural history of disease with and without therapy so as to determine if there are factors in individual patients or subgroups that influence outcome. After processing all of that information, choices then are made between alternative management strategies, the goal being to select the strategy that represents the greatest likelihood for an optimal patient outcome. Despite the conceptual appeal of this decision-making approach, the results may be imperfect because of the limitations and vagaries of memory and biases that may be introduced as a result. Often the most recent experiences or particularly good or bad outcomes are best remembered and, thus, may be overweighted in the analytic process. A vast amount of experience contained in the unremembered 'average' patient may not enter into the decision equation.

Wallace and Rosati [4] gave a classic example of a 1970s decision problem: *A white male in his middle 50s has crushing substernal chest pain which radiates to the left arm. The pain is brought on by exertion and is relieved in 23 minutes by rest or nitroglycerin. These pains began 1 year ago and occur with a frequency of 35 times/week. The resting ECG shows 56 PVSs/min and is otherwise normal. Two minutes after entering Stage II of the Bruce test, 3 millimeters of S-T depression developed in leads V4, V5 and V6. Chest*

*x-ray normal. Blood pressure 160/95. Cholesterol 220. Triglycerides 50. Cardiac index 3.3 liters/min<sup>2</sup>. Ejection fraction 60%. Coronary arteriogram shows a 90% occlusion of the left anterior descending, and 80% occlusion of the right coronary artery, and a 'clean' left circumflex. The patient smokes 2030 cigarettes per day and has for 25 years.*

The underlined words and phrases represent potentially significant independent predictive variables. The authors asked, 'Can you remember such a patient?' If you can, how was he treated? What was the result? Other questions might be: What is the probability he will be alive five years from now? Asymptomatic or symptomatic? Would angioplasty or bypass surgery increase the probability of being alive five years from now? Asymptomatic or symptomatic? Do other therapies provide a better prognosis? Would stopping smoking change his risk? If bypass surgery is done, should he have internal mammary or saphenous vein conduits? The questions are almost endless. Physicians make decisions about such patients daily. Where do they get the necessary data? Wallace and Rosati [4] proposed that a computer-based record system could make information readily available as to what happened to patients like this one with varying therapies and risk factor modification over time and, thus, could provide a 'living textbook' of cardiology. Other applications might include: technology evaluation (demonstrating that technology utilization favorably alters patient outcome), quality assurance (observed outcomes are the best that can be expected given a patient's risk profile) and cost-effectiveness analyses (the effects of alternative management strategies on patient outcomes relative to cost).

Some of the benefits of databases that were envisioned by Rosati *et al.* [3], Wallace and Rosati [4] have been achieved through registries such as the Coronary Artery Surgery Study Registry, the NHLBI PTCA Registry, Seattle Health Watch and the Duke Cardiovascular Disease Databank and several large randomized clinical trials [5]. The impact of many of these studies, however, and randomized clinical trials, in particular, has been primarily to establish broad therapeutic principles. In strict interpretation, the results apply only to the patients enrolled in the study and generalization to other groups is questionable to say nothing about the individual patient. Risk-stratification is not provided for in randomized clinical trials; rather, risk-averaging through the use of large numbers of patients so that the efficacy of a particular therapy can be assessed is the goal. Observational databases, on the other hand, can be structured and analyzed to provide the data for risk-stratification, subgroup definitions and patient outcomes according to risk profile [6,7].

Despite the potential advantages to be derived from observational databases, there have been significant barriers to their use that caused clinicians and clinical investigators to seek other alternatives to their data handling problems [8]. Clinicians could not see that computer databases offered sufficient advantages that it was worth changing patterns of practice to incorporate them. Clinical investigators also tended to avoid clinical research necessitating large database analyses. Perceived barriers included cost, inconvenient access, lack



of available expert consultation and computer programming assistance, a linguistic gulf between investigators and computer scientists and a less than receptive attitude of scientific review bodies. The firmly held belief that randomization was absolutely required for valid studies on patient populations relegated observational databases to less than 'scientific' status. Funding for clinical studies based on observational data was almost impossible to secure. If, then, there was little utility of observational databases for clinical decision-making, and research questions could not be properly addressed by observational database methodology, who needed them? Observational database development did not flourish despite their conceptual advantages.

Other factors intrinsic to database science acted as deterrents to the use of computerized database systems. The ultimate database for any clinical question cannot be prestructured or all possible significant qualitative or quantitative variables can not be anticipated at the time a database is established. New questions are constantly being asked and new data are constantly becoming available as new clinical technologies are developed and older ones become more quantitative. Constant revision of a database is the rule rather than the exception. This has been until relatively recently very time consuming, personnel intensive and frustrating to clinicians and computer scientists alike.

### **Developments and current status**

Although observational databases did not become common in clinical practice settings in the 1970s and early 1980s, interest continued in academic centers. Database management systems were created that allowed new questions to be asked and new data to be added without the need to create new data structures [8]. Database linking became possible. The systems also became more user-friendly. Clinicians could create files and query the data without needing to interact with a programmer. Computer literacy increased in clinical personnel. Examples of the utility of observational databases became increasingly frequent.

Advances in computer technology expanded the potential for observational database applications. The computers used for data input, data access and storage became readily affordable, the software required to enter data directly on-line was developed, data transfer capabilities via floppy discs became feasible eliminating the need for expensive data communication links and, 'off-the-shelf' statistical packages for data analyses that run automatically and have considerable sophistication can be readily acquired [6]. While not many private hospitals or other non-university affiliated practice groups had been willing previously to make the investment in biostatisticians, computer scientists and computer hardware necessary to manage clinical databases, it now became feasible for them to do so. Even small hospitals now can monitor their performance in terms of patient outcomes particularly if their data are structured so as to be comparable with other data sources. Multicenter

collaborative databases provide for larger numbers of patients thus increasing the validity of the observations.

An important impetus for increased interest in the development of observational databases comes from the cost-containment initiatives of governments and third party payors and the resultant need to assess whether quality care is being infringed upon by cost-cutting strategies. Governments, reimbursement agencies, physicians, and physician groups need vastly more information about medical costs as related to the long and short term beneficial or detrimental effects of alternative managements on patient outcome as a basis for policy decisions. Information needs to be provided as to what types of care are possible at certain reimbursement rates and what practices might need to be foregone if funds were not available [9]. It needs to be known if the trade-offs between cost and resource utilization make sense for patients and for society. Such information must be data-driven, preferably by appropriately analyzed observational databases. Physicians 'opinions' are no longer acceptable. They are considered to be self-serving.

Randomized clinical trials can not provide all of the data required in the current cost-containment and quality assurance environments. The cost of such trials is enormous, the scope often restricted, their relevance in a rapidly changing technological environment limited, the data often become obsolete before the trial is ended, and their applicability to the 'real world' questionable. Often after clinical trials, questions remain about important patient subgroups that were not included in the study, such as that of elderly women [5, 10]. Observational databases can assist in such issues by addressing key clinical and policy questions in a timely manner and there has been a resurgence of interest in such databases.

### **Principles of observational data collection and analysis**

Some of the strengths of the randomized clinical trial have been the rigorous experimental design, precise definitions for data elements and comprehensive patient follow-up [5, 10]. To fulfill the need for policy as well as clinical decisions, observational databases must match the accuracy of the randomized clinical trial. They must provide for accurate, objective, consistent data collection. In addition they must provide for risk-stratification, risk prediction, and population matching analyses. Many of the principles applicable to observational database structuring have been put forth by Lee and Goldman [11].

The experience of most who have been involved in clinical database development and utilization has been that data collection is facilitated when integrated into routine clinical care. The multicenter chest pain study as commented on by Lee and Goldman [11] provides such an example. Instead of writing a note on the emergency room form or patient chart, the physicians entered the clinical information on a data form. The physicians benefitted by

having a convenient alternative to note writing and the study benefitted from having prospectively collected data. Others have found that if a catheterization laboratory report or physician referral letter is generated from data entered by the physician at the time of catheterization, for example, data collection is more compulsively and accurately done.

Other considerations relative to data collection are that the data must be objectively collected and with clear definitions for the variables and outcomes of interest. The definitions should be readily available to the data entry person either on the data form or computer screen if there is direct data entry. Element definitions and adherence to those definitions is the most difficult part of observational database analyses. When does an acute myocardial infarction become a recent myocardial infarction? When does a recent myocardial infarction become a post myocardial infarction? When does an emergent procedure become a semi-emergent procedure? What criteria indicate congestive heart failure? Data definition is not an easy task.

### **Data analysis**

The observation that patients treated with digitalis after myocardial infarction have a higher mortality rate than those not taking digitalis illustrates a major problem in analyzing observational databases as discussed by Lee and Goldman [11]. It is probable that the deleterious effect of digitalis is related to the fact that the high risk patients are being treated with digitalis. Digitalis becomes, then, a confounding third factor [11]. It is obvious that the population-matching process becomes unwieldy when there are many such potential confounders as is true in multifactorial, multidimensional disease states such as coronary artery disease and its various manifestations.

The classic methodologies to identify factors with independent predictive significance and, thus, those that provide for risk-stratification and risk-prediction are the multivariate and non parametric analytic techniques such as recursive partitioning [11]. Other approaches are derived indices and 'propensity' scores [11, 12]. All of these techniques are quite complex and beyond the scope of this article. They require considerable statistical sophistication to understand their values and limitations.

An approach that we have used, as well as one of similar concept applied by the Duke University group [6, 7], deserves further consideration because of its relative simplicity and potential for comparative and predictive analyses. The underlying concept is that of pattern recognition, the physicians' approach [6, 13]. Pattern analysis of clinical problems starts by presenting the patient as a data-set (pattern) rather than as a simple collection of discrete features. The patient with acute myocardial infarction, for example, is a data-set; the electrocardiogram and serum enzymes are discrete data [14]. Patterns represent an entity that is greater than the sum of the parts and take into account features of a patient or clinical problem that often are interactive and, as such, those that

may be obscured by the weight of more frequently occurring features. Certainly responses to therapy or risks from a procedure or disease depend on critical combinations and interactions of factors present at the time therapy or other interventions are undertaken and to any perturbations in that combination that occur along the way. It would seem that measures used to describe patients and patient populations should make no assumptions that are dependent on fixed interactions of variables or linearity of those interactions.

When patients are presented as patterns (data-sets) within which the interactions of multiple variables are not known, questions as to the characteristics of patients that do or do not respond to any particular therapy can be asked without the bias inherent in methods that operate via preselection of significant variables. For example, while two significant variables, as generated by regression analysis, might explain a certain number of patients, three factors a few more patients, if seventy-five factors are available as derived by pattern analysis of a large patient population, the vast majority of all possible interactions are represented without bias. While similar interactive combinations of variables also might be assessed by regression analysis, the effort would be nearly exhaustive and bias still would not be eliminated as the combinations, by perforce, would be preselected.

Pattern analysis also may provide for population-matching [6]. In order to evaluate the effect of any particular risk factor, therapy or outcome, it must be known whether the patients being evaluated were comparable to those in the knowledge base group. Comparability of patient groups classically has been handled by randomization; or, when patients are not randomized, statistical methods have been used to correct for imbalances in known prognostic factors. While patient-matching at the time of entrance into a study can be assured by these methods, matching at the onset of an intervention or therapy does not assure that the matching persists throughout the study. If the presumption is that the two groups would have had identical outcomes if given the same treatment, the risk factors must be stable and not influenced by anything other than the therapy itself. Often this may not be a valid assumption as the therapy may induce or unmask an interaction in the data-set that changes risk and, thus, unmasks the populations during the course of the study. For example, suppose that a drug used to treat angina or ventricular tachycardia depresses contractile function, and that depressed contractile function is a risk variable for ventricular tachycardia [14]. The group receiving the drug will be at increased risk over and above that existing at the time of entrance to the study. The group receiving an alternative therapy might have a more favorable outcome not because the therapy is necessarily more effective per se but because it did or did not cause a change in contractility and, therefore, the risk profile. Other well known clinical problems could be similarly used to illustrate the need for continuous risk-matching, a capability not possible using standard analytic techniques.

Perhaps the most compelling argument for pattern recognition is that it most nearly replicates physician behavior.

## Examples of observational database applications

A few examples from the literature on observational database applications highlight their uses and potential abuses and serve to reemphasize the material presented above.

In a letter to the editor of the *New England Journal of Medicine*, Scher [15] presented a comparison of variables in 1500 consecutive abdominal operations by twenty-eight surgeons and one other surgeon with results that were worse than those of the other twenty-eight. Careful analysis disclosed that the patients of the 'worst' surgeon were older, poorer anesthetic risks and more likely to have undergone emergency procedures. Properly adjusted for risk, the results were acceptable. Scher [15] appropriately emphasized that initial comparisons of institutions or individual practitioners are misleading unless the data are adjusted for severity-of-illness.

Califf *et al.* [16], using data from randomized clinical trials as compared with data from the Duke databank, demonstrated the risk-profiling benefits to be had from analyses of observational databases. The study was of two patients with triple vessel coronary artery disease and ejection fractions of greater than fifty percent. Based on additional factors, such as frequency of angina, peripheral vascular disease, location of stenoses, electrocardiographic findings, the predicted survival rates differed by sixteen percent with medical therapy and after five years by fifteen percent for surgical therapy. Both patients, however, would have been put into the same risk category using data from randomized controlled trials.

Knoebel and Lovelace [6] described a method that allows for risk-stratification, risk prediction and population matching from an observational database. The outcomes of patients undergoing angioplasty at two different hospitals were coded in order to ascertain the variables influencing risk and then to compare the performance of the two hospitals by the risk profile of the patients undergoing therapy. Significant risk variables by descending order of importance included: the hospital where the procedure was performed, whether or not it was a repeat angioplasty, the angiographer (doing the angioplasty), the reason for the angioplasty (acute or chronic), preceding thrombolytic therapy, angina status and concomitant disease. There was a difference in the two hospitals that was not accounted for by the risk variables of the patients undergoing angioplasty. That difference is presently being assessed. Early evidence suggests that cigarette smoking, a variable not included in the first analysis, may be highly predictive of risk and the prevalence of that variable was significantly greater in the hospital with less than optimal results.

There are many other recent studies in the literature and the value of data emerging from observational databases is impressive. Applications of data derived from observational databases include decision modeling, cost effectiveness analyses and policy formation [17, 18, 19, 20, 21].

## Conclusion

Computer-based records interfaced with powerful statistical techniques have contributed importantly to advances in clinical cardiology through the definition of subgroups of patients who will benefit from therapy, those who will not, the risk factors influencing outcomes, and the factors that influence the appropriate timing of interventions. The present climate of cost-containment and government involvement in health care policy throughout the western world expands the need for such information. Data relative to the effectiveness of alternative strategies of health care must be available to provide a basis for responsible policy making [21]. The impact of changes in technology utilization and practice patterns on the outcomes of care must be quantitatively known so that resource allocation decisions can be made in the best interests of the patient and in accordance with national priorities.

While the randomized clinical trial has been the traditional method for delineating the effectiveness of diverse therapies on patient outcomes, it is not practical to use this methodology to answer all questions of clinical and policy interest. Clinical observational databases are likely to become increasingly important as a means for providing data as to the effectiveness of varying technologies and management strategies.

A number of developments have made utilization of observational databases feasible: dramatic increases in microcomputer capabilities and decreasing costs of microcomputer systems, increased computer literacy, database linking capabilities and the interest that has been generated by increasing medical costs and increased competition between hospitals and physicians. The latter serves to provide a clearly defined benefit to those who have computer databanks that can be used to generate mortality statistics and other quality care data that are useful in marketing.

A number of problems need to be overcome. Among these are: the need for common definitions of data elements; newer biostatistical methodologies; continuing financial and professional support. Physicians will need to play a prominent role as supporters of observational databases. Such support has not been enthusiastic in the past, probably because physicians could not see that the availability of such information provided them a major advantage. Currently this is changing and must change for the future. Data are required to demonstrate the effectiveness of technology utilization relative to patient outcomes, and gross mortality figures need to be viewed in light of the risk profiles of the patients being served. It is vital that the past patient experience data, the knowledge base data, must be an unbiased representation of the problem and that the data must be in consistent and interchangeable forms. It is suggested that analyses of the data might be appropriately done by pattern recognition techniques particularly in view of the fact that the competent physician undoubtedly analyzes past experiences relative to decision-making via pattern comparisons. Some believe that the judgement of experienced clinicians still represents the best approach to clinical decision-making.

Certainly it has not been duplicated by any currently available information processing systems. For physicians to be listened to, however, data and models based on data are required. The observational database will provide the required numbers, probabilities and temporal correlations.

## References

1. Coope R (1952) *The Quiet Art: A Doctor's Anthology*. Edinburgh, London. ES Livingston, Ltd. 17, 28
2. Coope R (1952) *The Quiet Art: A Doctor's Anthology*. Edinburgh, London. ES Livingston, Ltd. 28
3. Rosati RA, Wallace AG, Stead EA (1973) The way of the future. *Arch Intern Med* 131: 285-287
4. Wallace AG, Rosati RA (1973) Computers can change cardiology. *Circulation* 57: 439
5. Hlatky MA, Califf RM, Harrell FE, Lee KL, Mark DB, Muhlbaier LH, Pryor DB (1989) Clinical judgement and therapeutic decision making. *J Am Col Cardiol* (in press)
6. Knoebel SB, Lovelace DE (1989) General use observational databases: A method. *J Appl Cardiol* 4: 59-67
7. Harrell FE, Lee KL, Califf RM, Pryor DB, Rosati RA (1984) Regression modeling strategies for improved prognostic prediction. *Stat Med* 3: 143-152
8. Knoebel SB, Lovelace DE, Heger JJ, Zipes DP, Fisch CF (1981) A computerized cardiovascular database system: Application in arrhythmia research. *Comput Biomed Res* 14: 482-492
9. Knoebel SB (1989) Modeling for policy decisions: Potential and problems. *J Am Col Cardiol* 14(3): 7A-11A
10. Hlatky MA, Califf RM, Harrell FE JR (1988) Comparison of predictions based on observational data with results of randomized controlled clinical trials of coronary artery bypass surgery. *J AM Col Cardiol* 11: 237-245
11. Lee TH, Goldman L (1989) Development and analyses of observational databases. *J Am Col Cardiol* 44A
12. Stoupel E, (1976) *Forecasting In Cardiology*. Jerusalem, Israel. Israel Universities Press
13. Knoebel SB (1986) *Perspectives on Clinical Decision-Making*. Mt Kisco, New York. Futura Publishing Co Inc
14. Knoebel SB (1988) Perspectives on clinical decision-making. The application of pattern analysis. *PACE* 11: 2077-2081
15. Scher KS (1987) Importance of severity of illness in outcome assessment. *N Eng J Med* 317: 171
16. Califf RM, Pryor DB, Greenfield JC Jr (1986) Beyond randomized clinical trials: Applying clinical experience in the treatment of patients with coronary artery disease. *Circulation* 74: 1191-1194
17. Moskovitz AJ, Pauker SG (1989) A patient with new Q waves: Methods for decision making in the individual patient. *J Am Col Cardiol* 14: 29A
18. Georgeson S, Meyer KB, Pauker SG (1990) Decision analysis in clinical cardiology: When is coronary angiography required in aortic stenosis? *J Am Col Cardiol* (in press)
19. Dittus RS, Roberts SD, Wilson JR (1989) Quantifying uncertainty in medical decisions. *J Am Col Cardiol* 14: 23A
20. Weinstein MC (1989) Methodologic issues in policy modeling for cardiovascular disease. *J Am Col Cardiol* 14: 38A
21. Young DA (1989) Payment policy, quality of care and decision making with inadequate information. *J Am Col Cardiol* 14: 3A

## PART TWO

### Tools and services



# Advanced instruments and methods for the development of databases applied to cardiology

FRANCESCO PINCIROLI

*Dipartimento di Bioingegneria, Politecnico di Milano Centro di Teoria dei Sistemi, CNR, Italy*

CARLO COMBI

*Fondazione Sestini and Divisione di Cardiologia of the Ospedali Riuniti, Bergamo, Italy*

## Abstract

The purpose of this work is to point out what we considered a possible advance from the point of view of instruments and available methods for the creation of database applied to cardiology according to the technological panorama.

Microprocessors used at present allow, also on Personal Computer, signal and image processing. The quantity of central memory which these computers can manage reaches 16 megabytes. Devices of inside permanent storage at one's disposal nowadays offer to the user more than 100 megabytes. With storage capacities of this kind it is possible to use the Personal Computer also for medical record management.

In the Personal Computer field parallel processing is now beginning. The transputer is composed by a microprocessor, one memory (2-4 kilobytes) and a device set of communication among processors which allows data transfer high speed (20 megabits per second). Many problems in the medical field can be taken to pieces in parallel processes, and executed by a transputer network.

Video-discs and CD-ROM put at the user's disposal such a storage capacity to allow the storage of a great quantity of bio-images. For instance, two sides of a video-disk can contain 109000 images or an hour of film shooting (30 images per second for 3600 seconds) [19]. Storage capacity of a video-disk, expressed in the usual measure, reaches 4 gigabytes [20]. Storage capacity of a CD-ROM is 600 megabytes.

In order to represent biomedical images it is necessary to dispose of high resolution screens. Today videos are compatible with the majority of graphic interfaces adopted on Personal Computers and offer a resolution of about 280-300 microns and a suitable grey scale.

MEDIX is the name under which IEEE-EMBS proposes a standard for the exchange of medical data. The main purpose of MEDIX is to specify and establish a strong and flexible standard for data exchange among health information heterogeneous systems.

Hypermedia is a general concept which refers to processor capacity to manage, in an integrated way, great quantities of alphanumeric texts, but also

database, suitable sounds and static and dynamic images, by means of videodiscs or CD-ROM. The most known and widespread implementation of Hypermedia is Hypercard. Application packages Hypercard allows a fast access to great capacity mass storages of hundreds of Megabytes, such as hard-disks and CD-ROM, through personal computer. In this system, man-machine interface is particularly simple and comprehensible.

The work consists of three distinct parts which follow one other. The first one illustrates problems believed still unsolved by operators in the area of database for cardiology, as users or makers. In the second part we make comments on the recent methods and advanced instruments which can be useful to sort out those problems. The third one is devoted to pointing out some experiences, that could be significant also for the field of database applied to cardiology.

## **1. Introduction**

It is common knowledge that data concerning a patient in cardiology are numerous and heterogeneous. On the one hand, in fact, graphic information is collected such as echocardiographic images and coronary graphs. There is a need to control and analyze signals, typically electrocardiograms. Also tests which give numerical results are in great numbers: just think of all laboratory tests. There is a great deal of information, often called semiquantitatives and qualitative i.e. in anamnesys, diagnosis and therapeutic prescriptions. The aims to pursue in gathering and analyzing this complex set of data are varied, as well as the problems present in this field: how is it possible to organize this very heterogeneous set of information so as to get the utmost benefit for the patient's health? How to use all this information in the diagnosis formulation process, in the following therapeutic prescription and in their necessary verification?

The computer is the typical available technological instrument for data management. It is already employed in many specific sectors inside cardiology. Mostly, its applications are developed in special fields, they lack coordination and do not form an organic body. For instance, echocardiographic image processing is made on dedicated systems which have few possibilities to be connected to hospital file management systems. The development of database in cardiology presents aspects of a very deep need which is still awaiting suitable and satisfactory solutions. The international scientific scene presents some experiences which are still actively working on database for cardiology. At a European level it is worth mentioning the CADANS-ICIN project. The CADANS project (Cardiology Data Network Structure) developed in Holland at ICIN (interuniversity Cardiology Institute Netherlands) is aimed to develop a database and a geographic network for cardiology and was created to be suitable for precise necessities [1] which can be summarized as follows:

- a) The need to increase total efficiency of cardiology departments by means of collaboration works among different centres with the help of advanced resources;

- b) the fact that clinical cardiology can thrive only as long as there is an intensive interaction with research;
- c) the fact that clinical research is strongly based upon data processing;
- d) the belief that a permanent communication structure will lead to research projects which can provide results otherwise unrealizable;
- e) the experience that the employment of information technology is a natural encouragement to standardize. This is of great significance for action and knowledge exchange.

The CADANS project aims to offer communication ease in order to process data in a network suitable for research, education and clinical practice fields applied to cardiovascular pathologies. Together with the development of this cadre-project, which initially takes into consideration communication of 9 university centres in Holland, CADANS intends to deepen some specific and particularly interesting research fields: the development of a cardiologic workstation, the effects of thrombolytic agents on prognosis of myocardial infarction, the development of a relational database of all patients submitted to cardiac catheterization, and the automatic and integrated analysis of electrocardiograms.

The purposes of CADANS can be taken as examples of widely present expectations in the cardiological field. Among these we note: the estimation of new techniques and methods to be applied to clinical practice, the uniformity and standardization of medical data to support research and clinical practice by means of an exchange and comparison of data having a precise and unambiguous meaning, further computer applications in medical practice and especially in the development of instruments aiding decision-making, the combination of different scientific disciplines such as Information Science, Epidemiology and Cardiology in order to increase research efficacy and efficiency. As far as international cooperation projects are concerned, we would like to mention the effort in the sphere of EEC devoted to define initiatives inside cardiology, able to get the benefits of cooperation among research groups of different countries. A preliminary step of this effort, 'Database for Cardiology' has been carried out in a study visit related to Erasmus Project (contract – SVS-87-0005-1) [2]. The purpose of the visit was to define the most urgent problems to be solved with reasonable intervention of informatics.

As far as international congresses are concerned, two annual events, traditionally important on the international scientific scene for Medical Informatics and Bioengineering have recently been held: these are the SCAMC Symposium on Computer Application in Medical Care and the IEEE-EMBS International Conference of Engineering in Medicine and Biology Society. Each conference gathered more than 2000 people and represents a remarkable potential source of information on database for cardiology. Other similar events which can convey very specific information suitable for our field are: the IEEE Computers in Cardiology Annual Conference which takes place alternately in the United States and in Europe but which last year did not have

any session devoted to database for cardiology, the Annual Congress of European Cardiology Association which this year has devoted specific space to computer applications in cardiology, the census on database for cardiology existing in Italy put in order two years ago which will have a new edition having an international dimension. The purpose of this article is to point out in a logical way, what we consider a possible advance from the point of view of instruments and available methods for the construction of database applied to cardiology according to the panorama of cited events.

## **2. Problems**

There are still many areas of medicine in which technological means are not considered completely appropriate ways of solving real problems. Needless to say that people working in this field wish to make solutions as soon as possible.

### *2.1. Standardization*

*Technological aspects.* The need for standardization in informatics arises in various situations: just think of a user who wants to make his applications operate on different computers or of the necessity to make computers of different trade-marks hold a dialogue.

In the medical field, those who use a computer, feel the need of a system which, on one hand, makes it possible to eliminate an excessive collection of information and, on the other, allows an efficient connection among centres located in different places. Hospitals generally own computers of different trade-marks with different application packages: despite these differences, the need to connect systems to the network is more and more widespread. This necessity urges standards which are respected by manufacturers, this way users could use application packages they have on any kind of hardware. From a strictly theoretical point of view, fixed standards should work at a level where it is not necessary to know the computer we are occasionally working on. The operating system MS-DOS in the personal computer set is de facto a standard, that is to say a reference for almost all manufacturers. But in this case, we deal with a standard which we might call imposed by force by IBM which obliged almost all other manufacturers to adapt to their internal standards. However, a manufacturer is not always able to impose standards on the outside. Let us think, for instance, of the operating systems for mini-computer or mainframe; every manufacturer adopts a different operating system: VMS or UNIX on Digital-VAX, RTE on HP systems, EXEC-8 on UNIVAC mainframe and so on. Users of these systems feel the need to communicate, by means of a computer, with other users having the same needs and, furthermore, do not want to be bound only to that particular manufacturing company as to the hardware: the standard need (for instance on communication among computers, on the structure of stored data, on controls of peripherals) to which

manufacturers should adapt comes from the users who would make applications machine-independent.

The necessity of precise and suitable standards in this field is felt also in medical applications: connections among different centres would be easier, as well as the integration of informatic resources for instance, inside the same hospital by means of local networks [3, 4].

*Clinical aspects.* A hypothetical presence of standards concerning any possible communication protocol among computers would not solve in every case the problem of application transportability or of communication among different hospitals: the need of another kind of standardization strictly bound to medical collected information would still be present.

Data concerning a patient are stored on paper for hospital files. This information is put together in different ways according to the kind of test the patient undergoes, according to the hospital, to the department and so on. For instance, in communication among different hospitals, the different organization of information has to be carefully taken into consideration: the more the information organization is different, the more difficult interhospital communication will be. Collected information, furthermore, can be very different from a quality point of view and is related to the physician who makes out the various parts which comprise the hospital file. Clinical information concerning a patient is almost never simply put in a computer and stored and managed: let us think in fact of the number of synonyms to define a pathology, of how it is articulated, to the point of becoming subjective, the use of adjectives in order to define the seriousness of a sickness. In this field, there is a very strong need of standards able to allow a true communication among different wards of the same hospital or among different centres [5, 6]. The need to have standards in communication protocols among computers as well as in defining and organizing transferred clinical information can be considered two aspects of one requirement: using an analogy with the telephone communication field, two computers placed in two different hospitals must be able to communicate (communication standard) but should also understand each other (Medical data interchange standard).

## *2.2. Capacity of storage supports*

The information entity in the medical field is great indeed: just think of, besides all papers, the quantity of images and graphs collected during a patient's hospitalization. The necessity of storage devices with more and more capacity is felt in many application fields. A bioimage requires a lot of memory. For instance, a radiographic image obtained from a film: usually this is digitized with a resolution of about 200 microns. For a 14 × 17 in. film each image is composed of 1680 × 2048 pixels; for each pixel 8-12 bits are used to code grey levels. Storage of an image, therefore, requires 28-41 megabits [7, 8].

Capacity of present storage supports is often regarded as a limiting factor

versus desired applications. Devices like Laser-Card could allow the patient to carry the most important and useful medical information in case of an emergency: identity data, blood-group, allergies to certain drugs, a minimum clinical history of the patient, possible graphs or important images in case of urgent assistance (radiographies, ECG, and others) [9]. In this case, the desire is to have at one's disposal the utmost possible quantity of memory so as to record on Laser Cards all the information of clinical interest concerning the patient.

In dynamic electrocardiography it would be desirable to replace magnetic tape on which a recording of an electrocardiogram during 24 hours is performed with chips as storage devices. At present, memories manufactured with C-MOS technology having a capacity of 4 megabits are available: this result sorts out the problems of storing a Holter recording of 24 hours only after the application of data-compression technics [10]. Without compression about 90 megabits should be stored, coming from 2 traces having a 24 hours duration, sampled at 100 Hz with 10-12 bits for coding each sample: this is not possible with a limited number of solid state memories. After having applied suitable compression technics it is possible to store the whole trace on 4-5 chips which can be easily contained in a device having the same dimensions as present Holter recorders.

### *2.3. Transmission speed of channel*

The need expressed in many medical fields of transmitting and processing through the computer signals and images of clinical interest crashes, besides above mentioned communication software problems, also with technological limits. Let us take into consideration, for instance, a geographical network carried out by means of a commuted telephonic network: transmission speed of this kind of network is usually of 6400 bits/second. Let us imagine we want to transmit the radiographic image we mentioned in the previous paragraph: the image is sampled by  $1680 \times 2048$  pixels with 12 bits for each pixel. More than 41 megabits will have to be transmitted: in fact, we have to point out that control bits also have to be transmitted since they guarantee a good transmission, besides bits of image coding. On the above mentioned network it will take more than one hour and 45 minutes to transmit one image: in many applications this waiting time can be unacceptable. Similar problems arise when, instead of images, we want to transmit biological signals such as ECG, EEG, EMG. Values, in this case, come to about 5-6000 bits for each second of monotraced signal (i.e. recording an electrocardiographic trace where usually, sampling is carried out by a frequency of 500 Hz and a resolution of 10-12 bits each sample).

### *2.4. Software portability and hardware obsolescence*

The software available to the physician is often developed and manufactured ad hoc for a special machine: this is a very rigid limitation. There are two factors

which restrict the possibility to make suitable application packages for various kinds of machines: adopted Hardware and Operating System. The Operating System, that is to say the basic software on which different applications rely, is the key element for transportability: in fact this software separates application packages from a certain Hardware structure making therefore, programs transferable on different machines which adopt the same Operating System. Let us note, in the Personal Computer field, the example of MS-DOS Operating System: the adoption of this operating system by the majority of manufacturing companies (IBM, Compaq, Toshiba, Olivetti, to mention only a few) allowed a great diffusion of different kinds of application packages. Yet, in the Personal Computer field too, transportability of some application packages, although possible in theory, might often be difficult: many application programs instead of being designed to rely on operating system, are intentionally made for Hardware structure of a special category of machine. All this is worsened by the fact that obsolescence times of hardware are shorter and shorter. Let us take into consideration, for instance, the very fast evolution of Personal Computers which adopt MS-DOS operating systems: in a three year period we moved from microprocessor 8086 to 80386 with clock frequencies of 4.77 MHz and 20-25 respectively through various intermediate microprocessors (80C86, 80186, 80286). Central memory from a maximum of 640 Kbytes reached already usual 2-4 megabytes, capacity of bulk memory from 20 megabytes of some years ago reached more than 100 megabytes available even on portable machines.

In the Mini Computer field adopted operating systems are in great numbers: RTE on HP systems, UNIX or VMS on Digital VAX, RSX on Digital PDP-11 and so on. In this situation, transportability of application packages is allowed only inside the machine family of the same trade-mark which uses the same operating system. Software transportability would permit a more frequent technological up-date.

The need to preserve all information collected during years of use of the same system together with the possibility of adopting the new progress in the technological field leads to a more and more precise division between Hardware and Software: application packages should be applied on different machines and should not bind the user to a particular kind of machine.

### *2.5. Graphic resolution of supports addressed to transient and permanent displays*

The physician, in everyday practice, has to analyze biological signals and images such as, for instance, electrocardiograms, electroencephalograms, radiographies, echographies and others. Devices that provide for display of this kind of information have to meet suitable requirements: let us think of the resolution with which a radiographic image, as the above mentioned one, is acquired. We know that the human eye can distinguish up to 4-5 dots per millimeter: in order to see properly the image, therefore, it will be necessary to

dispose of display supports with resolution of about 200 microns. Besides this, we need to take into consideration the fact that images should be represented by a color or grey scale of a certain extent, typically by scale of 256 or more gradations.

The use of computers in processing biological signals and images requires the utilization of transient (video) and permanent (plotters, printers and so on) display supports which should be much more sophisticated than needed in order to represent alphanumeric information. Reaching the necessary fairness [11] has provoked the need to supply also these individual work stations with high resolution video and printers.

## *2.6. General and specific staff training*

In medicine, as in many other professional sectors, the obsolescence of knowledge is faster and faster. Consider, for instance, the various progresses that a physician witnesses during his professional life: technological evolution puts at his disposal more and more sophisticated and powerful aids which require a learning phase not only of the people in charge of their use, but also of the people who will handle newly available information. Research in the medical field which gives more and more precise answers to causes and physiological events connected to particular pathologies, proposes new therapies, new kinds of intervention, new drugs. The whole body of therapies is often characterized by new discoveries. In this situation of rapid evolution, the continuing education problem and learning time are of great importance. To what extent is this time compatible with the physician's usual professional activity? Which devices can help in permanent instruction? In order to face efficaciously all these problems we have to involve various sectors: among the most important, medicine and engineering applied to the study of knowledge structure and development of new instruments of learning. Learning time required to master traditional programming languages such as C, FORTRAN and others is very long: you can think of two-three full-time months before being able to write programs of some importance. A physician who carries on a daily consulting or department activity cannot devote all this time to learn. Also we cannot think of pushing beyond a certain limit the learning time: the complexity of these languages requires a constant practice for remarkable periods of time.

At present, there are application packages which require, in order to be used properly, very reduced learning times, typically about ten-twenty hours. Think, for instance, of Personal Productivity Tools such as LOTUS 123 or Symphony which allow easily archive management for personal use and statistical applications, or real database management systems available on PC such as dBase (I, II, III, III plus and IV) or PARADOX. The so-called Author systems which allow, for instance, consecutive screening to use for a class or a demonstration represent other kinds of languages which require a reduced training time.



Above mentioned software instruments can be directly managed by the physician after a short preliminary training. By virtue of these new languages there is a new possibility for the physician to personally manage available information resources without being forced to attend a long and tiresome learning phase. Availability of these new means gives rise to new needs and new investigation fields: which are the best periods and ways for suitable education of a physician in the sectors of informatics he is really interested in?

CAI Systems (Computer Aided Instruction) have been very successful in medicine as a valid help and complement to traditional teaching methods. Learning with the computer's help has many advantages: for instance the requirement, often strong, to tailor as much as possible to learning rhythms of a single student is met. By the implementation of programs which allow the student to verify his level reached during various learning phases [12] or even which schedule different study courses according to the student level, we can really exploit the extra resource that the computer puts at one's disposal. Thanks to the computer it is possible to put at the student's disposal a complete collection of clinical cases of the different sectors of medicine on which he can practice in an interactive way, which was not possible with the aid of previous paper files, seeing the accomplishment of theoretical notions previously learnt. The need of implemented learning systems on computers draws attention to problems already present in other fields of medical informatics: the use of the computer is effective if the available memory to record all interesting clinical cases from a didactic point of view has a great capacity. In CAI systems it is important to store and properly represent signals and images. Besides storing devices having capacity, it is therefore necessary to have peripherals able to reproduce precisely signals and images. Learning, furthermore, is again connected to the standardization of medical knowledge, to the definition of a common vocabulary of pathologies, to the specification of seriousness and danger degree of a pathology.

### **3. Emerging new instruments and their capacity as contributions to the solution of still unsolved problems**

#### *3.1. Hardware*

Technological instruments used to make applications of medical interest are now numerous. These instruments can be divided into two groups: on one hand computers, on the other the whole group of peripherals which allow more and more frequent use of computers themselves.

##### *3.1.1. Computers*

*Personal computers.* Among the most important features to be taken into account in estimating performances of the Personal Computer, there are:

processing power of used microprocessors, capacity of inside mass storage, and communication speed among various mentioned devices. The available microprocessors also allow on this kind of machine, applications which require a high number of operations. An example is given by the signal processing: computational power of these machines allows these applications very easily. Intel 80386 microprocessor is the youngest in the family of Intel microprocessors which includes also the previous 8088, 8086, 80286 and is used on many different computers: for example, the Toshiba T5100 or the Compaq 386 family.

The quantity of central memory which these processors can manage reaches 16 Megabytes: this feature, together with the power of microprocessors, is important, for instance, to make operations on great quantities of data amounting to tens of Megabytes which are called upon in image processing.

Devices of inside permanent storage at one's disposal nowadays offer to the user a bigger and bigger storage capacity: in the Personal Computer field, models with inside Hard-disk of more than 100 megabytes are marketed (see for instance some models of Compaq: 386/20 and 386/25). With storage capacities of this kind it is possible to use the Personal Computer also for applications which require a storage of great quantity of data such as archives management of data concerning patients of a certain department [13] or biological signal processing.

*Minicomputers.* Computers included in the category of Mini are the most widespread in hospitals: from administration offices to intensive therapy units for continuous monitoring of hospitalized patients. The attention of manufacturing companies is addressed to the study of systems able to communicate also with different computers such as, for instance, the Personal Computers. Digital, with VAX [14] guarantees the possibility to connect its mini to computers of the most important trade marks such as HP, IBM, Apple. Technical features of the newest minis are really remarkable: system memory up to thousands of megabytes, mass storage of tens of gigabytes, up to 128 connected terminals represent the most striking data. In the environment of minicomputers, considerable powerful single-user work stations are of great interest: especially in the medical field these machines are used for applications devoted to image processing. The model VAX station 3500 [14], for instance, has a relative performance of the processor (compared to VAX-11/780 which is a reference in the field of minicomputers as far as performances of processor are concerned) equal to 4.2, 32 megabytes of central memory up to 560 megabytes of bulk memory, a screen with a resolution of  $1024 \times 864$  Pixels and a scale of 256 colors, graphic co-processor, UNIX operating system: all this makes this category of machines suitable for applications which require the utmost speed possible, a great number of operations on great quantities of data and high representation quality of obtained processing.

Sometimes, as to Personal and Mini Computers, processor performances are defined as Mips (Million of instruction per second): instructions to which we refer are those belonging to the set of elementary instructions of processor.

*Transputers and neural networks.* New manufacturing philosophies are gaining ground in computer design: neural networks are so-called because they are similar to our nervous system. Our nervous system allows very short processing times, especially if compared to the relatively moderate conduction speed of our neurons, because there are parallel processing processes [15]. Why should we not apply this philosophy to the computer's design as well? On mini and mainframe parallel processing, obtained by virtue of several processors, this is already carried out: in the Personal Computer field parallel processing is now beginning. The word transputer means the hardware core which allows parallel processing of several processes [16]. This is composed by a microprocessor, similar to the 80386 Intel or to the 68030 Motorola, one RAM (24 kilobytes) and a device set of communication among processors which allows data transfer high speed (20 megabits per second). By the connection of several transputers we obtain a multi-processor system in a Personal Computer environment. Manufacturing companies (Computer System Architects, Levco and others) which have put on sale hardware for parallel computers also supply the necessary software in order to develop concurrent programs. Many problems in the medical field can be taken to pieces in parallel processes: just think, for instance, of the acquirements of ECG signals and of the various kinds of processing on the signal itself which can be carried out while acquisition is still under way or again, of a decision aid system which acts simultaneously on different subproblems which compose the problem we are dealing with.

### 3.1.2. *Peripherals*

Peripheral differentiation at the physician's disposal to storage and/or transmit information to the computer is nowadays remarkable.

Video-disc and CD-ROM which exploit technology setup for compact discs already widespread in the Hi-Fi field, put at the user's disposal such a memory capacity to allow the storage of a great quantity of bio-images [17] [18]. For instance, two sides of a video-disc can contain 109000 images or an hour of film shooting (30 images per second for 3600 seconds) [19]. Memory capacity of a video-disc, expressed in the usual measure, reaches 4 gigabytes [20].

Memory capacity of a CD-ROM is 600 megabytes. These instruments, where reading is performed by means of a laser, still do not cancel and recall images at will: image storage phase is initially carried out by the user's control.

Possibilities offered by video-discs and CD-ROM have been exploited mainly in the CAI (Computer Aided Instruction) system field: thanks to this technology we are able to organize image archives in order to permit the student to know all collected surveys, for instance, on a certain pathology [17, 18].

Another sector where modern technologies have offered new application fields is represented by the collection and the storage of data concerning patients: hospital files and health cards are two examples of a group of patient's data on which are studied new ways of collection and storage. Laser card is a support for storage and up-date of the patient's clinical curriculum and is

becoming more and more appealing. This card has the same dimension and structure as a credit card and can contain about 2 megabytes. It is obvious that with this storage capacity it is possible to store not only suitably structured alphanumeric data, but also, for instance, an identification picture or electrocardiographic traces up to radiographic images suitably compressed [19].

In order to represent biomedical images, as we already mentioned, it is necessary to dispose of high resolution screens. Today we can find on the market videos which are compatible with the majority of graphic interfaces adopted on Personal Computers which offer a resolution of about 280-300 microns and a suitable grey scale (or color scale): in comparison with above mentioned dimensions concerning representation of biomedical images, we can affirm that these screens permit a satisfactory display of acquired images. Above mentioned videos (among manufacturing companies we cite Princeton, Mitsubishi, Nec and others) have a price that can be put together to the Personal Computer without making the video more expensive than the computer: from 500 to 1500\$ compared to 2000-4000\$ for a Personal Computer.

In order to reach this purpose we have to take into consideration the fact that the majority of information that the physician has to handle daily, images, graphs or texts, are written on paper: it is therefore obvious that the interest brought about by the device which allows the computer to read and manage information written on a piece of paper is so great. The scanner permits to read by means of a computer, any kind of paper report, usually reports are typed sheets, but with medium quality scanners, having a resolution of about 300 dots per inch and a scale of about 34 shades of grey, we are also able to acquire traces and biomedical images. This way we can deal, in a unified way, with information of a various nature coming from different clinical examinations: what we acquire through a scanner can be later perfectly stored on laser cards [9].

### 3.2. *Software*

Software proposed nowadays covers a very wide range of needs and skills. In order to be clearer, we can divide the software into two groups: general development software, that is to say not directly oriented towards medicine which includes operating systems, programming languages, database management systems (DBMS), personal productivity tools (PPT), and software especially made for medical applications such as Medical decision making packages: management packages of hospital files, packages for continuous monitoring during intensive therapy and so on.

*Development software instruments.* In the personal computer environment new performances allowed by hardware have suggested new developments of used operating systems: on more powerful computers with 80386 microprocessor, RAM of 4 and more megabytes and hard-disks having more than 100 megabytes, some versions of operating systems created on machines of superior categories (mini and mainframe) such as Unix, Xenix and others are available.

These operating systems allow to overcome a lot of limits innate in MS-DOS

such as, for instance, the possibility of managing only 640 kilobytes of RAM memory or the impossibility of managing a network, or again, the limitation of not being able to make multitasking programs. The evolution of MS-DOS, the OS/2 of IBM undergoes the same trend: with this system, in any case, it is not possible to manage a network, while it is possible to manage more than 640 kilobytes of memory and multitasking applications. We should notice that by virtue of these new available opportunities the concept of Personal Computer is less distinct.

Programming languages are included among the instruments for the database development: on one hand we assist to more and more up-to-date versions of languages (for instance DBASE III, improved later by DBASE III PLUS, until the very last DBASE IV) which allow personalized and easier development of a database for inexpert users as well. For instance, DBASE IV offers, compared to previous versions of DBASE III PLUS, a more powerful and easier user interface as well as the compatibility with the programs made with DBASE III PLUS without any other modification [21]. By this new language, furthermore, program execution is faster since before each execution programs are compiled, while with previous versions they were interpreted. An important feature of DB IV is that it is possible to make a multi-user database for network applications of distributed database. With DBASE IV it is also possible to take out data from archives of Mini or Mainframe network connected. With these possibilities, the power of personal computers increases even more. On the other hand, more and more sophisticated and powerful programming languages are used; they often come from the design field of expert systems.

On this subject, an example is given by the use of the programming language Prolog created to be applied to problems of artificial intelligence and database management in the medical field [22]: this use is justified by the fact that Prolog available features to interrogate database are close to a logical way of formulating inquiries. For instance, statistics on patients are obtained by means of suitably defined reports: a report can be oldpatv, defined by attributes such as the year, number, percentage considering old patients inside the group of patients specified by attributes [22]. Let us assume that we want to know the number of examinations of old patients and their percentage versus the whole group of patients in 1986. The Prolog inquiry will be as follows: Which (x,y: oldpatv (1986,x,y)).

C language is used in different medical applications: signal processing is the natural application field of C and some extensions of it (C + +), by virtue of the fact that it is very easy to work on single bits. C language also allows, furthermore, thanks to remarkable use flexibility, management of alphanumeric information [23].

In the medical field, as well, application packages commonly called Personal Productivity Tools are gaining ground: these programs (Lotus 1-2-3, Symphony, Open Access and others) allow the single user, besides handling database of personal interest, also minimal but useful statistical data

processing: after data input, usually organized in bidimensional tables, it is possible to obtain the value of derived quantities, such as pondered average on some special values of the table, besides usual statistical quantities such as average and variance of all present values. These instruments are very efficient in the representation of various obtained results: besides traditional displays, such as spreadsheets, we can also easily obtain pie charts, bar charts, even tridimensional, which are usually more effective for result comparing.

*Medical software.* The computerized management of clinical information gains special attention from the physician: during recent years, management programs of hospital files have been carried out and became too complex for daily use. In order to obtain an effective use of available hardware resources, it is necessary to dispose of software which can meet present specific requirements, is easy to use and, at the same time, allows speed of use. In Italy too, a great number of hospital file management systems are spread out [24]. Until some years ago, management applications of clinical information were available only on mini computer: nowadays the trend is to implement management systems of hospital files also on Personal Computer which are already powerful enough to allow this kind of application [13]. Besides implementation of programs of hospital file management, a lot of attention has been paid on design development instruments of these application packages: in this respect the MUMPS example is remarkable.

MUMPS (Massachusetts General Hospital Utility Multi-Programming System), which has become considerably widespread in the USA, is a high level interpreted language particularly suitable to the development of management systems of hospital data using interactive applications. This language, initially developed for laboratory computer management, is characterized by a general importance for on-line medical information systems [25]. MUMPS has been developed for management of character string data [26]. This kind of data appears very often in many sectors, also non medical ones. MUMPS has become a standard approved by ANSI (American Standardization Institute). Its standardization has encouraged its diffusion on different systems.

COSTAR is a computer-based management system for ambulatory practice and has been developed at the Massachusetts General Hospital since the 1970s [27]. COSTAR is an example of how to face complex problems innate in this application field. COSTAR has been designed in MUMPS. A language for database inquiry has been created ad hoc for the physician inside COSTAR. This language (MQL, Medical Query Language) has been developed in order to allow simplicity and suitability in inquiring database, even for an inexperienced user. In more detail, MQL is easy to use, does not require long training time to learn its use, requires only a general knowledge of database structure and can be used for several databases [27]. We should notice that above mentioned features can be applied, at least at intentional level, to the the majority of Computer Based Medical records systems. Diffusion and effective use of COSTAR have gained ground and a COSTAR users' group has even been created. This example seems

remarkable in order to point out steps to follow to avoid the premature obsolescence of such a system, or even worse, the physician's refusal to use it.

### *3.3. Hardware and software environment*

Macintosh Personal Computers of Apple, together with new NEXT, represent the only alternative to the numbers of IBM-compatible Personal Computers based upon MS-DOS operating system. For these computers we can talk generally of hardware and software since they form a unit which is not easily separable. Available operating system and development environment, the use of a mouse, an icon-structured menu, as well as a menu management of peripherals, characterize this category of machines to the point that a distinction between hardware and software would not allow a correct estimation. Macintosh personal computers use microprocessors of Motorola 68000 family (68020 and 68030) and are available in different patterns (with or without hard-disk, possibility to assemble mathematic co-processor or others) just like personal computers that use MS-DOS. Use of mouse, besides keyboard, is widespread and is one of the typical features of this kind of machine. Operating system and available development environment on Macintosh stand out for their menu-structured user interface on which mouse can operate. This mouse makes a small arrow move on the screen; by positioning the arrow on label or icon, which means the action to be undertaken, and pushing the button placed on the mouse, we select desired operation. File erasement is also made the same way: by means of the mouse we take icon which indicates file to be deleted to the trash can displayed on lower-right side of the screen. Use of these machines is, therefore, much more intuitive than it is with MS-DOS. By this very simple use, peripherals also are managed, such as printers, video-disc players [17], laser cards reader/writers [9] and so on. Among available development environments on Macintosh, we cite Hypermedia which permits management of great quantities of data coming from peripherals such as CD-ROM and Videodisc as well as realization of applications which can include graphic parts, database management, sound processing. It is interesting to take into consideration application package of Computer Aided Decision for Macintosh called Decision Maker [28]. It was developed especially for physicians: this program has the purpose to help the physician to make a decision by the creation of a 'decision tree'. Nodes which compose the tree represent different kinds of situations that can arise during the decision process: choice between two or more processes, final action (i.e. patient's death) characterized by a number which indicated expected utility and utility of action (typically, death has null expected utility) as well as the occurrence probability of the action itself, branching out node which represents an action which can give rise to two different results, also characterized by an occurrence probability, and other auxiliary nodes (for instance logical node). After having created graphically the decision tree which represents the particular situation of the physician, by means of the above mentioned mouse,

the physician gives various occurrence probabilities to actions as well as expected values of each final situation. Program at this point, shows the best decision to be made, that is to say the decision that puts together the best way expected utility value with occurrence probability.

NEXT computer, announced as the computer of the 1990s has some features that make it very interesting for the physician, in general, we can affirm that many characteristics which are not often found on one single machine are here put together [29]. Used hardware is extremely powerful and rich: 68030 Motorola microprocessor, 68882 mathematic co-processor, digital signal processing chip Motorola ASP56001, central memory of 8 megabytes that may be increased, besides communication busses allowing to transfer data at 4 or more Megabyte speed, these are some of the used components. Mass storage is formed by a magneto-optic cartridge reader: each of these allows the user to store 256 megabytes of data. These cartridges are made of the same material as CD-ROM. Besides this, a hard-disk of 670 megabytes is available. NEXT video, 17 inch monochromatic, offers a  $1120 \times 832$  pixel matrix having a greater number of pixels than the one usually available on 19 inch screens. But only 2 bits are used to code grey levels: reading these data from the point of view of a physician we can say that, by virtue of available calculation power, it is possible to process signals and images even though later screens will allow a wonderful signal display thanks to the high number of pixels and not of images owing to the scantiness of grey levels. The NEXT printer is a laser one and reaches a resolution of 400 dots per inch. This computer works under the UNIX operating system, which is already gaining ground in the personal computer field, it puts at the user's disposal a complete development environment which is easy to use. Similarly to what happens in the Macintosh world, NEXT allows menu management of various operations by means of mouse. Among software instruments supplied for NEXT there are a word processor, an ANSI-C compiler, a package to make a menu, through which it is possible to move in various applications with the mouse. Compared to Mac environment, in NEXT one, an inexpert user can easily operate through the menus controlled by mouse, but the programmer can also easily move in an environment similar to the traditional systems adopting UNIX. Thanks to used hardware it is possible, besides above mentioned signal processing, to synthesize sounds having a quality which is similar to the Compact Disk's.

### *3.4. Emerging standards*

ISO is the institution that, at an international level, deals with standard formulations allowing communication among different computers respecting the so-called implementation freedom of each manufacturing company. ISO standards for interconnection of open systems, called ISO-OSI are structured in 7 communication logical layers and each layer relies on lower logical layer: physical layer, data-link, network, transport, session, presentation and application ones from the ISO-OSI reference where application layer relies on



presentation layer which relies on session layer and so on [30]. Three lower layers are oriented towards the device allowing communication: at these three layers, data physical structure is defined (for instance words length of transmitted messages), structures which allow control of good result of transmission are defined as well as the communication of possible transmission mistakes. Higher layers, on the contrary, are oriented towards applications: summarizing we can say that, while lower layers assure that the transmission is correct, higher layers have the task to verify that transmission makes sense, i.e. two computers speak the same language.

MEDIX is the name under which IEEE-EMBS proposes a standard for the medical data exchange. The MEDIX project (IEEE Standard Project – 1157) was begun in October 1986. The main purpose of MEDIX (IEEE P1157) is to specify and establish a strong and flexible standard for data exchange among health information heterogeneous systems [31]. MEDIX is based upon and integrates itself with ISO-OSI standard application layer. The aim is to define a standard of interface transactions allowing information systems to exchange data. Proposed standards should assure data integrity, foundation, certainty, reliability and privacy. One of the problems faced inside MEDIX concerns the glossary of words: an important aspect of a standard in medical informatics is the definition of a common word vocabulary to which we should refer. When this is lacking, the clinical data input can be fruitless because of possible synonyms and subjectiveness in describing certain symptoms. MEDIX is a very wide standard project: besides mentioned subjects, fields of interest refer to: image transmission, data organization, inquiry languages, data privacy, and methods to correct transmission mistakes.

#### **4. Some important experiences**

##### *4.1. Hypermedia in echocardiography*

Hypermedia is a general concept which refers to computer capacity to manage, in an integrated way, great quantities of alphanumeric texts, but also database, suitable sounds and images, static and dynamic, by means of videodiscs or CD-ROM [32]. Several manufacturing companies have produced application packages which implement the above mentioned concept: IBM, NEXT, Sun, Digital, are some manufacturing companies which are implementing packages oriented towards this trend in different environments.

The most known and widespread implementation of Hypermedia is Hypercard [33]. Application packages Hypercard, developed by Apple, operates on Macintosh hardware. Hypercard, like other mentioned packages, allows a fast access to great capacity mass storages of hundreds of Megabytes such as hard-disks and CD-ROM through personal computer. In this system, man-machine interface is particularly simple and comprehensible: various controls are chosen from menus formed by graphic symbols (icons) having an

obvious meaning. Choice is made by using a mouse which moves a cursor on the screen until it indicates and selects desired command. Mouse can be replaced by a touch-screen. On the screens called touch-screens controls are selected also by pushing a finger on the screen on the area where they are displayed. Besides this, systems that implement Hypermedia are able to approach images stored on videodisc, besides managing and processing digital sounds. The wide image availability and ease of use after a short training period have encouraged the use of this system in the field of Computer Aided Education.

We believe it is necessary to illustrate briefly the 'Echocardiology Hypermedia Project' developed at Yale University-School of Medicine. The system has been developed on an Apple-Macintosh II Personal Computer with hard-disk. Through serial communication port, the personal computer controls a normal video-disc drive (in this case a Pioneer LDV 6000) loaded with an especially ordered video-disc on echocardiography contains 54000 images which can be directly selected. Stored material has been selected among more than 14000 diagnosis studies carried out during 12 years at Yale-New Haven Hospital. These numbers show clearly how, with systems of this kind, it is possible to put at the student's disposal a whole range of pathologies which can be diagnosed by echocardiographic investigation. Pathologies found more frequently are easily described with several images. This means that already in learning phase complete various modes and qualities with which one can really present an echocardiographic image in clinical practice, are available. Furthermore, most stored images are made through the so-called bi-dimensional echography, in the archives there are also other kinds of echocardiographic images: M-mode, pulsated and continuous wave Doppler, Doppler with colors and also examples of transesophageal echocardiography. Access to images contained on video-disc is possible by an index: a traditional database placed in the hard-disk of personal computer contains index file of video-disc as well as information concerning clinical situation of patients to whom images refer. For each described pathology the system gives several images with relevant data (patient's age, sex, clinical cardiac report and so on). Besides this, learning modules contain a part of text devoted to formal clinical description of the case and some animations. These have been made in order to illustrate the clearest and simplest way pointing out the part of cardiac muscle affected by a certain disease, correlations between echocardiographic image and found pathology. Animations, relationships among various screenings and text files are structured and defined through a high level language available in Hypercard environment: the HyperTalk language.

HyperTalk is an interpreted language and belongs to the category of object-oriented languages. The other language category is the one of traditional procedural languages such as C, Basic, Pascal. LISP and Prolog belong to the object-oriented languages. In procedural languages there is a strict distinction between programs and data. In object-oriented languages, central branch point is represented by a concept, the data of which can be an example. Programs make actions on these concepts. In Hypercard, the main concept is the 'card',

several cards are organized in structures called stacks and each card has inside smaller structures called fields, buttons, etc. To each card a 'writing' is allotted, for instance a piece of specific program of one card and some data as a text or a chart. As a first approach we can affirm that a screening relates to every card with all actions that can be chosen in that situation: every card includes, therefore, management of menus on video, graphic representations of all figures related to a certain situation, management of all text files, help files and similar which can be used in various learning steps with computer aid. Cards and, in a more general way, HyperTalk, are also a very good illustration of peripheral integrated management: in fact, by this language all information which has to be used every time is managed in a similar way, whether these images come from video-disk, or texts stored on hard-disks, charts or figures made with computer or obtained by scanner.

In this CAI system for echocardiography, by virtue of management ease of different information allowed by Hypercard, there are also sonorous helps such as, for instance, stethoscopic sound related to a certain pathology. In making a learning system with these instruments, special attention should be paid to card structuring and to the path or possible paths from one card to the other: these paths are studied so as to make learning the most efficacious.

As already mentioned, this system requires a Macintosh II or SE with hard-disk and a video-disc drive with suitable monitor. But it is important to point out that minimum configuration of this system does not require a video-disc drive. Among peripherals that integrate immediately with a system of this kind we have to mention the LaserWriter print which allows creation of high quality paper documents including charts, and the scanner which make the whole system versatile. A medium quality scanner, as, for instance, the Micro Tek 300 C which reaches a resolution of 300 dots per inch (dpi) and covers a scale of 64 grey levels, is already able to read in a satisfactory way some images of biomedical interest. In the future we can think of integrated systems which have recourse directly to motion-picture cameras to acquire various kinds of biomedical image.

#### *4.2. Geographic computer network among laboratories of different hospitals*

In Utah, a geographical network among laboratories of some hospitals has been developed [4]. This connection has been carried out in order to allow a more efficacious control of patients, to help data exchange of patients among different and isolated institutions from a geographical point of view.

In this geographical network computers of different trade marks such as IBM, Tandem, Prime and Data General, on which operate different software, have been connected: software helping decision making, software of administrative data management, of patient's clinical data management and so on. Besides a geographical network, there are several local networks. In order to develop such a system, attention has been paid in adjusting the majority to national and international communication standards: this operation allows

both possible further connections with processors of other companies and also a system obsolescence time non directly bound to hardware technological obsolescence of the same system. In this case, for communication between Prime and Tandem hardware, both installed in hospitals of this network, reference has been made to the seven ISO-OSI layers. For the first three layers, physical layer, data-link layer and network layer, software packages based upon X.25 protocol have been used. Transport layers and session ones have been created locally according the IEEE standards while presentation layer uses X.409 syntax (CCITT standard, later completely adopted also by ISO). From this experience we can draw many remarkable comments concerning communication standards: protocols to which one should refer are numerous and give the feeling, confirmed unequivocally by authors, that the number of proposed protocols is of such an extent that it causes confusion and bewilderment in a hypothetical user. At the moment of implementation, these protocols do not allow any essential operation, such as, for instance, transmission of images from different centres while they also lack indications, especially in presentation and application layers.

#### *4.3. A System oriented to clinical Data Management through a Laser Card*

One of the most important problems in hospital systems is preparation, maintenance and taking out of documents on which activities concerning a certain patient are recorded in a chronological order. In our society this document is the file kept in USL (Local Health Units) archives which contains documents of various hospitalizations, information on the most important illnesses, on possible special treatments concerning a patient, and so on. This kind of document often contains illegible, hand written remarks and is sometimes incomplete. When needed, these documents are often not available either because they are passed from one place to another to be filled in or because people lose them. Furthermore, in emergency situations, it is not possible to go to the archives where these papers are kept and to look for those of the patient in question. Computers have made this kind of document available inside and outside the hospital, when suitably stored, contrary to above mentioned situations. Storage of clinical information of a patient by means of a computer can be carried out with an organization of data for problems, pathologies, and not in a mere chronological order: this can become profitable in clinical practice. The idea which puts together above mentioned remarks, developed at the University of Houston, is to allow the patient to carry with him/her his clinical information stored on a suitable device [9].

Laser card has some features that make it very interesting for all illustrated purposes: it can contain 2 megabytes, it is not easily damaged, has small dimensions (54×85 mm) and requires a cheap device for its writing and reading.

In this application, a system which uses Laser Card has been implemented on an IBM Personal Computer. Special attention has been paid to software

development: a physician who is not expert in informatics should also be able to use this system. Patients, also, have to be convinced that it is necessary to carry the card all the time with them since it allows hospital staff to have access to their clinical data.

Data stored on a card are divided into several groups, staff of different categories can have access only to information they are concerned with: administrative staff can have access only to social economic data, chemists can only look into prescriptions, physicians can have access to all data. These different professional categories are divided by means of codes which allow access to the system. Inside information groups (individual data, social and economic as well as family data and so on) one can have access to different kinds of data: identification, anthropometric, laboratory data, ECG, immunizations, etc. Interface towards user is organized as menus: it is possible furthermore, to have access to information concerning different periods of a patient's life. For pathology definition, this system adopts ICD codes (International Codes of Diseases). This coding goes in between attempts to meet above mentioned requirements to have a vocabulary of pathologies to refer to. Implementation of this system on a personal computer, to conclude, allows a progressive diffusion and costs which can be borne also by a single physician.

## **5. Conclusions**

Several problems concerning application of informatic technologies in cardiology are still unsettled: many applications tend to sort out special problems in special sectors such as echocardiography or out-patients control. Although this is a desire of all people working in this field, we still have not reached products which integrate all information collected on a patient, whether they are images, signals or alphanumeric information. Neither integration among these different sectors of the same hospital or among hospitals has been reached: applications in this sector show how complex this problem is and point out many technological and/or methodological questions which have to be sorted out throughout the whole path: network transmission speed, communication protocol standardization, medical knowledge standardization both at a structural level and a terminology level. New technical products like NEXT give hope that these problems will be sorted out: in this solution we feel that we can forecast that Personal Computers will play a more and more important role. These computers are gaining a remarkable role in medical informatics: first of all they have a much more friendly user interface than any kind of Mini or Mainframe; furthermore, they have acquired versatility and calculation power to such an extent that they are able to meet different requirements: from signal processing to distributed database management, to data statistical processing, to peripheral management of various kinds such as video-disc, CD-ROM, Scanner and so on.

## Bibliography

1. CADANS (1988) The Information network for the Dutch University Cardiology, presentation book, Holland
2. Pincirolì F, Meester GT (1988) EEC Commission – ERASMUS Program, Description of the Form and Content of the Study Visit, Critical Aspects and Conclusions. Final Relation
3. Shafaman M, Mapp T, Yip R (1988) Evolution of the UCSF Distributed Hospital Information System. 12 Ann Symp Computer Applications in Medical Care. 632-636
4. Huff SM, Evans RS, Gandhi S, Jensen B (1988) Development of a Regional Laboratory Healthcare Network. 12 Ann Symp Computer Applications in Medical Care. 643-647
5. Barr CE, Komorowski HJ, Pattison-Gordon E, Greenes RA (1988) Conceptual Modeling for Unified Medical Language System. 12 Ann Symp Computer Applications in Medical Care. 148-151
6. Sager, Friedman, Lyman (1987) Medical Language Processing: Computer Management of Narrative Data. Addison/Wesley Ed.
7. Arenson RL, DeSimone D, Seshadri SB, Kundel HL (1988) Satellite Image Transmission at the University of Pennsylvania: a Status Report. 12 Ann Symp Computer Applications in Medical Care. 470-472
8. Taria RK (1988) A Picture Archiving and Communication System Module for Radiology. 12 Ann Symp Computer Applications in Medical Care. 478-482
9. Brown JHU, Vallbona C (1988) A New Patient Record System using the Laser Card. 12 Ann Symp Comp Applications in Medical Care. 602-605
10. Bertinelli M, Castelli A, Combi C, Pincirolì F (1989) Data Compression Applied to Dynamic Electrocardiography. Medical & Biological Engineering & Computing. 33-40
11. Cerutti S, Bianchi A, Bontempi B, Comi G, Gianoglio P, Liberati D, Micossi P, Natali Sora MG (1988) Quantitative Analysis of Heart Rate Variability Signal in Diabetic Subjects. IEEE EMBS 10th Ann Conf 145-146
12. Pincirolì F, Castelli V, Mosca G (1988) D.E.K.G.-Trainer son P.C. Machines: a CAI System for Arrhythmias in Electrocardiography. Proc of MIE 87, Roma.
13. Rossi R, Pincirolì F (1988) GPC. A Time Oriented Medical Record for Cardiological Patient. Abstract in Experiences of Computer Based Management of Cardiac Units, Satellite Meeting of 1st Word Congress of Pediatric Cardiac Surgery. Int Conf.
14. Digital-VAX technical information book, (1988) USA
15. Futami R, Hoshimya N (1988) Recognition of Temporal Patterns using State Transition of Neural Network, IEEE EMBS 10th Ann Conf 1488-1489
16. Wilson P (1988) Parallel Processing Comes to PCs. In BYTE, 213-218
17. Stensaas SS, Sorenson DK (1988) HyperBrain and Slice of Life: an Interactive Hypercard and Videodisc Core Curriculum for Neuroscience. 12 Ann Symp Computer Applications in Medical Care. 416-420
18. Chueh HC, Barnett GO, Hayes WC, Beck WS (1988) Hemavid: a flexible Computer-based Interactive Video Resource for Hematology. 12 Ann Symp Computer Applications in Medical Care. 421-425
19. Friedmann RB (1987) Interactive videodisc. In Images, Signals and Devices, MD Computing.
20. Felician L (1988) Image Base Management System: a Promising Tool in the Office System Environment. Data Base, SIGBDP-ACM No. 1
21. Ashton-Tate dBase IV technical information book (1988) USA
22. Allison WA (1988) A Physician's Approach-designing in Prolog: a deductive Database of Patient Statistics of Patients with musculoskeletal Disorders. 12 Ann Symp Computer Applications in Medical Care. 579-582
23. Pincirolì F, Pozzi G. (1989) A drug-to-drug interaction package. Computer Methods in Biomedicine. 183-189
24. Milazzotty F, Martiny W, Matta F, Pincirolì F. (eds) (1987) Banche di Dati Integrate in Cardiologia, Centro Scientifico Torinese

25. Beaman PD, Justice NS, Barnett GO (1979) A medical information system and data language for ambulatory practices, *Computer* 9-17
26. Kuzmac PM, Walters RM, Penrod G (1988) Technical Aspects of Interfacing MUMPS to an external SQL Relational Database Management System. 12 Ann Symp Computer Applications in Medical Care. 616-620
27. Morgan M, Piggins JL, Zielstorff RD (1988) Issues in Computer-Based Ambulatory Medical Record. 12 Ann Symp Computer Applications in Medical Care, TUTORIAL
28. Digital Medicine, Inc, D-Maker, Decision Analysis for the Macintosh, Handbook.
29. Thompson T, Baran N (1988) The NEXT Computer. In *BYTE*. 158-175
30. Le Moli G (1988) *Telematica: Architetture, Protocolli, Servizi*, vol I, ISEDI
31. Medical Data Interchange Committe (1988) IEEE Standards Project – 1157–MEDIX, Internal Reports
32. Jaffe CC, Lynch PL (1988) HyperMedia in Medical Instruction. 12 Ann Symp Computer Applications in Medical Care, TUTORIAL 416-420
33. Goodman D (1988) *The complete HyperCard Handbook*. 2nd Edition Bantam Books, NY

# A query language for medical statistical analysis

R. BRENNECKE, H.J. RUPPRECHT, R. ERBEL, J. MEYER

*2. Medical Clinic, Johannes Gutenberg-University, Mainz, West Germany*

## **Abstract**

While standard query languages support primarily the definition of single queries, in the evaluation of medical studies one usually formulates large sets of interdependent queries. A set of this type is called an integrated transaction. Our system for the definition of integrated transactions is based on the observation that in medicine a large number of statistical evaluations is founded on a conceptional model that can be structured as a tree. We describe a screen oriented tree editor for the relational data base system DBase and report on our experience with its application in the evaluation of the success rate of PTCA interventions.

## **1. Introduction**

Relational database systems become more and more the standard tool to obtain, maintain and manipulate patient data for clinical and scientific applications. These database systems usually incorporate also a query language. The user interacts with the system by submitting requests for data selection (queries) expressed using this language. A sequence of queries which is logically a single unit of interaction with the database is called a transaction. In typical query languages, a transaction has usually the size of one to several lines. This type of query gives only a limited (atomic) access to the data stored in the relational database. If we have the task to answer a general question such as the success rate of percutaneous transluminal coronary angioplasty (PTCA) interventions, we will usually define a large number of these atomic queries. If they are entered into a batch file, this can be considered as an integrated transaction. However, this method of entering an integrated transaction is time consuming and often typographical or logical errors will be encountered. Moreover, it will generally require some computer experience to define, enter and use such a transaction.

Our system for the definition of integrated transactions is based on the



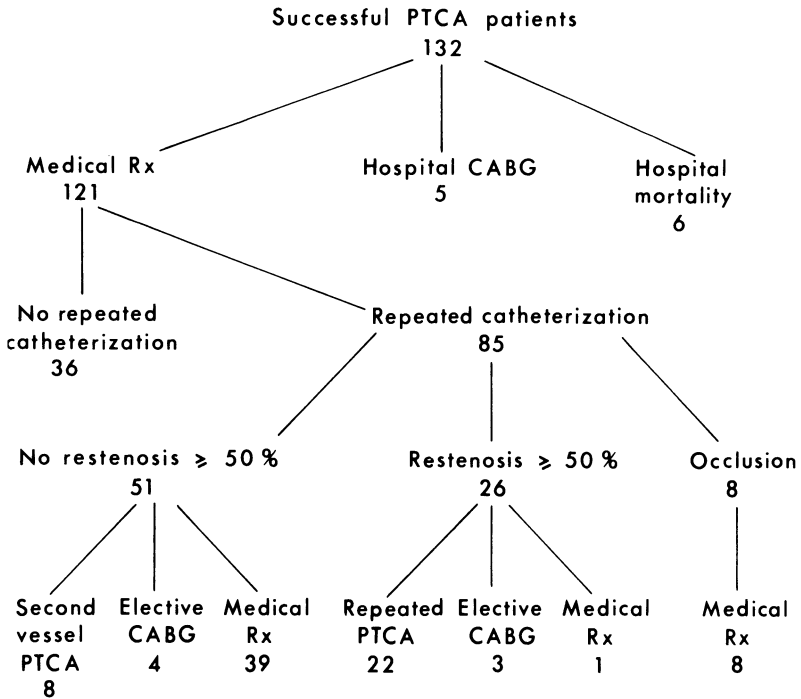


Figure 1. Example of a flow chart of follow-up results of successful percutaneous transluminal angioplasty (PTCA). Modified from Rothbaum *et al.* [1]

observation that in medicine a large number of statistical evaluations on the data stored in a relational database is founded on a conceptional model that can be structured as a tree. An example for these tree structures is the flow chart of follow-up results of successful PTCA shown in Figure 1. The numbers shown in this flow chart could be obtained by an integrated query or transaction that would consist of about 16 atomic queries. Here we describe a new method to enter all queries necessary for flow charts like this, that has the following advantages:

- a) one-to-one correspondence between the pictorial representation of the tree structure (Figure 1) and the definition of queries
- b) minimum amount of typing
- c) minimum risk of errors.

The described query language transforms the underlying relational database into a statistical database systems, i.e. into a database system that supports statistical data analysis. In this chapter, we discuss the design of the screen oriented query language for statistical databases, called Tree-Base. This query language was constructed as a highly secure and user-friendly statistics database system. While the data entry, maintenance and manipulation is based on the time proven advantages of the relational data model, a medical user mainly interested in statistical results from the database is provided with a man-machine interface to these data, that is based on the tree structure. Since this type of

structure is not only found in a large number of medical studies (Figure 1), but seems also nearly self-evident even for people without computer knowledge, the definition of large integrated transactions is simplified and many sources of errors are eliminated.

## **2. The PTCA-database**

### *2.1. Relational data storage*

In our clinic, the results of PTCA-interventions are stored using a relational database system (DBase III Plus, Ashton Tate). The PTCA database consists of 6 database files. These files contain a total of about 200 fields (attributes). Key fields (one, two or three per record) serve as synchronizing information. Index files based on these keys are maintained during data acquisition and form also a basis for processing queries of the type shown in Figure 1.

While demographic data such as the date of birth of the patients are entered only once (one record per patient in the patient data file), several records of anamnestic data may have to be stored in the anamnestic database file, one for each intervention. The file containing measurements of the extent of stenosis may even contain several measurements for each of several interventions of one patient (e.g. in the case of multi-vessel dilatation). Presently, the database contains data of 1500 patients.

The data are entered into these database files using special shell programs. In this step, only tools that are commonly used in relational database systems are needed. For the integrated queries necessary in the statistical evaluation of the data stored in this way, however, the user applies the interface described in the next paragraphs.

### *2.2. Principles of tree- base*

In the introduction, the metaphor of the tree was introduced since it corresponds to an important class of queries necessary in statistical evaluations of relational databases in medicine. For the incorporation of this metaphor into a computer program such as the already mentioned statistical database interface, it is an advantage, that the properties of 'trees' are also extensively studied in computer science. The elements of a tree are known as the nodes and the branches (Figure 2). This figure shows at the right the simplest tree, a binary tree.

Figure 3 compares an example of a general tree structure (Figure 3A) with a binary tree (Figure 3B). The binary tree is used in the design of our database query language in order to simplify the design of queries and thus reduce errors to a large extent. The binary tree possesses two branches or 'children' from each branch or 'father'. The filter conditions are attached to the nodes. Only one of the two conditions B and C have to be entered, since in the binary tree the

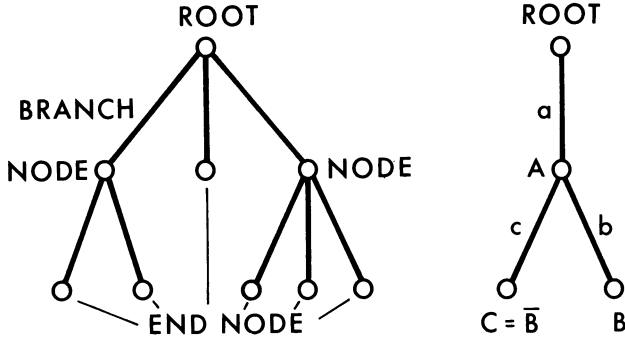


Figure 2. Basic structure elements of a tree (left). In a binary tree (right), the condition associated with node C is complementary to the condition entered by the user for node B.

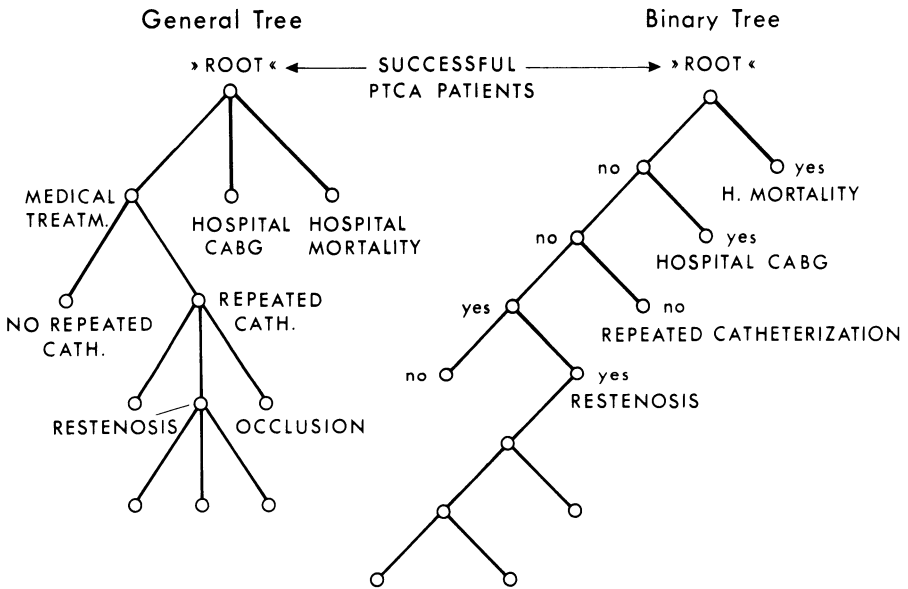


Figure 3. Examples of a general tree structure (A) and the equivalent binary tree (B).

second condition is the negation of the first.

During experiments assessing the value of the tree structure as a metaphor for integrated transactions we discovered, that the underlying metaphor of the query tree should be represented to the user in the slightly modified form of three classes of trees:

(a) The *join-tree* merges fields from records of several databases into a single database (Figure 4). It is a quite common practice in the evaluation of complex relational database structures consisting of a large number of files with different keys to merge prior to statistical evaluation into a single rectangular

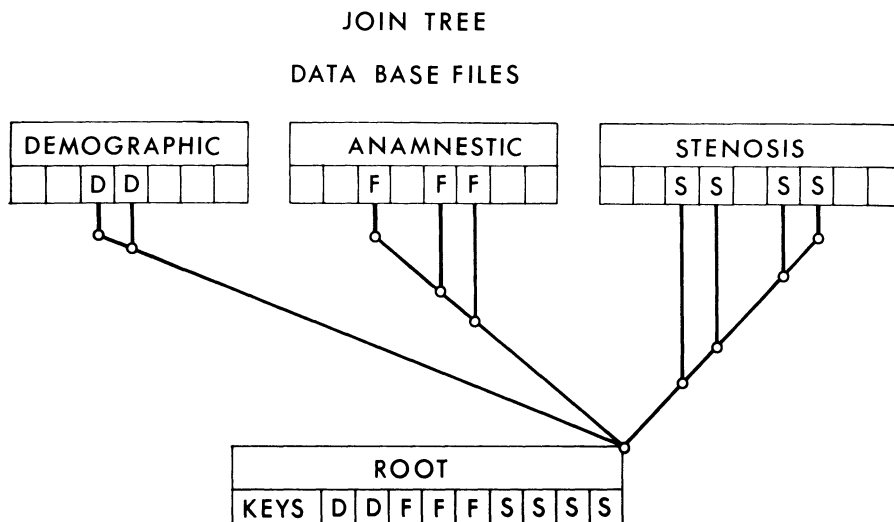


Figure 4. Example of a join-tree that selects tupels from several databases to construct the root database. In the same process, related time dependent fields from different records are combined into a single record in the root database.

file the relevant fields from these separate data files. For this purpose, the JOIN- and PROJECT- operations are provided by relational database systems [2]. The join-tree offers a pictorial representation of these operations to the user. Since joining the databases is a straightforward process, the join-tree is not as essential for our concepts as the other two trees described below. The result of performing the operations defined in the join-tree is a merged rectangular database that we call the 'root', since this database forms the root for the trees described below. An extension of this join operation also takes into account temporal relations as described below (Section 2.4.).

(b) The *selection-tree* (Figure 3) symbolizes the hierarchical structure of the queries that evaluate the 'root-' database provided by the join-tree (Figure 4). The underlying command in the general scheme of relational databases is the SELECT- operation [2]. The selection tree is the central concept in the query language described. Figure 1 gives a typical example for the purpose and the structure of a selection-tree. The result of each select operation is a table containing the data selected by the query. In our language based on the idea of integrated queries, each branch of the tree represents one table. The branch leading to the second end node in Figure 3B e.g. means: 'table of all cases with hospital CABG, contained in the cases with no hospital mortality'.

(c) The *statistics-tree* (Figure 5) represents recurrent statistical evaluations performed at the branches of the selection-tree. When we consider the node as a type of filter that allows only records with a certain combination in the contents of the data fields to pass the preceding branch, then the statistics-tree does the evaluation of the filtered records. The principal function of the statistics tree is to enter so called *aggregate functions* such as COUNT, SUM,

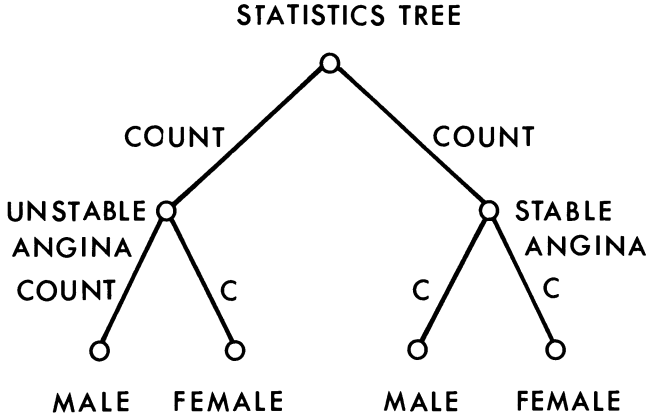


Figure 5. Example of a statistics tree. The root of this tree is associated with a branch of a selection tree (Figure 3).

MAX. Usually this type of aggregate functions have been incorporated in an ad hoc manner into the query languages of databases such as DBase or Oracle. Ozsoyoglu *et al.* [3] have pointed out, however, that the relational algebra and the relational calculus query languages introduced by Codd [2] do not formally incorporate aggregate functions. Although it is certainly possible to add aggregate functions to the relational calculus query language, we see it as an additional advantage of the structuring of the new query language into three types of trees that level (a) and (b) are only concerned with operations allowed by relational algebra and that the separate level (c) is primarily concerned with aggregate functions.

Here, the terms 'join' and 'select' are used in the sense of definition of relational database schemes [2]. In the programming language provided with the database system DBase, we find the same terms. In standard SQL, the 'join'- and 'project'-operations are hidden in a special type of 'select'-command.

### 2.3. Time oriented query processing

In the relational model, data from a single patient that describe the results of different interventions are stored in separate records ('vertically'). In a typical case, there may be five records for one patient in the stenosis data file, describing stenosis data before and after his first and second PTCA and data on a six month follow-up. In some cases, the stenosis data obtained before a second PTCA can be interpreted as follow-up data for the first PTCA.

For queries based on combinations of these data, commonly used relational query languages do not provide for direct access to the data stored in another record. We have therefore extended the above mentioned mechanism for the creation of the root database (primarily based on the concept of the join-tree) to include a transformation ('vertical' to 'horizontal' storage) of time related

fields thus simplifying the subsequent definition of time related queries by standard relational calculus.

As already mentioned, the join-tree merges fields from the records of several databases into a single rectangular database. This concept is extended in order to take into account the additional problem of time related queries. In a first step, time-dependent fields can be marked. These are then duplicated, so that for each intervention, variables of these fields describing subsequent events can be stored in a single row instead of vertically (e.g. 'stenosis after PTCA' and 'stenosis at follow-up'). Moreover, from the relative dates of the interventions, the system automatically computes a variable that classifies each intervention as a first PTCA without restenosis, a PTCA with restenosis and the necessity of a second intervention, as a second intervention without restenosis and so on.

The process described is related to the concept of the JOIN-operation both conceptually and also with regard to the type of database commands used, since the standard join operation can be applied for the described transformation. Thus, the inclusion of the transformation described into the join-tree seems to be a decision that is in accord with the underlying logic of relational systems.

### **3. Editing join-, selection- and statistics- trees**

The selection tree (Figures 1, 3) is the central entity of the statistical database described. A special program serves as a user- friendly editor for the creation of selection trees. Starting with the root node, one atomic or incremental query after the other is entered.

Instead of a query, a selection tree or a statistics tree created earlier may be entered at any new node. The nodes are automatically provided with a code number. When all conditions have been entered, the program synthesizes the total tree. This can be displayed on the screen or printed out in graphical form.

A useful extension of this scheme is the possibility to enter for each node a name, that characterizes its function in natural language. These incremental comments are also automatically concatenated. Thus, when afterwards navigating interactively through the nodes of the selection tree, the system not only displays for each node the integrated query expressed in terms of the query language of the underlying relational database system, but it also supports the user with a 'sentence' consisting of the incrementally entered comments. This automatically generated sentence is much easier and more rapidly interpreted by the user than the same command as expressed in the language of the original database system. Table 1 gives an example.

The statistics tree at present cannot be entered interactively. Based on our experience, we have included the evaluation of a set of database fields such as stenoses diameter into the program interpreting the selection tree. Thus, e.g. the mean stenosis diameter before and after PTCA is evaluated. A report is generated that contains for each node both the internal and the commented form of the integrated query related to the node and the data generated by the

Table 1. Comparison of incremental and integrated queries (left) and comments (right). These are related to the selection tree shown in Figures 1 3B.

INCREMENTAL QUERY	INCREMENTAL COMMENT
control>0	.and. repeated catheterization
INTEGRATED QUERY	INTEGRATED COMMENT
.not. hospex = .T.	.and. .not. hospital mortality
.not. CABG = .T.	.and. .not. hospital CABG
control>0	.and. repeated catheterization
reducpro<50	.and. restenosis > 50%
occlu = .T.	.and. .not. occlusion
dil = 3.or.dil = 5	.and. repeated PTCA

statistics tree positioned at the respective node.

The join tree is defined by entering the names of the database files to be merged. Then the user is presented with a screen showing the names of all fields (attributes) of those database files he or she has selected. The user accepts from this list the attributes he or she has incorporated into select- or statistics trees. The key fields are known to the program and these fields are automatically included into the root database.

#### 4. Discussion

The system has been accepted by users inexperienced with data processing as a tool simplifying to a great extent the entry of large and often changing sets of queries. The tree metaphor supports the intuitive design of queries by inexperienced users. The fundamental role of the tree property of a large class of queries has been mentioned in the literature before.<sup>4</sup> More specifically, in our work the tree metaphor and the graphical display of the trees was found to clarify the process of query design at an early stage. The selection tree editor makes it easy to enter into the computer selection trees, that have been designed with paper and pencil. Important numerical results can be directly written or printed into the graphical scheme printed out by the computer (compare Figure 1) offering a structured synopsis of results. The join-tree editor and the concept of the statistics tree simplify other tasks to be performed in the definition of integrated queries.

When comparing Tree-Base with other user friendly query languages [4] such as query-by-example (QBA), one finds several differences. The most important is, that while most other languages simplify the creation of single queries, tree-base supports the creation of *sets* of queries ('integrated queries'). Selection trees with a depth of up to nine generations and more than 50 nodes have been in routine use. A representative tree was equivalent to a set of nearly 200 queries concerning a general medical topic [5], the results of primary and secondary

dilatations in stable and unstable angina pectoris.

The concept of the statistics tree is not directly related to the basic structure of a query in a relational database. As mentioned above, the statistic tree essentially contains the so called aggregate functions that have to be added to the kernel of query processing especially in applications to statistical databases. A simple statistic tree contains the function COUNT (i.e. number of all records appearing at this branch), or the names of those fields that have to be evaluated and functions such as AVERAGE or MAXIMUM to be performed on the fields specified. The function LIST has to be added to allow the selective storage of tables that are to be processed using standard statistics packages such as SAS.

For a statistics tree editor, it may be more appropriate to use the metaphor of a fiber bundle [6] instead of the metaphor of the tree, that is ideal for the description of the selection tree.

At this time, we have applied the concept of a tree-structured query language only to the PC-oriented database system DBase. The application to the emerging standard SQL is planned. For the future, it is also our goal to replace the join-tree editor used for definition of the join-tree by a program that automatically generates the join-tree based on the data stored in the selection tree and in the statistics trees or statistics fiber bundles. Since the names of all fields referenced by any incremental query are contained in the definition of the selection tree and the statistics tree, the join-tree can be derived from the selection- and statistics-tree and from the additional information provided by a data dictionary specifying the field names of each database file. Index files are maintained in conventional ways during relational data entry.

## **Conclusion**

The statistical database system described above has been in use at our institution for more than 2 years. During this time, the system has always been accepted as a tool simplifying to a great extent the entry of large and often changing sets of queries. The tree metaphor supports the intuitive design of queries by inexperienced users. This graphical interpretation of the structure of integrated queries can help to discuss the design, the modification and the results of a set of queries by people without computer experience.

## **Acknowledgement**

This work was supported by funds from the Deutsche Forschungsgemeinschaft and from the Robert Müller-Stiftung.



## References

1. Rothbaum DA, Linnemeier TJ, Landin RJ, Steinmetz EF, Hillis JS, Hallam CC, Noble J, See MR (1987) Emergency PTCA in acute myocardial infarction. A 3 year experience. *J Am Coll Cardiol* 10: 264-272
2. Wiederhold Gio (1983) Database design. Mc Graw Hill
3. Ozsoyoglu ZM, Ozsoyoglu G (1975) A query language for statistical databases. In Kim W, Reiner WS, Batory DS (eds) *Query Processing in Database Systems*, Springer 1985, 171-190
4. Jarke M, Koch J, Schmidt JW (1985) Introduction to query processing. In Kim W, Reiner WS, Batory DS (eds) *Query Processing in Database Systems*. Springer 3-28
5. Rupprecht HJ, Brennecke R, Bernhard G, Erbel R, Pop T, Meyer J (1990) Analysis of risk factors for restenosis after PTCA. *Cath Cardiovasc Diagn* (in press)
6. Jung D, Brennecke M, Kottmeyer M, Hering R, Clas W, Erbel R, Meyer J (1988) Tree- Base. A query language for a time- oriented cardiology database system. *Comp Cardiol IEEE press*, 381-385

# Enhancing relational database management systems by applying artificial intelligence techniques

WILLEM R.M. DASSEN, ROB G.A. MULLENEERS and  
WILLIAM P.S. VAN BRAAM\*

*Department of Cardiology, University of Limburg, Maastricht, The Netherlands*

## **Abstract**

In this chapter the possibilities of using expert systems to interrogate large databases will be discussed. Two applications will be highlighted where the expert system increases the functionality of the database, by enabling a more intensive search or by supporting its use by non-experts. Finally new developments in this field of artificial intelligence research will be presented.

## **Introduction**

Databases used in medicine are growing explosively, both in size and complexity. This development has accelerated the search for tools and techniques to facilitate the management of such databases. In this chapter the use of artificial intelligence techniques, especially expert systems and natural language handling in relation to databases, will be discussed.

## **History**

Starting in the sixties attempts were made to develop computer programs that could think, that is, 'solve problems in a way that would be considered intelligent, if done by a human' [1]. These attempts formed the birth of a new science: the Artificial Intelligence (AI).

The first generation AI researchers tried to find general modes of solving broad classes of problems. But the more classes of problems such a general-purpose program could handle, the more disappointing results were noted on any individual problem. In the seventies one realized that this approach would not be fruitful, and more emphasis was put on the representation of the

\* The STRAIT Foundation (STimulation of Research in Artificial Intelligence Technology), Amsterdam, The Netherlands.

problem to solve, and on more intelligent search strategies. But the final breakthrough came at the end of that decade as AI researchers more and more realized, that it were not mainly the inference schemes used, that determined the problem solving capacity. Programs could be made intelligently by incorporating lots of high-quality domain specific knowledge. This understanding represented the birth of the expert system.

### **Expert systems**

In an expert system the knowledge of a human expert is represented in rules and stored in a separated part called the knowledge base. The inference engine uses this knowledge while solving a problem. Generalizing expert systems differ from conventional programs in a number of ways:

They must demonstrate expertise in their domain to be accepted. Expert systems represent the knowledge they use in a symbolic way using complex rules. Finally these systems can optimize their reasoning and explain their operation.

This knowledge representation makes it relatively simple to add new rules, and edit or delete old ones. Apart from the symbolic knowledge represented in the knowledge base, all expert systems contain a smaller or larger database with information on the real world.

### **Databases**

Since an expert system usually contains a database, there must be more informational value attached to the expert system than to the database. A relational database consists of entities, entities consist of attributes. Entities are connected by relations. Relations can be one to one, one to many or many to many. An example of a one to many relation between entities could be that between doctor and patients: one doctor treats many patients. An example of a many to many relation between entities could be that between patients and medicine: one patient may take many forms of medicine, while at the same time a particular medicine may be taken by many patients. A doctor, patient or medicine are thought of here as entities, while 'treats' and 'takes' are the real world equivalents of relations. Generally one can think of an entity as a noun and of a relation as a verb. Now the reason to split up these real world data into entities is one of convenient maintenance of the database. When the name of a medicine changes, then one has to change that name in the database only once, instead of having to change it in the record of every patient who takes that medicine. This means that the relations are not so much established reflecting the real world, but more as the result of technical considerations. One speaks of logical relations.

Relations, such as expressed in the rules of an expert system however, are

semantic rather than logical. A semantic relation is not concerned with ease of maintenance of data, but much more with reflecting meaningful real world relations. For instance, although there may be a logical relation defined between patient and medicine, but none between doctor and medicine in the database, in real life there may well be one. Therefore, an expert system that takes care of the communication between user – experienced or not – and database, must be able to translate semantic into logical relations, so as to be able to retrieve the relevant data. In this sketchy example, the expert system should be able to figure out that, to establish this semantic relation between doctor and medicine, it should equate it to two logical ones, namely first that between doctor and patients, and next the one between patient and medicines.

Further, more detailed examples will follow below.

## **Applications**

Two major applications of expert systems in databases could be distinguished:

- a) Expert systems designed to assist an experienced researcher in examining the information represented in the database extensively. These systems therefore contain knowledge on the domain represented in the database.
- b) Expert systems who enable a large public to interrogate a database and to obtain knowledge using questions formulated in natural language.

An example of an expert system of the first category has been described by Porenta [2]. This expert system establishes a set of heuristic rules linking attributes in a database of patients with arrhythmias. In 1986 Blum described expert system assisted design of clinical studies using a routine clinical database [3]. With the help of a knowledge base of medicine and biostatistics the expert system was able to identify confounding variables, to determine methods for controlling them, and to build a statistical model. Furthermore it determines patient eligibility criteria and retrieves the essential data from the database. With this approach routinely collected clinical data may become a new resource for generating medical hypothesis.

An example how expert systems can make information, stored in a database, available to a large public has been described below. In this project natural language analysis techniques are applied to interpret the search request and translates it in a command sequence that will accomplish the corresponding search in the database.

## **Examples**

In these examples a relational database management system will be described, using six databases. Together they are intended to form a system to control all appointments in an outpatient clinic.

The six databases are:

- DOCTOR** a database identifying all physicians represented in the system,
- PATIENT** is a database containing all patient data,
- ECG** which contains information on all ECGs known to the system,
- FACILITIES** stores all information on special facilities required during the outpatient clinic visit, which are not available at all locations (pacemaker programmers etc),
- OUTPATIENT CLINIC** is the database linking patient and doctor, time and place to one appointment, and finally
- POSTMASTER** contains all messages generated by the entire system.

These six databases are linked together in one relational database management system. The user can communicate with the system with the help of an expert system.

This expert system can interpret commands given in natural language and translate it in commands in the symbolic query language used by the database.

The relational database management system **APPOINTMENT**

DB DOCTOR				
Dr#	Name	Room#	phone	Cluster
1	Kildaire	637	658	A
2	Watson	655	381	B
3	No	528	847	A
4	<b>Rossi</b>	<b>529</b>	<b>844</b>	<b>A</b>
5	Zivago	34	122	C
.	.....	....	....	.

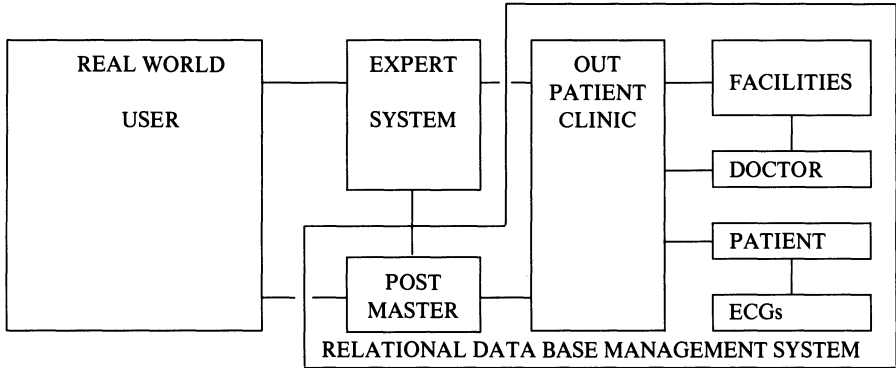
DB PATIENT					
Pat#	Patname	Address	City	Chart#	Complaint
62339	Jones	23 North 3 St	New York	652881	AP
43998	Smith	12 East 7 St	New York	642292	PM
<b>62887</b>	<b>Davis</b>	<b>66 South 1 St</b>	<b>New York</b>	<b>593433</b>	<b>Inf</b>
58835	Miles	19 West 6 St	New York	673994	Tach
61194	Parker	88 Middle St	New York	622175	Inf
.....	.....				

DB DOCTOR			
Pat#	Date	Time	ECG#
69845	04/18/90	16:34	1937533
54974	05/02/90	08:23	2183766
<b>62887</b>	<b>05/04/90</b>	<b>10:12</b>	<b>2213642</b>
63190	05/30/90	08:45	2242119
<b>62887</b>	<b>06/03/90</b>	<b>20:58</b>	<b>2329292</b>
.....	.....	.....	.....

DB FACILITIES	
Cluster	Loc.
<b>A</b>	*
<b>B</b>	*
<b>C</b>	6
<b>D</b>	*
...	...

DB OUTPATIENT CLINIC						
App#	Dr#	Date	Time	Dur.	Pat#	Loc#
63293	4	11/08/90	14:15	15	54228	2
<b>63294</b>	<b>4</b>	<b>11/09/90</b>	<b>10:30</b>	<b>30</b>	<b>00000</b>	<b>3</b>
63295	5	11/12/90	15:30	30	64229	6
<b>63296</b>	<b>4</b>	<b>11/12/90</b>	<b>14:30</b>	<b>15</b>	<b>62887</b>	<b>2</b>
63297	1	11/12/90	16:15	45	65289	5
63299	4	11/12/90	15:00	30	62217	1
.....	.	.....	.....	...	.....	.

DB POSTMASTER					
Mess#	Date	Time	From	To	Message
<b>144363</b>	<b>08/04/90</b>	<b>10:15</b>	<b>SYSTEM</b>	<b>Dr Rossi</b>	<b>Could you ..</b>
<b>178297</b>	<b>08/04/90</b>	<b>10:45</b>	<b>Dr Rossi</b>	<b>SYSTEM</b>	<b>Yes. I will...</b>
187878	08/05/90	14:35	Dr Kildaire	All	Nest friday..
.....	.....	.....	.....	....	.....



CATEGORY	ITEMS
Personname	Name, Patname, Who, From, To
Pers#	Dr#, Pat#
Location	Room#, Loc#, Where
Calendardate	Date, When
Clocktie	Time
Elapsedtime	Duration
.....	.....

All items at the right side belong to the same category indicated at the left side. In two examples the philosophy behind this approach will be demonstrated.

*Example one*

In the first example the following question:

**‘When is Mr Davis visiting Dr Rossi?’**

will be analyzed.

The expert system analyzing this natural language sentence classifies **‘WHEN’** as a calendardate, and assumes both **‘Mr Davis’** and **‘Dr Rossi’** to be a personname. **‘Is visiting’** is an operation that is related to the OutPatient Clinic database.

**When – Calendardate**

**Mr Davis – Personname**

**visiting – OutPatient Clinic relation**

**Dr Rossi – Personname**

In the Outpatient Clinic database no personnames, but two personnumbers are used: Pat # and Dr #. In this example two personnames are known: Mr Davis and Dr Rossi. First the ES interrogates the DB to determine if Dr Rossi is known in the DOCTOR database, and Mr Davis in the PATIENT database. Since the expression **‘is visiting’** is used, the system knows that the second name should be the doctor. If, for instance, instead of **‘is visiting’** the expression **‘is talking to’**

would have been used, it would not have been clear, who was the doctor and who the patient. The expert system will generate the following Symbolic Query Language commands:

in DOCTOR retrieve Dr # **where** name = 'Rossi'.  
in PATIENT retrieve PAT # **where** name = 'Davis'.

This query will result in Dr # = '4' and Pat # = '62887'

Subsequently, with the acquired information the expert system can determine the date requested:

in OUTPATIENTCLINIC retrieve Date **where** Dr # = '4' **and**  
Pat # = '62887'  
resulting in **Date** = '11/12/90'

Finally, a new sentence is generated by the expert system and sent to the user.

<b>When</b> 1	<b>is</b> 2a	<b>Mr Davis</b> 3	<b>visiting</b> 2b	<b>Dr Rossi</b> 4
<b>Mr Davis</b> 3	<b>is</b> 2a	<b>visiting</b> 2b	<b>Dr Rossi</b> 4	<b>on 11/12/90</b> <b>Date</b>

In this example, the correlation between the real world description and the items represented in the databases is relatively simple. If the search definition becomes more complex, the expert system cannot interpret and execute the operators used in the natural language sentence immediately, but has to design a schedule of database symbolic query language commands to perform the task.

#### *Example two*

Consider the following request in natural language:

**'Advance the next outpatient visit of Mr Davis and Dr Rossi, preferably with the same doctor'**

To perform this task the expert system has to perform a number of tasks in a row. A possible scenario could be:

Identify Dr Rossi and Mr Davis:

in DOCTOR retrieve DR # **where** name = 'Rossi'  
Dr # = '4'  
in PATIENT retrieve PAT # **where** name = 'Davis'  
Pat # = '62887'



If one or both items cannot be retrieved, try a search for the item 'Rossi' in PATIENT and 'Davis' in DOCTOR.

If successful identify the appointment number:

**in OUTPATIENT CLINIC retrieve APP # where DR # = '4' and  
PAT # = '64229'**

Try to find an earlier appointment by tracing for a slot in the schedule, cancelled by another patient:

**in OUTPATIENT CLINIC retrieve APP # where DR # = '4' and  
PAT # = '000000'**

If possible:

- Make new reservation for the correct location
- Delete old reservation
- Send Message via POSTMASTER to confirm change to doctor and patient

If not possible:

Try to find an open slot for an appointment with a doctor from the same cluster:

**in OUTPATIENT CLINIC retrieve APP # where cluster = 'A'  
and PAT # = '00000'**

If possible:

- Make new reservation for the correct location
- Delete old reservation
- Send Message via POSTMASTER to confirm change to old doctor, new doctor and patient

If still not possible:

Try to find an appointment anyhow:

**in OUTPATIENTCLINIC retrieve APP # where PAT # = '00000'**

In this way open slots caused by canceled appointments are filled efficiently. It is of course also possible to generate directly a new appointment with the doctor of choice.

## **Discussion**

Expert systems as mentioned above will soon become commercially available. Significantly more research effort will have to be invested to realize the application described below. Since the number of specialised databases is growing at an alarming rate, a problem will arise as to determine in which case to use which database. The problem tends to gain more weight as it has done in

other, non-medical areas since the most interesting databases in each particular case may well be located geographically far apart. Consulting these databases may well be a time consuming affair, even when one has access to local or global networks, and therefore these communications will be expensive. Apart from this geographical difficulty, there is the unfortunate circumstance that different database systems often speak different languages to query them with. Most users will not be interested in learning these un-English languages like SQL, DATATRIEVE and such, and neither should they be forced to learn to use them. Expert systems that take care of these problems, as the hypothetical one described below, have been successful for some time now in diverse areas like disaster control, identification of unknown chemical substances and the running of day to day traffic in large harbours.

Generally speaking a number of expert systems would be required to handle three different aspects:

- a) One system should be able to determine which database has the highest probability of a hit, that is of an answer of the highest possible quality to a particular query.
- b) Furthermore one system should shield the user by taking care of the syntactic and communicational idiosyncrasies of different databases;
- c) To be able to use the expensive communication as effectively as possible, one would require a semantic layer between the user and the databases.

This layer must be able to analyse a user posed problem, and to divide it into subproblems. For each of these subproblems an optimal answer may be anticipated by the expert system mentioned first. Then the second system takes over and actually retrieves those data from often very diverse databases and database systems. Finally, by the semantic layer these data are assembled into an intelligible and meaningful answer to the user. In this way the most efficient and effective communications are being used, while at the same time succeeding in determining the most high quality answer to a query.

## **Conclusion**

In new generations of database management systems, databases will not only contain logical entities and the logical relations between them, but also a body of semantic rules. These rules are representations of relations in the real world. An expert system executing these semantic rules, translates them into logical ones that in turn are used to retrieve data. Also it can add procedures to the data. For instance, if in such a medical database already patient length and weight has been stored it is useful to add knowledge how using length and weight also the body surface area could be calculated. By incorporating knowledge into the database the relational database can evolve to an object oriented one.

## References

1. Waterman DA (1986) A guide to expert systems. Reading, Mass Addison-Wesley
2. Porenta G, Binder T, Pharinger B, Anvari A, Weber H (1988) Evaluating antiarrhythmic strategies: a knowledge-based system for exploring clinical data. PACE 11: 2110-2113
3. Blum RL (1986) Computer-assisted design of studies using routine clinical data. Annals Int Med 104: 858-868

# Electronic information in cardiology: review of external databases and services

LUCIANA LIBUTTI

*Institute of Studies on Research and Scientific Documentation, National Research Council, Via Cesare de Lollis, 12 00185 Roma, Italy*

## **Abstract**

This paper reviews the current situation in the online database industry and describes various aspects of this market (online databases, providers, host computers, gateways, document suppliers). The paper also discusses the development of some user-friendly services (Dialog Medical Connection, BRS/Colleague) available in particular to the biomedical community. The second section of this paper describes in detail a range of databases available both online and on CD Rom for use in the cardiology sector. A comparison is made between the two media.

## **1. Introduction**

Today, more than in any other period, the possibility of accessing information sources is having a profound effect on the quality of our work. Never before have we had access to such an enormous mass of information. Nevertheless, it is still often a difficult task to find the right information quickly, when we need it.

The increased quantity of data available has thus revealed the inadequacy of the traditional system of information retrieval. This is particularly true in those sectors, such as scientific research, where it is essential to be able to apply an immediate check on the state of the art. Technological progress has brought about fundamental changes in traditional methods of producing, processing and retrieving information.

Nowadays, all that is needed is a terminal or personal computer, a printer, a telephone line, a modem or an acoustic coupler in order to use telecommunications networks specialized in data transmission which are linked to the switched telephone network and provide access to computers all over the world. These computers place at our disposal databases containing information related to all branches of knowledge and these data can then be used as the basis of further processing.

## **2. Online databases: typology and functions**

According to the type of information contained in them, there may be two main categories of databases: reference databases (bibliographic, referral) and factual or source databases (full-text, numerical, etc.).

Reference databases provide access to reference information (bibliographic references to publications in various fields, references to research projects, agencies and organizations, etc). They thus allow the primary information to be located and perform an extremely important function, i.e. that of indicating to the user the existence of documents published in his field of interest.

The user must then take further steps to find the original texts or obtain photocopies of them.

Source databases, on the other hand, supply the text of the document directly and provide a work tool that the researcher can use immediately. In recent years there has been a constant increase in the supply of source databases and also a tendency to produce mixed databases containing information of various kinds (e.g. bibliographic and full-text information).

Reference and factual databases should nevertheless be regarded as complementary to each other: in some cases, after searching a bibliographic database, the user may go on to a full-text database in order to retrieve the text.

## **3. The information market (providers, host computers, gateways)**

A database provider, who is responsible for the content of the database, gathers, selects and processes the data, indexes them and transfers them on to a magnetic medium. The host computer represents the commercial organization which, after acquiring the database from the producer, sells it on the market on certain conditions.

The host has the task of inputting data into its computers, updating them with material provided by the producer and to provide for the query language. The user wishing to have online access to a database must therefore apply to these organizations and enter into a user contract with them.

Databases may be accessed through hosts using information retrieval software which allows the user to interact with systems without any great difficulty.

Each host has its own query language, represented by a set of commands and procedures which allow the user to make his own searches. There exists a wide range of query languages, although they are all based on the same operating procedure: selection of terms, logical combination of terms (using the Boolean operators OR, AND, NOT), display/printout of results.

The differences between the various languages and the frequent assignment of different meanings to formally identical commands nevertheless represent an obstacle for the user. It is consequently advisable to learn the commands of a specific language by attending the one or two-day training courses periodically organized by the hosts.

Research languages vary considerably in complexity and the user may often have at his disposal extremely sophisticated commands, such as restriction of the search to certain years, specific countries, thesaurus display (if available), sorting of the retrieved data by year, etc.

In setting up their databases, several producers give priority to information retrieval using controlled languages (thesauri), and often provide online access to such dictionaries. Other producers allow only free terms to be used and others again recommend the combined use of both methods.

### *3.1. Intelligent systems for searching: gateways*

One problem to which database users are especially sensitive is that of being billed separately for each service for which a contract has been stipulated. In order to eliminate this inconvenience, network distribution services, known as gateways, have been introduced in recent years.

A number of host computers contract to provide the users of one service with simultaneous access to the files of another host without having to deal directly with the latter. The advantages for the user are that he has to make only one connection for which he receives a single bill. One further obstacle standing in the way of even the expert user is, as already mentioned, the problem of the different query languages used by the host computers which forces the user to memorize numerous commands.

An attempt has been made to solve this problem by setting up 'intelligent gateways'. This kind of service acts as an intermediary between the user and a certain number of host computers. It is thus possible for several databases distributed by different hosts to be accessed using a unified language and with the user being billed only once.

The user, who no longer needs to be familiar with the retrieval system, selects the best search strategy from among the options displayed in the various menus. The selected database will then be accessed by the system in a way that is totally transparent to the user.

The originality of the system consists precisely in the way in which the strategy is selected, i.e. quite independently of the query language specific to the individual database.

The first example of this type is the US Telebase System which has developed the *Easynet* intelligent gateway. The query language used has a tree structure linking more than 8000 menu displays. This system allows about 15 international host computers to be accessed, for a total of about 900 databases. If no data are retrieved by the user there is no charge. Payment may be made by billing or using a personal credit card. To help the user in developing his strategy, Easynet provides a round-the-clock 'SOS' system. When the user uses the 'SOS' command, a specialist provides him with online interactive help in formulating a strategy and selecting a database. No charge is made for this service.

Several countries have licensed national versions of this system from Telebase. The Italian equivalent is Magic On line.

Another gateway service known as *Intelligent Information Facility*, developed by the Geonet System and Infotap, with support from the European Economic Community, has recently become available. This initiative is in line with EEC policy to promote European products and services inside the Community. The service thus provides access to about 70 databases available on European host computers. This is a multilanguage service which allows the user to select the query language used. It is also possible to select both the host to be accessed and the search strategy, either using menus or the host query language.

#### **4. User-friendly services in the biomedical sector**

In the early 80's, with the rapid spread of microcomputers, online services began to develop simplified or 'user-friendly' query systems in which the end user was encouraged to make his own searches without necessarily involving an intermediary. These systems make searching database very easy and do away with the need to learn a complex set of commands related to the procedure for querying a particular database. A number of services of this type are available in the biomedical sector:

##### *4.1. Medical connection (DMC):*

Medical Connection is a Dialog service which can be accessed through a single contract also by subscribers to the traditional system and it is specially designed to cater for physicians, biomedical researchers and health practitioners in general.

There are two methods of access to this service: menu-driven and command-driven.

The first system, intended for inexperienced users, leads them step by step through the search: using the second or command-driven system, a number of commands which differ slightly from those used in normal Dialog searches, allow the entire query system to be exploited. At any stage the user can go from the menu to the command system and vice versa. On-line help is also available.

Through DMC the user can log into 4 information systems, which may be accessed singly or all together. These systems offer specialist information of various kinds and are divided into the following 4 sectors or 'libraries':

- Medical Reference Library, comprising different types of databases, (EMBASE, MEDLINE, PSYCINFO, INTERNATIONAL PHARMACEUTICAL ABSTRACTS, DRUG INFORMATION FULLTEXT, CLINICAL ABSTRACTS, HEALTH PLANNING & ADMINISTRATION, CANCERLIT);
- Biosciences Reference Library, referring to the biological science sector, (e.g. BIOSIS PREVIEWS, AGRICOLA, CAB ABSTRACTS, FOOD SCIENCE & TECHNOLOGY ABSTRACTS, LIFE SCIENCES COLLECTION);

- Sci/Tech Reference Library, referring to the mainly scientific and technological aspects of biomedicine (e.g. SCISEARCH, CASEARCH, COMPENDEX, INSPEC, NTIS);
- General Reference Library, containing databases referring to new technologies, computer applications; full-text of important encyclopedias (e.g. MAGAZINE INDEX, NEWSSEARCH, NATIONAL NEWSPAPER INDEX, ACADEMIC AMERICAN ENCYCLOPEDIA, BOOKS IN PRINT, TRADE & INDUSTRY INDEX, MICROCOMPUTER INDEX, MARQUIS WHO'S WHO IN AMERICA, ETC.).

#### 4.2. *Colleague medical search service*

Also this service provided by BRS consists of a user-friendly information retrieval system designed for health care professionals. It provides access to specialist databases at lower rates than those charged for standard BRS Search Services searches. The user is guided through the search process by a menu.

The service offers two systems, which also in this case are known as 'electronic medical libraries': *Bibliographic Indexes and Abstracts* and *Comprehensive Core Medical Library*. These two electronic libraries contain respectively bibliographic information and full-text databases. The service allows the medical subject headings of the National Library of Medicine to be searched in hierarchically ordered groups.

Online Help is available, as well as an electronic mail service and bulletin boards (*Medlink Mail Services*). The latter is an electronic service comprising more than 15 biomedical subjects that Colleague users can use for discussion amongst themselves: (AIDS, anaesthesia, cells, drugs, dermatology, geriatrics and gerontology, hackers (computer users), human genetics, immunology, medical ethics, neurology, child psychiatry, future events of interest to the medical community, etc.). A videotape training film illustrating the system may be purchased.

High-performance services like those described above, although having the considerable merit of bringing the online information market within the reach also of novices and occasional users, can in no way be considered as an ideal tool for all: simplification of procedures entails also a reduction in the exhaustiveness of the search since the user cannot use all the commands and procedures that would otherwise be at his disposal if he were using a query language. These systems also involve rather long access times which means higher costs. In conclusion, whether or not these services are to be recommended depends on the degree of sophistication of the user's search.

### 5. **Online databases for cardiology**

The mass of information available, particularly in the biomedical sector, has in recent years led to a proliferation of specialized information sources, such as



abstract bulletins (Index Medicus, Biotechnology Abstracts, etc.). To facilitate their management, many of these publications have been automated and are now available in online versions, thus providing a much more powerful research tool. Information transfer technology not only allows large masses of data to be filed away but also to be related, thus enabling the user to analyse the problems of interest to him from many different points of view.

The growth of the online information market has brought with it a considerable increase in database production, sometimes with considerable overlapping of the subject matter. This may hinder the user in choosing the product best suited to his needs. It is thus important, in order to make the right choice, to get in touch with the producer and ascertain the exact characteristics of the product as well as the source from which the data has been obtained. There are also several online commercial directories of online databases which provide the user with a useful tool for making his choice. One of these is the 'Directory of Online Databases' published by Cuadra/Elsevier.

The databases presented here provide a means for keeping abreast of developments in professional fields and share several characteristics which distinguish them clearly from other databases: they are available for online interactive accessing; they are publicly available, and therefore at the disposal of private individuals on certain conditions, e.g. by subscription, hourly consumer rate, etc. and are available through commercial services linked to one or more national or international telecommunications networks.

Listed below are several on-line databases of particular interest to the cardiology sector. However, I also deem it of interest to mention several which are available in other than online form (e.g. diskette, batch access, etc.).

### **Clinical abstracts**

Database produced by Medical Information System, Reference & Index Services and distributed by Dialog Information Services. It contains bibliographic references accompanied by abstracts from more than 300 medical journals covering the fields of paediatrics, internal medicine, general and cardiovascular surgery, genetics, infectious diseases, cardiology, dermatology, pharmacology, neurology, anaesthesiology, psychiatry, urology, allergology. It is available also on microfiches. The printed version is Abstracts in Internal Medicine.

It may be accessed also through Easynet and Medical Connection.

### **Cardiovascular disease prevention database**

This is a subset of the COMBINED HEALTH INFORMATION DATABASE and can be accessed online through the latter. It is produced by the US National Heart, Lung and Blood Institute Information Center of Bethesda and is

available through BRS and BRS/Colleague. It contains references and abstracts of published material concerning the education, prevention and health promotion aspects of blood resources, cholesterol, hypertension and smoking. The producer also provides a query service upon request.

### **Comprehensive Core Medical Library (CCML)**

Produced and distributed online by BRS Information Technologies, it contains the full text of current editions of selected medical reference works and textbooks, retrospective issues of more than 80 English language medical journals published worldwide: Acid base and potassium homeostasis (Brenner and Stein); Chronic renal failure; Manual of antimicrobial therapy and infectious diseases (Eisenberg); Brain failure and resuscitation (Grenvik and Safar); Manual of admitting orders and therapeutics (Larson and Eisenberg); Recent advances in intensive therapy (Ledingham); Principles and practice of infectious diseases (Mandel, Douglas and Bennett); Birch's emergencies in medical practice (Ogilvie); Textbook of surgery (Sabiston); Principles and practice of emergency medicine (Schwartz), etc.

Among the journals: *Annals of Internal Medicine* (from 1982); *Annals of Neurology* (from 1986); *British Medical Journal* (from 1983); *Coronary Artery Disease* (Boucek); *Coronary Care* (Karliner); *Journal of Neurology, Neurosurgery and Psychiatry* (from 1986); *Lancet* (from 1983); *Neurological Clinics* (from 1986); *Neurology* (from 1987); *New England Journal of Medicine*; *Seminars in Neurology* (from 1986), etc.

It is available also through Easynet and BRS/Colleague.

### **Embase plus**

Contains bibliographic references, about 4000000 records, to the 44 abstract journals and 2 indexes (*Adverse reaction Titles* and *Drug Literature Index*) produced by Excerpta Medica.

Reviewed articles are divided into 43 sections, each corresponding to one speciality and for each of which an abstract is produced by a specialist on the subject. The emphasis is laid on the action and side effects of drugs and related chemical compounds.

The data are abstracted also from monographs, talks and more than 3500 important journals in the sector; books have been included only since 1980. Patents are excluded. Four controlled dictionaries are used to index and retrieve data:

1. EMCLAS, a decimal classifying system with subdivisions corresponding to the biomedical areas; it consists of a five-level hierarchical system. The indexer can assign up to ten classifications for each record.
2. EMTAGS consists of 200 controlled terms representing a wide range of concepts.

3. MALIMET (Master List of Medical Terms), a controlled dictionary comprising over 500000 selected terms and more than 240000 synonyms. Malimet is available both online and on microfiche. It is updated retrospectively so that if the name of a disease changes or is subdivided into several intities, all the previously classified articles with the previous name are updated.
4. EMTREE has a structure similar to that of MeSH and consists of a hierarchical thesaurus, with the possibility of term expansion.

Furthermore, the text contains non controlled indexing terms such as the trade names of drugs. For drugs the file uses the WHO International Nonproprietary Names Scheme. It can be accessed through BRS, Data-Star, Dialog, Dimdi, as well as through Easynet menu-driven systems, BRS/Colleague and Dialog Medical Connection.

### **Journal watch**

Produced by Massachusetts Medical Society, contains citations and abstracts to clinical studies reported in 24 leading medical journals and other medical literature. Among sources are American Heart Journal; American Journal of cardiology; Annals of Internal Medicine; Circulation; Journal of American College of Cardiology, etc. It is distributed online by BRS and BRS/Colleague.

### **MEDIS**

Produced and distributed by Mead Data Central, Inc. (USA), it contains the text of more then 60 biomedical publications, newsletters and handbooks. This service, available to users since 1985, allows free language to be used to retrieve the full text of a large number of biomedical publications. The database is subdivided into 5 main 'libraries':

*Genmed* (general topics) containing the text of more than 40 specialist reviews and handbooks subdivided according to clinical speciality (e.g. Cardiology, Paediatrics, etc.);

*Pharm* (with information on drugs) containing Drug Information fulltext (DIF) and Handbook of Injectable Drugs;

*Cancer* containing, among other things, PDQ (Physician Data Query), a database giving the description of more then 1000 research programmes regarding the treatment of cancer;

*Medline*, online version of Index Medicus;

*Admin* containing the F-D-C newsletter and those of the American Health Consultants.

Each 'library' is subdivided into 'files': the reviews, the newsletters and the handbooks, which are also grouped by medical speciality and type of publication.

A file consists of 'documents' (articles, drugs monographies, abstract, etc.) with the full-text of the tables. Each document is also divided into 'segments' (title, authors, references, etc.). The user has the choice of making highly targeted searches.

### **Medizinische Technik (MEDITEC)**

Produced by FIZ Technik and distributed online by the producer and by Dimdi. Contains bibliographic references, with abstracts, to German and international literature on biomechanics, biophysics, biocybernetics, biomedical measurement, medical diagnostics and therapeutics, artificial organs, etc. The data sources consist of journals, proceedings of meetings, books, reports and patents. Most of the abstracts are in German while the descriptors and titles are in both English and German.

### **MEDLINE**

The largest medical database is produced by the National Library of Medicine of Bethesda. It comprises the Index Medicus, Index to Dental Literature and the International Nursing Index. It contains bibliographic references to world biomedical literature, including research, clinical practice, odontoiatrics, nursing, administrative problems, policy, etc. References are drawn from some 3200 specialized journals published in the United States and in 70 other countries. Also included are chapters and articles taken from selected monographs up to 1981. For 60% of the references the authors' abstracts are from selected monographs up to 1981. For 60% of the references the authors' abstracts are available (about 250 words, starting from 1975), which are not supplied in the hardcopy version. Terms are indexed using the controlled dictionary of the Index Medicus, Medical Subject Headings (MeSH), which is also available online. MeSH descriptors may be sought directly as terms, or as codes for searching the sections of the tree structure of this thesaurus. Each month about 20000 new records are added to the database. There is an approximately 40% overlap with Embase information. The database is distributed by BRS Information Technologies, Dialog Information Services, Inc., Data-Star, Dimdi, Mead Data Central, Inc., NLM and Telesystemes Questel. It is available also on the facilitated Easynet, Dialog Medical Connection, BRS/Colleague and Dimdi services.

### **Scientific American Medicine (SAMM)**

Produced by Scientific American, Inc., contains the full-text of reports providing information on current developments in clinical medicine, broken down in 15 subspecialties of internal medicine: cardiovascular medicine;

endocrinology; metabolism, respiratory medicine, etc. Distributed online by BRS and BRS/Colleague. Corresponds to the print publication *Scientific American Medicine* (monthly), which is also available in microform.

## **SCISEARCH**

This is a multidisciplinary bibliographic database with several characteristics which distinguish it radically from the others. It is particularly useful for a number of analyses and evaluations of world scientific literature. The main difference lies in the fact that each document is accompanied by a list of its bibliographic references. It is thus possible to know whether an article or an author has been cited in subsequent publications and, if so, how often. It is also possible to know the total number of times a journal has been cited and to select the publications on the basis of the number of bibliographic references.

The information it contains is drawn from a breakdown of some 4500 journals. The topics treated include life sciences, physical sciences, chemistry, earth sciences, agriculture, biology, environmental science, biomedicine, internal medicine, engineering, technology, applied science. In particular it is an excellent source for research on new technologies in the biomedical, pharmaceutical and bioengineering fields.

The database is produced by the Institute for Scientific Information (ISI) and distributed commercially by the vendors Data-Star, Dialog, Dimdi and Orbit. It can be accessed also through the Easynet system and Dialog Medical Connection. The equivalent hardcopy publications are *Science Citation Index* and *Current Content*.

## **New England Journal of Medicine**

Produced by Massachusetts Medical Society, Medical Publishing Group, it contains the complete text, from 1983 to the present, of the *New England Journal of Medicine*, a weekly journal covering current research in all areas of medicine. It is a subfile of *Comprehensive Core Medical Library*, available online through BRS and BRS/Colleague.

## **Microcomputer Index**

Produced by Learned Information, Inc., it contains references and abstracts of articles appearing in specialized microcomputer and computer journals. Includes also software reviews. Online availability through Dialog and Dialog Medical Connection. Also available in batch access.

## **Pacing and cardiac electrophysiology retrieval system (PACERS)**

This database is produced by Intermedics, Inc. and it is available in batch access

through the producer who offers search services without charge to the medical community. Contains bibliographic references to journal literature on artificial cardiac pacing published since 1966. Consists primarily of all relevant references indexed in MEDLINE database which are reindexed and supplemented with subject descriptors and other information.

### **Doctors facts**

Produced by Melissa Data Company (USA), it contains 390 statistical data items, from 1975 to 1983, describing the health and medical profession for all USA counties. Covers medical schools and medical professionals. It is available for purchase in diskette which operates on IBM and compatible personal computers.

## **6. The document delivery**

One facility offered by the hosts is that of ordering directly by terminal photocopies of the articles retrieved while searching bibliographic databases. Practically all the hosts provide this kind of service, the best known being *Dialorder* of Dialog, *Primordial* of the European Space Agency, *ADRS*-Automatic Document Request Service of Blaise. The host thus acts as intermediary by placing electronic orders on behalf of the user with the document delivery organizations.

Also database producers can sometimes act as document suppliers in the case of documents contained in their own archives. Significant examples of this are OATS of the Institute for Scientific Information, ERIC, Educational Resources Information Center and CDST (Centre de Documentation Scientifique et Technique) of the French CNRS which, as producer of the interdisciplinary database PASCAL, supplies primary documents both in hardcopy form and on microfiches.

In order to make the service more efficient, a number of consortia have been set up among libraries as well as several automated library networks.

In the medical sector, one example of this is the *DOCLINE* system of the National Library of Medicine of Bethesda. This is an automated library lending system linking up some 25000 libraries participating in the project both as users and suppliers of documents. Since more than 90% of the demand is for periodical articles, Docline has instituted an automated link with the Medline database. Using this link and a unique citation number (UI) the bibliographic data referring to the periodical can be identified and the order sent to the library closest to the requester's address. Documents are then delivered by post. In Europe the Docline document supplier is the British Library Document Supply Centre.

As the online information market has expanded in recent years, other more commercial information brokerage services have grown up side by side with the

traditional document suppliers such as libraries and producers. The US *Information on Demand (IOD)*, in particular the exclusive supplier of Easynet, is one example of this. The service is based on a network of professionals located in selected libraries in the USA.

Several publishers have also begun to try out electronic publishing techniques and to use optical memory storage. One significant example is that of the European *ADONIS* project. Several surveys carried out starting in 1980 had in fact pointed to the existence of very strong world demand for biomedical articles, in particular those written in the preceding three years. The Adonis project use CD ROMs to supply selected document supply centres with the full text and illustrations of about 300 European biomedical journals.

## **7. Search services and aids**

In information searches carried out, for instance, on factual databases, it is important not only to retrieve the data but it is often of great assistance to the user to be able to compare the data retrieved with others or to process them using statistical analysis techniques. For this purpose the hosts provide the possibility of using a suite of programs which allow analyses and comparisons to be carried out which would be impossible to do manually. These programs allow the user to download data and to create his own files using the data thus retrieved.

The distributors often sell their databases and programs on other media such as floppy disks, magnetic tape and CD ROMs.

## **8. Cost of accessing data banks**

The costs to the user of accessing a bibliographic database are made up of the following components:

- Subscription or guaranteed minimum;
- Database connect time and display/print charges;
- Telecommunications costs.

### *Subscription or guaranteed minimum for the producer or host*

The majority of hosts of bibliographic databases request no subscription payment or minimum guaranteed use. Current practice is to bill on the basis of hourly consumption.

Recently several hosts introduced a new form of billing based exclusively on the volume of information actually retrieved by the user. One example of this trend is the European Space Agency at Frascati and by the German host Dimdi, both of which have introduced a new system known as 'pricing for information' which is thus independent of searching time.

Sometimes access to certain databases is reserved to subscribers to the hardcopy version or to other services provided by the producer;

### *Cost of access time and display/printout of results*

The cost of access time varies from one host to another and from one database to another. Most hosts offer a number of options which impose various conditions on the user in exchange for a reduction in access time. Online and offline printout costs vary according to the number of paragraphs or fields printed (e.g. printout of only bibliographic references, title, complete record, etc.) and the cost of offline printing out may include an additional cost per page or a supplement for postal charges;

### *Telecommunications costs*

In addition to query charges, the user must also pay telecommunications charges, which vary according to the type of telecommunications network used. There are several different kinds of network, although the main procedures used for accessing databases online are as follows:

*a) Through the switched public telephone network.* The user dials into the host computer and uses a modem to set up a telephone link over which to transmit the data. This type of solution costs the price of a telephone call and varies according to distance and call time. It is therefore not to be recommended for accessing distant hosts;

*b) Through a dedicated line.* A fixed direct link is set up between the user and a specific host. This solution is more economical for users needing long and frequent connections to the same host;

*c) Through a packet switched network.* Based on a technique in which a DTE (Data Terminal Equipment terminal or computer) linked to the network establishes a logical communication link with its counterpart (in this case a host computer) for the purpose of exchanging data transmitted in the form of 'packets' or blocks of characters. The user's messages are sent over the network in a large number of packets, the maximum length of which is fixed in advance and each of which is associated with the destination address. Since these networks have been set up specifically for the purpose of data transmission both the performance and the cost are much higher than for the other methods.

## **9. CD ROMs**

Optical disk technology has been acknowledged as extremely effective for both memorizing and retrieving information. A CD ROM has a storage capacity



equivalent to about 300000 typewritten pages.

With the advent of this new medium many online databases are now produced also in a CD ROM version and some 500 products are already on the market.

The minimum equipment required for data retrieval is a microcomputer, a printer and a CD ROM player. The search software is supplied on the CD ROM together with the database. Although not yet standardized, this software retains the characteristics of normal information retrieval systems. The subsequent addition of special menus and tutorials makes them even more user friendly and, unlike online database searching, tends to cut out the intermediary.

The only cost in addition to that of the initial outlay for the CD ROM player is the annual subscription and for updates (periodically delivered to the user). Unlike online searching, the totally local management of the CD ROM requires no connection time, telecommunications networks and thus no unexpected expense. This makes CD ROMs a valid alternative to online searching also in the developing countries where telecommunications networks often still do not exist.

In the biomedical sector several on-disc versions of the most important databases are already available. It is therefore always important to know the exact performance of the various different kinds of CD retrieval software. With regard to this problem we assist a very fast succession of new versions of CD software. In particular it is useful to know: retrieval speed, whether all fields are indexed and searchable, whether the thesaurus structure can be displayed, whether the search is menu or command driven, whether the search strategy can be saved for subsequent execution in the corresponding online database, the time coverage, etc. A few examples of databases available on CD ROM are:

*Dialog on-disc Medline* – produced by Dialog Information Services; the query software is the same as the online version and is integrated with it in such a way as to allow subscribers to the online service to switch from one system to the other at any time. This ensures that complete and up-to-date information is always available. The data have been available since 1987. The thesaurus can be displayed and the user can switch between command and menu mode.

*Bibliomed* – produced by Digital Diagnostics, Inc.; a subset of Medline with abstracts of the 450 most important biomedical journals selected from a shortlist prepared by universities and medical centres. It only allows searches on titles, sources, authors and MeSH fields. The search profile cannot be saved. It is therefore not suitable for very complex searches. It is available for 1984 on.

*Compact Cambridge Medline* – produced by Cambridge Scientific Abstracts. It is equivalent to Medline in content.

*Comprehensive Medline/EBSCO CD ROM* – produced by Ebsco Electronic Information. It incorporates Medline from 1986 on. Search software Ebsco-CD.

*Core Medline/Ebsco CD ROM* – produced by Ebsco Electronics. It is a subset of the previous two years of Medline as well as of the current year. It

comprises more than 560 reviews and more than 330000 references. Search software: Ebsco-CD.

*Medline and Medline Compact Med-base* – produced by Online Research Systems, Inc. It contains references, starting from 1966. Search software: CD-Plus and CD-Plus.

*Medline Knowledge Finder* – produced by Aries Systems Co. It contains bibliographic references to the current year as well as to the preceding four years. Search software: Knowledge Finder.

*Medline on SilverPlatter* – produced by SilverPlatter. It contains references going back to 1966 on 4 disks. Search software: SPIR.

*ClinMED-CD* – produced by SilverPlatter. It is a subset of Medline. The emphasis is placed on clinical medicine and it comprises 320 journals listed in the Abridged Index Medicus as well as the Brandonhill List from the Bulletin of the Medical Library Association and the Library for Internists List. It contains 10 titles from the Year Book Series, including: anaesthesia, cardiology, etc. Search software: SPIR.

*Excerpta Medica Library Service* – produced by SilverPlatter. It contains references to the latest year of the equivalent hardcopy version and is available solely for subscribers to the latter. Search software: SPIR. It offers many help screens and search samples. Tutorial screens are available at the press of a help key. Malimet is not available on disc.

*BRS/Colleague Disc* – produced by BRS Information Technologies, Inc., with help screens and tutorials. Search software: BRS/Search.

*ISI CD* – produced by the Institute for Scientific Information, it is equivalent in content to Scisearch database.

## 10. Conclusions

With respect to the past each user can now choose from among a much wider range of information accessing methods. In addition to the traditional hardcopy, the media now available include online databases, floppy disks and, for the past few years, also optical media such as CD ROMs.

However, current CD ROM technology has a number of drawbacks. While its capacity to store large quantities of data makes it an excellent tool for resources with a low updating frequency such as catalogues, encyclopaedias, handbooks, guides, etc., updating becomes a problem when the information to be retrieved must be updated frequently. Another problem is that the user interface systems currently available are still rather inadequate. Retrieval performances are much lower than in online searching, particularly in the case of Boolean searches with a combination of large sets. Lastly the still high costs of subscription to a single database mean that the user pays a lot for a single resource, while online connection affords him the possibility of exploiting numerous resources simultaneously.

However, while it is true that the CD ROM cannot replace the online services,

at least for the time being, it certainly is a very important complement to online services and to printed books.

## **Bibliography**

1. Giorgi M, Libutti L (1987) Online databases for biomedicine. *Acta Medica Romana* 25: 390-410
2. Giorgi M, Libutti L, Valente A (1988) Basi dati in linea per la Medicina. Atti della sessione tutorials, APIM '88 - Attualità e Prospettive dell'Informatica in Medicina, Firenze, pp 109-125
3. Bianchi G, Giorgi M, Libutti L, Valente A (1988) Full-text databases and biomedical Libraries. Proceedings of the Second European Conference on Medical Libraries, Bologna, November 2-6, 1988, München, New York, London, Paris, KG SAUR
4. Basch R (1989) The seven deadly sins of full-text searching. *Database* 15-23
5. Tenopir C (1985) Systems for end-users: are there end-users for the systems? In *Library Journal* 15: 40-41
6. Bonham MD, Nelson LL (1988) An evaluation of four end-user systems for searching Medline. *Bulletin of the Medical Library Association* 76: 22-23
7. Bonham MD, Nelson LL (1988) An evaluation of four end-user systems for searching Medline. *Bulletin of the Medical Library Association* 76: 171-178
8. Stern B (1986) ADONIS - publishing on CD-ROM in mixed mode. In 10th International Online meeting, London, 2-4 December 1986, Oxford, *Learned Information* 23-31
9. Ogilvie-Parfrey W (1987) The Excerpta Medica Database on CD-Rom. In *Medical Libraries: Cooperation and new Technologies*, Elsevier Publishers B.V. 177-180
10. Bakker S, Bleeker A *et al* (1989) Medline on CD-Rom: a comparison. *Online Review* 13: 39-50

# Computer-aided clinical problem solving as an educational paradigm for teaching preclinical cardiac pathophysiology

ROBERT A. GREENES, BRYAN P. BERGERON, MARK S. DICHTER

*Decision Systems Group, Harvard Medical School, Brigham and Women's Hospital, Boston, Massachusetts, USA*

JOHN T. FALLON

*Shriners Burns Institute, Massachusetts General Hospital, Boston, Massachusetts, USA*

## **Abstract**

Clinical problem solving exercises have long been used as a way of providing simulated experience with clinical medicine. Multimedia capabilities now available extend the potential for realism of these exercises. In addition, with the use of hypermedia methodologies, it is now possible to link components of a clinical case – e.g., the tests, the images (or sounds or tracings) of the actual results, or the textual descriptions of findings – with other related information. This information can amplify on the tests themselves, the range of possible results that could alternatively have been obtained, the differential diagnosis of any particular finding, or the profiles of diseases in the differential, and can provide discussions and references about the diseases. Thus the case problem solving exercise may have potential for more than providing practice with making clinical judgments; it may offer an alternative paradigm for accessing much of the same medical content traditionally delivered in didactic lectures and in textbook presentations. We have implemented an authoring environment for case problem solving exercises that enables access to multimedia presentation of clinical results, and related clinical and pathophysiologic content material. Case simulations designed with this approach are currently being used in teaching first-year medical students in the HST track of the Harvard Medical School.

## **1. Learning by problem solving**

A far-reaching document on the future of medical education that was published by the American Association of Medical Colleges and known as the GPEP report (GPEP 1984) predicts that the activities required of physicians of the twenty-first century will bear little resemblance to those of physicians today. These physicians of the future will require a completely new set of skills, knowledge, and attitudes and will be heavily reliant on information technology. Learning how to learn – how to investigate a premise, associate ideas, analyze

new concepts, and synthesize information into a schema against which new data can be evaluated – will replace the current practice of simply learning innumerable facts. The validity of these projections is supported by the current state of medical practice, in which resources are no longer recognized as unlimited; where many priorities, some of them non-medical, compete for these resources; and where prepayment and other reimbursement mechanisms alter both patient and physician incentives – all of which increase the complexity of medical decision making (Eddy 1982).

Given the desirability of a problem solving-oriented rather than a memorization-based approach to medical education, it is useful to consider the implications on the nature of the preclinical medical curriculum. Clinical medicine remains largely a problem solving activity in which one learns by experience, but ideally this should follow upon preclinical training which provides a framework of essential skills, knowledge, and attitudes on which to build. An approach that appears to be particularly valuable in this regard is the ‘case presentation method,’ in which cases are constructed to address specific problems and have particular learning objectives. This approach has been the foundation of the New Pathway curriculum reform at Harvard, which began in 1985 (Tosteson 1990). The investigation of a case by students, individually or in groups, and the ensuing discussion under the guidance of a faculty tutor serve to provide a focus for both learning a variety of essential knowledge and gaining experience with the use of information sources, analytic tools, and other resources that can help in future problem solving.

We address here the use of the computer to facilitate the acquisition of problem solving and analytic skills and to serve as an information resource in the learning process, with particular focus on the case-based exercise. It should be noted that in general, an experienced and charismatic teacher can transfer both domain knowledge and an enthusiasm for learning in a way that no textbook, videotape, or computer program can. However, a variety of tools now available to educators can supplement the basic student-teacher interaction. We believe that these tools are particularly useful in the kind of learning by problem solving that is intrinsic to new curricula addressed to the needs of the future physician. These supplementary tools are also generally useful when there is a large student-to-teacher ratio (Novak and Glase 1987), when there is a vast amount of material to be learned in a short period of time, when the concepts involved are difficult for most students to grasp, or when independent learning is to be fostered.

## **2. Goals**

In this chapter we focus on the role of clinical problem solving exercises as a means of teaching cardiovascular pathophysiology to preclinical students of medicine. We use this as a paradigm for a general approach to preclinical computer-aided instruction (CAI) in medicine. The approach taken has the

following objectives:

- (1) augment didactic instruction of cardiovascular pathophysiology with clinical problem solving experiences;
- (2) use simulated cases to illustrate typical clinical entities;
- (3) present the problem solving process realistically, not only in terms of viewing results and making decisions, but by demonstrating the inconsistencies and uncertainties encountered, accounting for the passage of time, and enabling the case to progress through stages, of initial assessment, test selection, data gathering, interpretation of results, problem formulation, development of a plan, patient management, and follow-up;
- (4) provide opportunity to explore related topics, e.g., other diseases to consider, underlying pathophysiologic mechanisms, procedure details, workup strategy and decision-making issues, risk factors, prognosis, and correlations with pathology;
- (5) enable case complexity or emphasis to vary with the level of experience and focus of the student, so that the same basic case can be used as a recurring example for courses in pathophysiology, pharmacology, epidemiology, and so forth;
- (6) explore methods for evaluation of student experience, to identify the role of the case in the curriculum, and to determine its effectiveness and efficiency for learning.

Cardiovascular medicine lends itself well to this approach, in view of the highly developed ability to identify pathophysiologic bases for clinical phenomena, the extensive use of various technologies for imaging and assessment of cardiovascular function, the general importance of cardiovascular medicine, and the high level of interest that students have in this field.

In the next section we describe the two principal methods utilized in developing our clinical problem solving approach to teaching – hypermedia and simulation; in later sections of the chapter, we illustrate, by describing an application being used at Harvard Medical School, how hypermedia and simulation-based teaching can be combined in cardiovascular pathophysiology. We then discuss issues of evaluation and a number of possible extensions and ramifications of the approach.

### **3. Methods**

Predominant among the instructional methodologies that are suitable for providing clinical problem solving experiences is the use of a combination of hypermedia and simulation methods.

#### *3.1. Hypermedia*

Graphic environments on modern microcomputers allow multiple ‘windows’ of information to be displayed concurrently; thus the range of activities available

to the user, or the information that can be viewed simultaneously, is increased – a situation that facilitates the correlation of various items of data or the integration of different functions. A familiar example of the latter is the ability to concurrently view numeric tabulations of data in one window and graphic plots of the data in another, while composing a narrative discussion of the analysis for a manuscript.

One of the ways this multi-window capability can be exploited for teaching and learning is the development of enhanced ways to present data and knowledge to a user. Rather than relying only on sequential presentations of textbook material or tabulations of data (as would occur on paper or with single-window, character-based computer displays), graphically oriented windowing environments on microcomputers make it possible to select pertinent areas of interest in a display, usually with a pointing device such as a mouse, and to branch in a nonlinear fashion to other windows that amplify on, illustrate, or otherwise relate to those areas of interest. These windows may be viewed concurrently (within the limits of available screen size), facilitating comprehension and allowing the user to pursue his or her own pathway through the content material, depending on interest or need. When textual material is made available in this nonlinear fashion, it is referred to as *hypertext* (Conklin 1987; Nelson 1987); when images, animations, and other material are included, the term *hypermedia* is used (Yanlelovich *et al.* 1985).

In the last few years, hypermedia has become enormously popular as an approach to the design of computer-based instructional and other interactive applications that present options and allow the user to explore various pathways. This popularity has been stimulated by the appearance of several commercially available authoring software environments that facilitate the creation and use of hypermedia-like approaches – e.g., GUIDE® (Owl Technologies, Inc., Bellevue, WA), HyperCard® (Apple Computer, Inc., Cupertino, CA), and SuperCard® (Silicon Beach Software, San Diego, CA). At present there is considerable research and development to extend the capabilities of hypermedia in various directions, including the combination of hypermedia capabilities with artificial intelligence-based expert systems, addition of free-text information retrieval capabilities, and development of methods for converting traditional print materials (e.g., books and journals) to hypermedia format.

With respect to instruction, hypermedia techniques can be used to provide experiences based on either free-form exploration or more didactic predetermined branching, and to perform examination and testing. A principal virtue of hypermedia is the ease with which content experts (e.g., physicians) can use the authoring software to develop attractive and useful applications for presentation of their material without the aid of computer professionals.

### 3.2. Simulations

Patient case simulations have been shown to be engaging and have many

advantages over conventional teaching modalities. The advantage of simulation as an instructional mode is that it facilitates individualized experimentation, exploration, and problem solving by the student. A range of possible alternatives may be investigated and their consequences evaluated without the limitations of inaccessibility or the inefficiency of performing these investigations in real life. The potential value of this capability is growing as modern microcomputer systems permit graphic, multimedia experiences to be portrayed, with realistic sound, images, and animation.

*3.2.1. Roles and uses of simulations.* Simulation can play a variety of useful roles, for example, to help medical students understand and analyze complex processes, acquire new skills, conduct experiments, and reinforce other learning experiences. By helping students visualize relationships, simulation-based courseware can make learning both more efficient and enjoyable (Piestrup 1984). Simulation-based applications can be used to supplement and, in some instances, replace conventional learning experiences. For example, patient case simulations can minimize the need for a real laboratory or a real patient, reduce student inhibitions, and permit a wider range of experiences (Bergeron and Greenes 1987; Friedman and Schwartz 1987; Gardner 1987; Mackenzie 1987).

Graphic computer simulations are extremely effective tools for teaching and manipulating ideas that can be readily conceptualized using visual paradigms. These visual paradigms serve as mental models, which offer the benefits of extending the student's memory capacity while presenting the significant information in a particularly usable form, thus fostering a deeper understanding of the situation (Fundt 1981; Meyer 1981). Research also suggests that simulated experiences are far superior to the learning of rules, especially during the early phases of learning (Ross 1982). Simulations can provide experimental data in cases where it is either impossible or unethical to perform the actual experiment on real patients (Cobelli and Mari 1985; Davis and Jack 1986; Krenz 1985; Nandedkar 1985; Salzsieder 1985; Starmer and Kerr 1985; Sun 1985; Van den berg 1985). The time compression afforded by simulations enables one to more easily trace the time course of a disease or other process under study (Gabay 1983). Furthermore, results of simulated experiments can suggest promising experiments on real patients (Steinhous *et al.* 1985).

Knowledge acquisition is enhanced through more active learning opportunities, rather than, or in addition to, more traditional didactic material in texts, journals, and classroom lectures. A simulation-based environment is ideal for learning concepts relevant to clinical medicine and eliminates those characteristics that reduce the value of the laboratory or the patient's bedside as a learning environment (e.g., distractions, expense, time limitations, poor patient cooperation, and the limited range of actual experiences available to the student).

Simulations also can perform a decision support role by providing a means of performing 'what-if' and sensitivity analyses (Pollack and Greenes 1985). In



addition, since simulations are in effect highly compressed databases, they can assist the clinical decision-making process by providing a ready reference of clinical problems, patient presentations, and pathophysiologic mechanisms (Van den berg 1985).

Simulations are also of use as part of an overall test of medical competence, since they can approximate the kinds of decision making that physicians will face and can aid in the assessment of clinical skills and judgment rather than just recall of factual knowledge (Clyman 1990).

Simulation methods can be valuable research tools that allow investigation of complex processes and allow the investigator to specify the values of parameters and view the resultant simulation. Often, simulation is the only technique that exists to explore models of complex, interconnected, and irregular biologic systems in a reasonable amount of time.

**3.2.2. *Obstacles to universal use.*** Due to the relatively large time investment required to implement systems that entail complex, interactive graphic simulations, it is not surprising that the great majority of computer-based instructional systems in medicine consist of text-only lessons and case scenarios and that until recently only a few incorporated sophisticated graphics (Barnett 1984; Beck 1983; Greenes 1983; Hoffer and Barnett 1975; Pollack and Greenes 1985; Smyth-Staruch and Littenberg 1983; Stulurow and Cochran 1983). Progress in reducing this time requirement has come about through the development of high-level authoring tools and shells that simplify and speed the coding of the nonspecific aspects of simulations and the availability of hypermedia capabilities (as discussed in Section 3.1), which can be combined with simulations.

**3.2.3. *Considerations in simulation design.*** For a simulation to be effective, its design must encapsulate the subject material (e.g., cardiac pathophysiology) and present it in a manner consistent with the student's expertise both in the subject domain and with the computer system and with the student's unique cognitive style and goals. Some of the highly interdependent factors that are generally recognized as having a direct bearing on the quality and effectiveness of computer-based simulations include the following:

(1) Choice of metaphor or model. Students develop their own internal or mental models of the simulated system; these models are derived from the images provided by the application and help in learning (Norman 1983). For example, a student learning the basics of arterial wall compliance might use a mental model based on an electrical resistor-capacitor network. The 'naturalness' of a model may be due in part to our common use of concrete objects in reasoning about abstract systems (Borning 1979; Norman 1984). For example, cardiac vasculature can be represented by rubber tubing – something that students can relate to in their everyday lives.

(2) Modes of interaction. The keyboard and pointing devices are the primary available means of conveying information *to* an application by the user,

whereas the graphics display, supplemented by sound, provides the primary information channel *from* an application to the user. Use of menu bars, icons, slide controls, and other ‘gadgets’ may be helpful for input, and visual and sound effects may be effective for output and reinforcement, but the cardinal rule is to make these choices based on pedagogical principle rather than ‘showmanship.’

(3) Power versus ease of learning. Graphics, on-line help, support for pointing devices, menu bars, dialog boxes, intuitive icons, informative error messages, and robust error handling are all components that contribute to an interface that is easy for students to learn. However, optimization for ease of learning may result in an application that is too slow to use regularly and too cumbersome for students to use productively, no matter how friendly. The optimal mix of power and ease of learning for a particular interface varies considerably among students and is influenced by the nature of the application. It also may vary over time, as experience is gained by the user, at which point ‘power user’ features, shortcuts, and more terse operation modes may be desirable.

(4) Use of color in graphic displays. The primary objective of the use of color in a program is to add meaning. Color is most effective when used to discriminate between different areas, show processes or applications that are functionally related, identify relationships, and identify crucial features.

(5) Role of sound. Like color, sound is used to add meaning. Sound can be integrated throughout a program to help make students aware of the state of the application. Similarly, sound can serve to alert students to unexpected events or the termination of a lengthy calculation or process.

(6) Appropriate level of graphic complexity. As the graphic complexity of a program increases to a level comparable to that of the real system, this may offset a major advantage of the simulation – that of focusing a student’s attention on the pertinent aspects of the system being simulated. Simulations that make use of simpler, less complex graphics will sometimes be easier for students to understand and use. On the other hand, the ability to increase the graphic complexity of a simulation can be an advantage, for example, if the intent is to stress students with an over-abundance of visual cues through simulation, as in testing.

(7) Modalities supported. Multi-modal simulations that are intended to increase a student’s skills in performing, say, a surgical procedure must contend with the complex problem of sensory mediation. There must be some way of providing students with an intuitive indication of force reflection, touch, proprioception, tissue compliance, and resolved rate.

(8) The need to provide situational awareness. Graphic simulations must at all times provide obvious clues – be they graphical, color, auditory, or otherwise – that inform students of where they are, where they can go, and where they have been.

(9) Characteristics of the hardware platform. The computer hardware is an integral part of any application, since changing the input device from a mouse,

trackball, or touch/pad to a light pen or substituting a color monitor for a monochrome one will have an enormous effect on the user interface.

### *3.3. Combining hypermedia and simulation methods to provide enhanced learning environments*

Hypermedia can be combined effectively with simulation presentations, since hypermedia makes it possible for large amounts of content to be brought to bear upon a problem. For educational purposes, hypermedia not only makes it possible to augment simulations with highly graphic visual information through still images, animation sequences, sounds, and graphic displays, but also allows the student to pursue other pathways, providing amplification of an experiment with discussions, examples, details, reference material, and so forth.

## **4. Previous related work in this laboratory**

The Decision Systems Group (DSG) is a research unit in Medical Informatics at Brigham and Women's Hospital and Harvard Medical School, the aim of which is to foster the use of computer systems to aid in medical decision making and education. Consisting of 17 computer scientists, physicians, and other staff, the DSG has been developing an extensive repertoire of tools and techniques for these purposes. Development activities include the following:

### *(1) Explorer-2*

An object-oriented 'knowledge management' authoring and presentation system for providing consistent access to a wide variety of forms of medical knowledge (Greenes 1986; Tarabar *et al.* 1989). Knowledge may be unstructured text, pictorial information (e.g., clinical algorithms, diagrams), structured information (e.g., disease, finding, or drug data bases), dynamic procedural knowledge (simulations, quantitative analyses, expert systems), or the documents in a document library (Greenes 1989). The knowledge is represented as a network of nodes and links, with each node's content under the control of a specific tool for producing and displaying it (e.g., a hypertext tool, an animation tool, a spreadsheet, a simulation, an expert system, or a data base retrieval tool). Explorer-2 provides a framework for browsing through the knowledge, for selecting particular nodes and links by use of keywords associated with each, and for retrieving or producing content and displaying it. Node content is displayed in multiple windows. If a particular term or topic of interest is chosen by selecting a 'hotspot' in a window, the user is able to branch to other windows that contain additional data. An overview map provides a navigational aid and a means for tailoring and saving personalized pathways through the knowledge for subsequent use.

Special features of Explorer-2 include the ability to present information from a textbook in an easy-to-use fashion (by expanding and collapsing topic areas), without the need for major reformatting of the original content; the ability to present medical images such as roentgenograms or pathology slides in full resolution; and the ability to incorporate various kinds of knowledge, both passive and dynamic (Greenes *et al.* 1989a). We have currently redesigned our knowledge management environment around a kernel set of 'building block' components, known as DeSyGNER (Decision Systems Group Nucleus of Extensible Resources), to facilitate customization and extensibility.

### *(2) Information retrieval by use of a Unified Medical Language System (UMLS)*

This project is a development effort of the National Library of Medicine (NLM) in which the DSG is participating. The goals of this project are to provide a consistent method for indexing and retrieving medical knowledge (Humphreys and Lindberg 1989). This framework is needed to more readily interface with hitherto independent applications and to facilitate the development and use of electronic knowledge resources and personal reference files. The UMLS taxonomy provides a basis for the indexed retrieval of knowledge by Explorer-2 (Komorowski *et al.* 1988). A natural language query system, with particular focus on its use in cardiovascular medicine, is being developed by William Hersh, M.D., a postdoctoral research fellow of the DSG.

### *(3) Explorer-2 knowledge bases*

'Knowledge bases' may be defined as collections of content that pertain to a particular subject domain. One of these knowledge bases, CASPER (Computer Aided Selection of Procedures and Evaluation of Results) (Greenes *et al.* 1989b), is a compendium of diagnostic strategies for a wide variety of common clinical problems based, in part, on a handbook of diagnostic strategy assembled by the Brigham and Women's Hospital Department of Radiology. CASPER gives information about general patient characteristics for a specific clinical problem, the approach to workup, a flow chart of the clinical algorithm, and data about the various tests available, including preparation requirements, advantages and limitations, and quantitative data about test performance. In addition, 'what-if...' procedures can be invoked to analyze the value of a proposed test for a specific patient. QMR<sup>®</sup> (University of Pittsburgh) (Miller *et al.* 1986) is a knowledge base and expert system for differential diagnosis that has been adapted by the DSG for use within the framework of Explorer-2. (Details of QMR and its knowledge base were made available to us through a special research arrangement with, and the generous cooperation of, R. Miller and colleagues at the University of Pittsburgh.) Other knowledge bases include adaptations of chapters of major medical textbooks, with cooperation of the publishers, and several specialized databases.

#### *(4) Multimedia content management tools*

These tools are being developed to support the organization, indexing, and sharing of high-resolution video, graphics, sound, animated sequences, simulations, and text, both within and among institutions. Our present system is designed to handle digitized radiographs, images of gross and microscopic pathology specimens, illustrations, and graphs within the cardiovascular domain. By initially working primarily with digital images, we are addressing many of the indexing, image processing, and network bottlenecks that generally occur when working with other multimedia content.

#### *(5) Simulations and tutorials for skillbuilding*

At present these materials are used to give physicians and students experience in understanding the complex interrelationships that form the basis for clinical judgment and decision making. The DSG has addressed this task in several areas: cardiac auscultation, electrocardiograph interpretation, and lung auscultation. HeartLab (Bergeron 1987), for example, which is in wide use throughout the United States and elsewhere, gives the student experience in detecting and recognizing heart sounds by simulated auscultation – first, by allowing user selection of sound characteristics and abnormalities and second, by generating unknown cases for user interpretation. Review material is made available to augment the exercises. EKGLab (Bergeron and Greenes 1987) provides similar experience in electrocardiographic analysis. An arrhythmia tutorial system, with high-quality graphic presentations of rhythm tracings and diagrams, has been developed by Robert McClure, M.D., a postdoctoral research fellow in the DSG, as an adaptation of a previous program in which the content had been largely paper-based. Several simulations of probabilistic decision-making concepts such as sensitivity, specificity, predictive value, Bayes's theorem, and Receiver Operator Curve (ROC) analysis have also been developed (Pollack and Greenes 1986).

#### *(6) Pattern recognition skills in radiology*

These skills are being enhanced through use of a computer-based digital imaging system. The system is aimed at providing users with experience in recognizing pertinent image features and in combining them appropriately into diagnostic categories. This work is being carried out in collaboration with John Swets, Ph.D., David Getty, Ph.D, and colleagues at BBN Laboratories, Inc., who developed a feature cluster model that was shown to be very successful (Getty *et al.* 1988).

#### *(7) A dynamic algorithm display generator*

This tool is being used to encode medical logic and to provide educational and

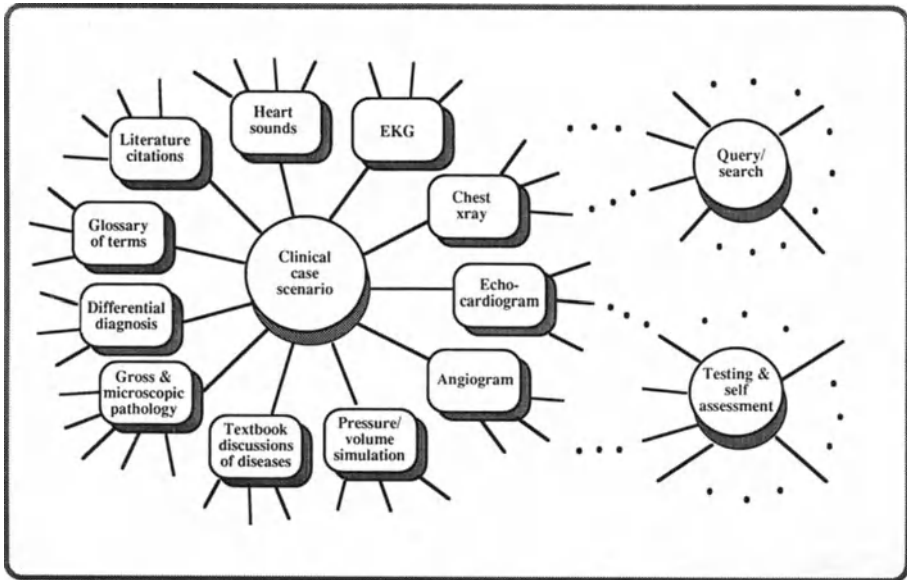
decision support (Abendroth and Greenes 1989). A wide variety of clinical problem solving tasks are described with clinical algorithms. The computer tool overcomes the limitations of paper-based algorithms by permitting arbitrarily complex algorithms to be dynamically generated and displayed in flow chart form, depicting only those branches pertinent to a specific situation. Prior data cause certain branches to be taken automatically and their logic collapsed into summary nodes. Sections of algorithms can be represented by single boxes that can be expanded if the user wishes to see the details and can be linked to other knowledge resources. The DSG is also exploring the use of clinical algorithms as an organizing strategy for an electronic knowledge resource for primary care and/or emergency medicine.

## **5. Cardiovascular problem solving exercises**

We utilize a combination of the above resources, others we can obtain elsewhere, and still others developed as needed, to design clinical case presentations that give the student an opportunity to explore realistic workup strategies, inspect results, determine differential diagnoses, treat the patient, and follow the patient long term. The clinical problem solving experience is made realistic through the accessibility of actual examination data such as heart sounds, roentgenograms, and electrocardiographic tracings. Explanatory material relates procedures, findings, and diagnoses to their pathophysiologic bases, and differential diagnostic aids and other decision aids are provided where appropriate. The collection of knowledge resources available to the case is depicted in Figure 1.

The approach we describe is currently being used in a first-year medical student course in cardiovascular pathophysiology. This course is given to a special track of the Harvard Medical School (HMS) class known as the Health Science and Technology (HST) track, which includes approximately 35 students per year (approximately 20% of the entire HMS class). The HST track, almost entirely separate from that of the rest of the HMS curriculum for the first two years of medical school, is taught by faculty who have appointments in the HST Division, a joint academic division of both HMS and Massachusetts Institute of Technology. In general, HST students are somewhat more technically oriented and interested in research than are their other HMS classmates.

Case presentations have been part of the HST 090 course for a number of years. The first half of the course is a didactic and laboratory-oriented introduction to quantitative pathophysiology. In the second half of the course, the students work through three case presentations, both out of class and in tutorial discussions. Each case is studied for approximately 10 days. In the past the cases were all presented in the form of paper-based handouts. The different stages of the case were extended over several days to allow the students time to find relevant background information on their own. Many images referred to in the handouts were not included, or were only available in the form of poorly



*Figure 1.* Integration of a variety of different kinds of knowledge resources (both content and presentation tools) around a clinical case scenario. Also depicted is the ability to ‘repurpose’ the individual knowledge resources for other user interaction environments, such as query and search or testing and self-assessment.

photocopied images. Of course, neither sounds nor any dynamic information were included.

We have created a computer-based system (Dichter *et al.* 1990) that supports a structured combination of simulations, hypermedia, and testing and that enables the students to have instant access to relevant data and knowledge. They can see images of test results in full resolution, including color or animation where appropriate, and they can inspect examples of normal test results. They can explore implications of various findings, differential diagnoses, alternative presentations, and the pathophysiologic bases for various tests and procedures. Our aim is to create a highly engaging exercise in which both directed instruction and exploration combine to impart an educationally effective learning experience.

### *5.1. Design considerations*

We use high-resolution, realistic renditions of the clinical findings by providing a multi-media graphic and auditory experience that simulates the live clinical situation. As we have continued to develop and enhance our own knowledge management environment, we have used a variety of other courseware development tools for prototyping or for gaining access to content material provided to us by various collaborators, including HyperCard, GUIDE, and SuperCard. We have also used content material from other sources to which we

have access – the University of Utah’s ‘Slice of Life’ videodisc (Stensaas and Sorenson 1988), QMR (Miller *et al.* 1986), PathMac (Stofer 1989), and HeartLab and EKGLab (Bergeron and Greenes 1989). Our work is facilitated by the availability of affordable content acquisition hardware described in Section 5.3. The prototype case described here is currently implemented in SuperCard.

The case is divided into stages. As the student progresses through a stage, selecting tests or treatments as appropriate, examining results or consequences, and considering related information, he or she can progress to the next stage only by updating a list of working hypotheses. Appropriate feedback is provided for the hypotheses chosen, and tests are made available on the basis of these hypotheses.

The tests and results pertaining to the patient are the triggers for accessing related information about other tests, other possible results, diseases that should be considered, and discussions of both tests and disease processes. A glossary of terms is also provided, as well as references for all material. The general structure of the case problem solving exercise is depicted in Figure 2.

The material used to construct the case is indexed and organized in a database so that it can be accessed where appropriate, possibly in multiple contexts.

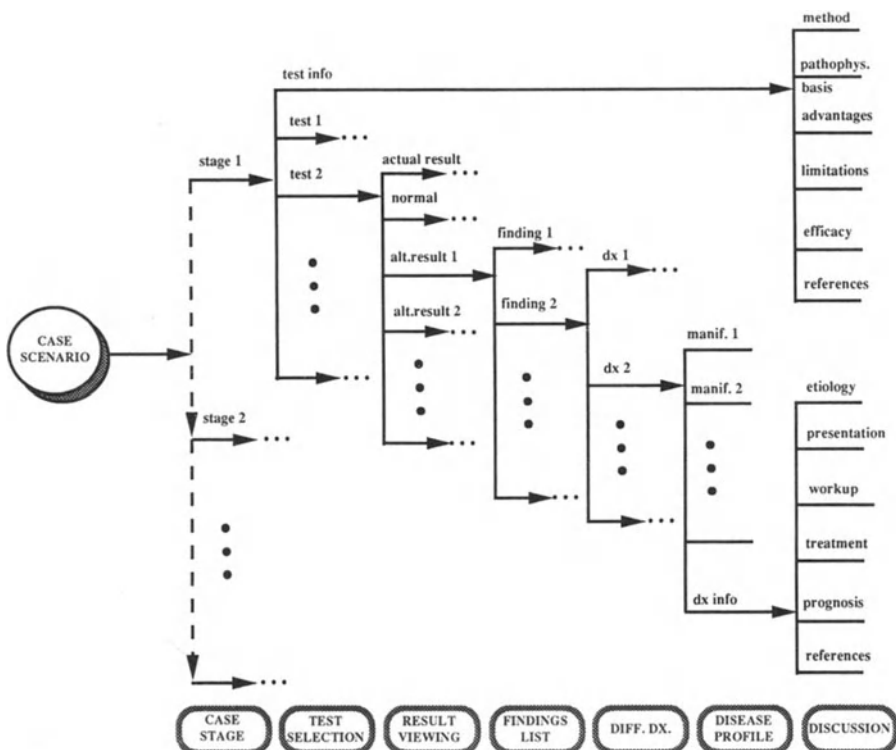


Figure 2. Structural organization of medical content for case problem solving exercise.



Quizzes, for example, can present material (e.g., textual, visual, or auditory) from the case itself or from the collection of related information in questions or in discussions of answers, thus ‘repurposing’ this material as needed (Figure 1). Although recent technologic advances in computer hardware and software facilitate the presentation of multimedia courseware, the acquisition and management of the content required to create such courseware remain complex knowledge management problems. We believe that the traditional approach of embedding content within an application and using nonstandard formatting and indexing schemes has no place in modern multimedia development. Given the effort and resource requirements associated with the development of multimedia courseware, there must be provision for using content from multiple sources. In addition, the content must be maintained in a form such that it can be repurposed as needed in other courseware.

### *5.2. Major features of the authoring and presentation environment*

In designing our initial case, we have built a ‘generic’ case presentation shell (which in this instance is configured to teach cardiac pathophysiology). In so doing, we established a list of desiderata that guided our implementation:

(1) *Multimedia content.* The case presentation shell supports a variety of multimedia content types, including digitized sounds, images, animation sequences, simulations, and hypertext.

(2) *Structural constraints.* To enable incremental testing we provide the ability to impose constraints on access to the content. For example, students must provide a correct differential diagnosis or management decision before they are allowed to progress from one stage to another within the case.

(3) *Provision for student tracking.* We record when students invoke particular functions, how long they study images, and which tests were ordered, to aid in evaluation both of the features supported by the software and of student responses.

(4) *Context-sensitive help.* On-line, context-sensitive help is provided, not only to minimize student frustration but to decrease the demand for supplemental, external resources, including faculty time and written manuals.

(5) *‘Hyperlinking.’* All significant terms in the textual descriptions of the patient case, including clinical findings and diagnostic tests, are linked to definitions. Other material, such as text discussion, disease profile (if applicable), and reference citations, may be accessed from the definition or from other contexts. For example, the term *dyspnea* in the description of the patient’s history is linked to a definition of dyspnea; a discussion (derived from HST faculty and textbooks) of the etiology, pathology, and treatment of

dyspnea; and a profile of dyspnea, including a list of associated diseases. Both the discussion and profile are linked in turn to reference citations. In addition, each term in the dyspnea profile is linked to a list of associated differential diagnoses. Each item in the differential diagnosis list is in turn linked to a definition and discussion.

In addition to hypertext linking, our case presentation environment supports linking to and from pictures, sounds, animated sequences, and simulations. For example, the chest roentgenogram taken upon admission is linked to a normal chest roentgenographic study as well as to a variety of chest roentgenographic studies with alternative findings. Similarly, a student studying the patient's auscultation findings has the option (by activating the multimedia links) of examining normal heart sounds and abnormal heart sounds related to a variety of cardiac abnormalities.

(6) *Display independence.* The relative proportions of the graphic interface and the size of the images displayed are automatically modified to suit the particular video display used. This allows us to deploy systems with either standard 13-inch color monitors or larger, 19-inch color monitors without changes in software.

(7) *Reusable content libraries.* Images, sounds, and other multimedia content are maintained in a database separate from the courseware system, as described in Section 5.1. While the files may be on the same physical disk as the courseware system, they can also reside on a server, linked to the student workstation via Ethernet. A central repository for images and other data minimizes local disk storage requirements, provides for efficient database maintenance, and supports content repurposing.

(8) *Multiple, simultaneous image display capability.* Since images generally provide more information when they can be viewed in relation to other images (e.g., as in comparing an admission chest roentgenographic study with a postoperative film), our system is capable of displaying multiple images simultaneously.

(9) *State memory.* Students may exit the courseware environment at any time without fear of losing their place, since the system state is automatically restored with subsequent log-ins. For example, a student may exit the system while examining a digitized pathology image to discuss the image with his or her instructor and return to that state upon reentering the system.

(10) *Graphic overlays.* Image-based question-and-answer sequences are simplified by the ability to use graphic overlays to identify relevant components of an image.

### 5.3. Content acquisition

Content for the cardiac pathophysiology materials embedded in the courseware was obtained from a variety of sources. For example, textual definitions and descriptions were distilled from textbooks and supplemented by HST faculty. Disease profiles were obtained from textbooks and from Miller's knowledge-based system, QMR (Miller *et al.* 1986). Sound information was digitized with the aid of an inexpensive audio digitizer unit, MacRecorder (Farallon Computing, Berkeley, CA). High-resolution color and grey-scale images were captured from NTSC videodisc and videotape units with a ColorSpace II framegrabber (Mass Microsystems, Sunnyvale, CA) and from 35-mm slides with a BarneyScan slide digitizer (Barneyscan Corporation, Berkeley, CA).

By capturing and archiving images in digital form, we are able to simultaneously display several different images. In comparison, a videodisc-based system allows the display of only one frame at a time. The use of digitized images makes it possible to use a central image server that is networked to workstations. Because there is currently no economical, effective means of networking videodisc players, our use of digitized images frees us from the expense of purchasing multiple copies of each videodisc title used and the requirement for one videodisc player per workstation.

### 5.4. Structure of the case

The case study, which concerns an elderly electrician admitted with a chief complaint of dyspnea during exercise, attempts both to introduce students to clinical medicine and to relate the clinical material to basic principles of cardiac pathophysiology. In the first stage of the case, students have access to the patient's history, physical exam findings, digitized radiographs, cardiac auscultation sounds, EKG tracings, and so forth. All important terms in the textual description of the patient have links to definitions of the terms.

When the student requests an examination (e.g., cardiac auscultation or a chest roentgenographic study), the appropriate result is provided (e.g., heart sound, or PA and lateral chest film images). Regions of interest may be determined by the student by inspection; alternatively, the actual findings may be revealed. The latter may be done by moving the cursor over the image to reveal overlays and labels that indicate the findings or by inspection of a textual list of findings, in which case selection of a finding causes any associated overlays and labels to be displayed.

For any finding the student may also request a list of candidate diseases to consider for that finding. For any of the diseases so indicated, the student may then access a disease profile that indicates the set of findings to be expected for that disease and a textbook discussion of the disease that describes its etiology, presentation, workup, treatment, and prognosis (Figure 3). For any test, the student may access for comparison an example of a normal result (e.g., normal heart sounds or normal chest roentgenographic study) or a variety of other

File Edit Differential Dx Help Comment

The patient was a 72 year old white male electrician who began to experience episodes of dyspnea and coughing during exercise six months prior to hospital admission. The patient was totally asymptomatic until these episodes began. There was no past history of hypertension, diabetes mellitus, pulmonary disease or rheumatic fever. He had a 75 pack year history of smoking. At age 68, during a routine physical examination, he was noted to have a murmur. Because of the increasing frequency of his episodes of dyspnea, the patient was admitted to the hospital for evaluation.

History  
 Admitting History  
 Physical  
 Heart  
 Auscultation  
 Radiology  
 Admission CXR-PA  
 Admission CXR-Lateral  
 Physiological Testing  
 Electrocardiogram

**Profile**

Diagnosis: Hypertrophic Subaortic Stenosis **Info**

Medical History and Current Symptoms  
 Age 26 to 55  
 Male  
 Physical Exam Findings:  
 Cardiac Auscultation:  
 Gallop R  
 Systolic  
 Holosyst  
 Systolic  
 Early sy

**Discussion**

**Term: Hypertrophic Subaortic Stenosis Info**

This disorder is characterized by cardiac enlargement with myocardial hypertrophy. The left ventricle is usually more involved than the right, a

**Differential Diagnosis**

**Term: Left Ventricular Hypertrophy Info**

Sympto  
 reduced  
 capacity  
 empty

Hypertensive Heart Disease  
 Aortic valvular stenosis, rheumatic  
 Aortic valvular stenosis, bicuspid valve  
 Aortic valvular stenosis, calcific  
 Aortic regurgitation  
 Cardiomyopathy, idiopathic dilated  
 Cardiomyopathy, restrictive  
 Hypertrophic Subaortic Stenosis  
 Mitral regurgitation, rheumatic

**FINDINGS.**

Left Ventricular Hypertrophy **Info**

**MAGNET** **QUIZ** **REPORT** **SHOW RELATED**




Figure 3. Case scenario (top left), list of available tests (lower left), chest film (middle) with one overlay visible; a finding can link to its differential diagnosis (right) and then branch to its profile and discussion (overlapped windows on right).

results that might have occurred (e.g., murmurs other than that associated with the actual case or other possible roentgenographic images such as those showing left atrial rather than left ventricular enlargement or mitral valve rather than aortic valve calcification). For each of these alternative results, the student can inspect the list of findings associated with that result and access the differential diagnosis, associated disease profiles, and discussions. The student may also access a discussion of the test procedure itself. For any of the above content presentations, the source of the content is provided, along with reference citations, if appropriate.

After entering a provisional differential diagnosis, students may request invasive tests and view the results. Digitized, high-resolution cardiac catheterization sequences, cardiac echocardiograms, plain films, and text descriptions of laboratory values are available to the student. Again, the student may access other content that amplifies on the findings or on those associated with other comparative results. The student continues through the case, managing the patient and reviewing laboratory and patient data along the way.

Simulations of pathophysiological mechanisms are provided when appropriate to help the student to relate the clinical findings to the underlying pathophysiology. As an example, the effective valve orifice is computed for a set of data on the cardiac output, heart rate, and mean systolic valve gradient of a patient (Figure 4). The student can change the values of these data and see the results immediately. These calculations allow the student to perform 'what-if...' analyses and visually explore the relationships among these factors. Another simulation consists of an animated line drawing synchronized with sounds of the heart that shows a beating heart with valves moving; coinciding with each heart beat is a pressure reading of the aorta and left ventricle.

The material in the case is supplemented by allowing students to launch external applications from within the courseware. For example, a cardiac auscultation simulation (HeartLab) and a program on cardiac anatomy and physiology (HyperHeart, University of Utah Medical Center) can be accessed by students from within the case presentation.

### *5.5. Interface design*

The program is designed to use color, and with large-screen monitors case presentation and exploration are particularly effective, although it can adapt to other monitor sizes. On the left side of the screen are two resizable text fields (Figures 3 and 4). The upper left area displays the clinical scenario of the initial presentation of the patient and accumulates new patient information as the student proceeds through the case.

The lower left area displays the tests and physical findings that may be interrogated at the current time (these change as the student proceeds through the case), grouped under major headings, which expand when selected or collapse when not selected. A test is requested by selecting it with the mouse.

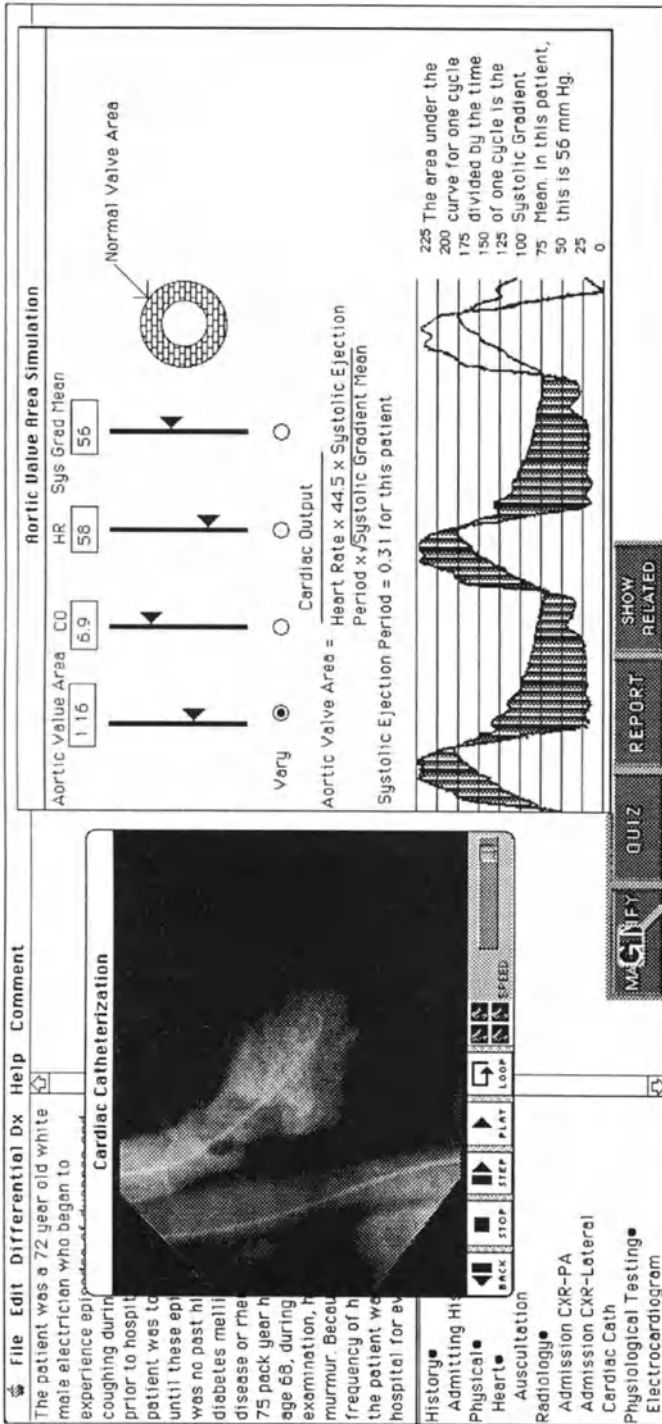


Figure 4. Case scenario (top left), list of available tests (lower left), cardiac catheterization animation window (middle), aortic valve area simulation (right).

The right side of the screen, up to the edge of the resizable text areas, is devoted to images. The bottom right area of the screen contains buttons that provide access to the official report or interpretation of an image – magnification (to view areas of the image in full resolution), images that represent normal examples of the test, a list of the findings in the image, a list of diseases with the same finding(s), related images, and descriptions of the diagnostic procedure.

There are additional buttons in this area that enable the student to branch to quizzes on the material when appropriate or to select treatment for the patient. A ‘help’ window provides instructions concerning the interface and suggestions to direct the attention of the student. This help window can be closed and called up again at any time. At each stage, the suggestions change to fit the circumstances.

### *5.6. Quizzing and self-assessment tool*

The student can be quizzed at any time by means of a hypermedia examination based on either the information so far presented or related information. To help construct the hypermedia quizzes, a program was developed that allows an author to create text-based answers only, image-finding answers only, or mixed text- and image-based answers. Where appropriate, text-based questions can be linked to digitally recorded sounds that the student can replay. After deciding upon an answer, the student enters his/her response by selecting an area of the image or by selecting the button next to the text-based answer. The author can set up the quiz to go to another question immediately, give the student feedback immediately, or allow the feedback to be seen after all questions have been answered. The author can establish how many of the entered questions the student is to answer, and the student will then see a random selection of the total questions. A time limit can be set so the student will not spend excess time on the quiz.

## **6. Evaluation**

Of the many challenges associated with providing quality computer-based systems for medical education and decision support, the design and implementation of adequate evaluation methodologies remain the most difficult. Those of us interested in introducing computer-based educational tools into the medical curriculum must justify the high cost of acquiring and maintaining computer hardware and the often higher cost of developing and maintaining the necessary software. That is, we must provide solid evidence that these tools are not only as good as, but better than the conventional alternatives.

Open questions regarding this mode of case presentation and problem solving with easy access to relevant associated material, include the following: (1) Is this mode of problem solving enjoyable, challenging, and engaging to the

students? (2) What are the patterns of access to the available material? What content is accessed and what is not? (3) What areas are found unacceptable, confusing, or inadequate in other ways? (4) How efficient is the learning process in terms of time spent per unit of information gained, and how effective is it in terms of the tasks of understanding the range of possibilities for differential diagnosis, workup strategy, underlying mechanisms of disease, treatment, and prognosis? (5) Returning to our original thesis, how does the case problem solving approach compare with more traditional didactic or textbook-oriented approaches in terms of acquisition of workable knowledge in these tasks? We are currently pursuing these questions, but cannot yet provide definitive answers, except with respect to the first question, for which the verdict is clearly positive.

One hypothesis is that this kind of approach to learning medical content can serve as a 'mental model' for a topic that can be retained for future use. The same case study not only could be used in the context of a particular course (e.g., pathophysiology), but also could be revisited in other courses (e.g., pharmacology or biochemistry) and considered from other perspectives (e.g., diagnostic strategy, clinical efficacy, therapeutics). At each subsequent encounter, the knowledge accessed might be different or the focus on particular aspects could occur in greater depth. A challenge in using this paradigmatic approach, therefore, is both to make a wide variety of content available while guiding the student to those aspects most germane to the current problem and to enable the student to navigate effectively and avoid being overwhelmed by material extraneous to the current problem. Meeting this challenge will require careful design of the user interface and modification of the courseware as a result of student experience and feedback.

Talking, both formally and informally, with peers, domain experts, and representative users is of value throughout the design, implementation, production, and maintenance phases of the lifecycle of any application. Formal and informal questionnaires are the commonest form of data collection, perhaps because of their apparent simplicity and ease of use. Although questionnaires can provide a large amount of data with a relatively small investment of time and energy, they must be carefully planned if the results are to be meaningful. Improperly designed questionnaires can provide vast amounts of worthless data.

Watching (or videotaping) representative users interacting with a system is usually more revealing than interviews and can provide data on interface design, general patterns of use, and indications of user frustration. Videotapes of user interaction provide a nonvolatile record of how the software is used and can be repeatedly examined to answer new questions. Busy students cannot reasonably be expected to voluntarily maintain a daily log of their computer activity. Fortunately, it is relatively easy to add code to a program to monitor how it is being used. The log of user activity, or audit trail, can provide a wealth of information (e.g., those parts of a program that are used most often, what commands are most or least often used, and what the most typical patterns of usage are).



In our laboratory we use program instrumentation techniques extensively to evaluate software projects. Compared with other common evaluation techniques, program instrumentation offers a number of benefits, such as the following: (1) Minimal tester involvement. Since the monitoring system is computer-based, human intervention is required only to evaluate the audit trail. (2) Nonbiased results. Unlike the human observer, who may be forced to make 'judgment calls' in certain situations, computer monitoring simply compiles a record of the actual events. (3) Automatic documentation. Documentation of the audit trail is automatic and, therefore, free of human errors that could occur during transcription. (4) No loss of data. Computer-based monitoring is not susceptible to the loss of data due to lapses of concentration or forgetfulness on the part of the observer.

Despite the power and simplicity of program instrumentation, this evaluation technique must often be supplemented with other methods that gather different types of data. For example, program instrumentation does not directly record user confusion or allow subjective input. To adequately provide for subjective input, we are developing a more elaborate system that will provide users with a means of entering suggestions at any point in the program. These free-form messages (e.g., 'I don't understand my options here') should help highlight deficiencies in program design that may not be obvious from a simple audit trail.

Program instrumentation has other limitations as well. For example, unlike the technique of observation, program instrumentation does not make allowances for (or even detect) multiple, simultaneous users. Another limitation of program instrumentation, in situations where large amounts of data are to be gathered, is that the computer hardware must include a hard drive or other mass storage device, since a floppy disk may be filled too quickly to be useful. Also, if students shut down the machine with the power switch, rather than through the operating system, data can be lost. For example, if the event-tracking program is in the middle of a disk write when the power is abruptly shut off, there is a high probability that the data file be lost or corrupted.

In the first year, only one of the three cases in the HST cardiac pathophysiology course is presented in computer form. While each of the three cases deals with a different class of clinical problem, we have the opportunity to measure student acquisition of image interpretation skills and to test for the amount of background knowledge (e.g., in pathology, physiology, and differential diagnosis) that is acquired with use of computer-based versus paper-based modes. These results will also serve as a baseline against which to compare the introduction of computer-based versions of other cases in the following year.

## **7. Future directions**

Our future directions in computer-based clinical problem solving focus on determining how to maximize the educational effectiveness of the experience.

For example, we are investigating the applicability of a variety of tutorial strategies, the potentials of intelligent tutoring, and how information gained from student modeling can be used to tailor courseware to suit student needs. These areas of ongoing research are outlined below.

### *7.1. Tutorial strategies*

The commonest tutorial strategies used in traditional Computer-Aided Instruction (CAI) applications have been the following: (1) drill-and-practice, which involves the repetitive presentation of materials and is most suitable for learning facts; (2) motivation, in which gaming techniques are used to make learning more enjoyable; (3) teaching by example, wherein the simulation presents an example as an illustration of the general case; (4) teaching by analogy, in which a concrete, familiar situation is used as a platform for the presentation of new material in related, familiar terms (for example, cardiac circulation could be explained at first in terms of a more intuitive hydraulic plumbing system so that the plumbing system serves as a mental model to help the student learn the new material); (5) discovery learning, in which an unstructured environment is created and students are free to test hypotheses and discover knowledge on their own; (6) spatial problem solving, in which graphics are used to build up mental representations of the task; and (7) directed learning or tutoring, wherein the system guides the student to the correct answers or strategies (Eberts 1986).

Of the strategies described above, the ones most frequently used in CAI applications in medicine have been drill-and-practice and tutoring. The drill-and-practice strategy assumes that the student understands the principles of the subject matter and needs practice to transform this general understanding into mastery and application of the material. The tutoring strategy makes use of the computer to deliver instruction and to mimic the student-teacher relationship. Although both drill-and-practice and tutoring closely parallel traditional teaching methods, these approaches may be ineffective for many students. For example, by highlighting misconceptions, faulty reasoning, omissions, and other errors, the tutor provides students with the information they need to adjust their thinking so that it is more in line with that of the tutor. And, since frequent but moderate corrections in the course of the study are preferable to infrequent but possibly major ones, it follows that for effective learning to occur, computer-based tutoring systems, like their human counterparts, should ideally verify the student's knowledge after each session of student-tutor interaction.

Computer-based systems, while currently not as effective as human tutors in promoting student understanding, can potentially provide superior insight into a student's perception of a domain. This is possible primarily because of the direct, dedicated, one-on-one communications channel between the student and the computer-based courseware. To fully exploit this intimate student-computer link, the courseware must maintain student models that contain

relevant information about particular students (a student's approach to learning, and his or her misconceptions about a subject domain). Because students may develop their own valid sets of rules and approaches to solving particular problems, it is preferable to use systems that can work in conjunction with the students' own reasoning, instead of those that use predefined algorithms distilled from domain experts. In fact, rewarding strict adherence to predetermined algorithms may serve only to stifle student creativity.

## *7.2. Intelligent tutoring systems*

The composition of an intelligent tutoring system (ITS) is commonly described in terms of four interrelated modules: a student model, the domain expertise, the tutoring component, and the user interface (Weneger 1987; Woolf and McDonald 1984). One objective of ITS research has therefore been to design programs that construct insightful models of a student's strengths, weaknesses, and preferred style of learning. To achieve this end, ITS researchers have made use of artificial intelligence (AI) work in areas of natural language understanding, knowledge representation, and methods of inference (Clancey, Shortliffe et al. 1984). The initial work in ITS focused on techniques of knowledge representation (Brown and Burton 1975; Carbonell 1970; Suppes and Morningstar 1972). In later work, generative tutoring systems were developed (Wexler 1970); these knowledge representation techniques were augmented with AI techniques and were used to formulate models of the learner in terms of issues or skills that should be learned (Burton and Brown 1976). Contemporary ITS programs use AI techniques to represent the tutorial strategies themselves, a technique that is said to have the advantage of flexibility and modularity of representation and control (Brown 1976; Goldstein 1977). Providing the student with feedback as to the appropriateness and completeness of his or her solution is central to the design of both CAI and ITS applications. Programs range from providing feedback at each decision point (Feurzeig 1964) to reporting only at the end of the solution (Harless 1971). CAI systems provided with frame-oriented systems cannot tolerate a random dialogue in which partial responses are offered to the system. Programs based on statistical tables (Harless 1971; Kirsch 1963; Steel 1978), while providing more flexibility than simple frame-oriented systems, have no way to justify correlations or characterize what are the logical, significant questions to ask. One approach that has had limited success in providing meaningful feedback makes use of overlay models (Brown 1976; Carr and Goldstein 1977). In this ITS technique, the tutorial program explains the student's behavior in terms of domain-specific rules. This methodology cannot, however, detect the use of incorrect rules or concepts, nor can it support reasoning that is different from that contained in the underlying domain rules.

A high degree of realism in the tutorial problem solving environment is important in ensuring transferability of learning from the program to the clinical setting. However, the simulation of real-world environments has not

been a major focus of current ITS research. Most systems are text-based and provide the student with a patient who can be interviewed (Harless 1971), place the student in a hospital setting where dialogues are established with an attending physician and the patient (Feurzeig 1964), or simulate a discussion with a resident physician or classroom instructor (Diamond 1974; Weber and Hageman 1972).

### *7.3. Student modeling*

In their comprehensive review of ITS, Sleeman and Brown identify several basic models for representing student knowledge (Sleeman and Brown 1982). Although these early models have had a significant impact on virtually all subsequent student modeling systems, they fail to provide an adequately personal view of the student. That is, they cannot diagnose the student's errors within the context of his or her own conceptualization of the problem, but must instead rely on a library of rules, examples, or plans compiled from domain experts. Clearly, the task of determining the student's misconceptions for the purpose of tailoring the system to the user's needs must make use of alternative approaches, especially when deeper reasoning is demanded of the student.

The resource requirements associated with providing courseware systems with comprehensive student modeling capabilities – including the large time investment required to develop and verify domain-specific knowledge bases for each specialty area and the difficulty in creating student models that can be used in more than one domain – are prohibitive. We have therefore focused on developing generic adaptive systems that can be used with actual courseware systems, albeit with less sophisticated student modeling capabilities than those demonstrated by more complex laboratory systems. In particular, we have investigated the applicability of neural network technology as a means of providing practical adaptability for our multimedia educational programs (Bergeron 1989).

### **Acknowledgment**

This work was supported in part by grants LM 07037, LM 04715, and LM 04572, and contract N01-LM-8-3513, from the National Library of Medicine, by grant CA 45574 from the National Cancer Institute, by an equipment grant from Apple Computer, Inc., and by funds from the Health Science and Technology Division, Harvard Medical School and Massachusetts Institute of Technology.

### **References**

- Abendroth T, Greenes R (1989) Computer presentation of clinical algorithms. *M.D. Computing* 6(5): 295-299

- Barnett O (1984) Computer-based simulations in clinical problem solving. *Med Inf* 9(3-4): 277-279.
- Beck JR (1983) Microcomputer aids to medical decision making: educational and clinical applications. *Med Decis Making* 3: 3-7
- Bergeron B (1987) Heartlab: A multi-mode simulation for teaching cardiac auscultation. *Simulation* 50(2): 50-60
- Bergeron B, Greenes R (1987) Heartlab and EKGlub: skill-building simulations in cardiology. Eleventh Annual Symposium on Computer Applications in Medical Care (SCAMC), Demonstration Digest 29-30
- Bergeron BP, Greenes RA (1989) Skill-building simulations in cardiology: Heartlab and EKGlub. *Comput Method Programs Biomed* 2/3(30): 111-125
- Bergeron BP (1989) Using a spreadsheet metaphor to visualize neural network behavior. *Collegiate Microcomputer* (in press)
- Borning A (1979) THINGLAB – A constraint-Oriented Simulation Laboratory. Report SSL-79-3, Xerox Palo Alto Research Center, Palo Alto, CA
- Brown JS (1976) Reactive learning environments for computer-aided electronics instruction. Report 3314, Bolt Beranek and Newman, Cambridge, MA
- Brown JS, Burton RR (1975) Multiple representations of knowledge for tutorial reasoning. *Representation and Understanding*. New York, Academic Press
- Burton RR, Brown JS (1976) A tutoring and student modelling paradigm for gaming environments. *Proceedings of the Symposium on Computer Science and Education* 236-246
- Carbonell JR (1970) Mixed-initiative man-computer instructional dialogues. Report 1971, Bolt Beranek and Newman, Cambridge, MA
- Carr B, Goldstein I (1977) Overlays: A theory of modelling for CAI. Report 406, Artificial Intelligence Laboratory, M.I.T., Cambridge, MA
- Clancey WJ, Shortliffe EH, Buchanan BG (1984) Intelligent computer-aided instruction for medical diagnosis. *Readings in Medical Artificial Intelligence – The First Decade*. Reading, MA, Addison-Wesley
- Clyman SG (1990) Medical school testing computer-based exam. *Computer News for Physicians* (February): 15-42
- Cobelli C, Mari A (1985) Control of diabetes with artificial systems for insulin delivery – algorithm independent limitations revealed by modeling study. *IEEE Transactions on Biomedical Engineering* 32(10): 840-845
- Conklin E (1987) Hypertext: An introduction and survey. *IEEE Computer* 2(9): 17-41
- Davis RS, Jack DB (1986) A surgical cut into computers. *MD Computing* 3(1): 21-28
- Diamond HS (1974) A computer assisted instructional course in diagnosis and treatment of rheumatic diseases. *Arthritis and Rheumatism* 17(6): 1049-1055
- Dichter MS, Greenes RA, Bergeron BP (1990) A medical case presentation shell for directed hypermedia. *Spring Conf of American Medical Informatics Association (AMIA)* (abstract)
- Eberts R (1986) Computer-aided instruction. *MD Computing* 3(3): 20-28
- Eddy DM (1982) Clinical policies and the quality of clinical practice. *N Engl J Med* 307: 343-347
- Feurzeig W (1964) Computer-aided teaching in medical diagnosis. *Journal of Medical Education* 39: 746-755
- Friedman D, Schwartz W (1987) A diabetes simulation to teach insulin therapy. *IBM ACIS University Conference* 164
- Fundt BV (1981) problem solving with diagrammatic representations. *Readings in Knowledge Representation*. Los Altos, CA, Morgan Kaufmann Pub Inc
- Gabay TV (1983) Automated teaching systems in the USSR. *Computer-Based Learning*. England, Pergamon Infotech Ltd
- Gardner D (1987) Virtual Cat: A student lab simulation with biological variability. *IBM ACIS University Conference* 103
- Getty D, Pickett R, D'Orsi C, Swets J (1988) Enhanced interpretation of diagnostic images. *Inv Radiol* 23(4): 240-252
- Goldstein I (1977) The computer as coach: An athletic paradigm for intellectual education. Report 389, Artificial Intelligence Laboratory, M.I.T., Cambridge, MA

- GPEP 1984 Physicians for the Twenty-First Century: The GPEP Report. Report of the panel on the General Professional Education of the Physician and College Preparation for Medicine. American Association of Medical Colleges
- Greenes R (1986) Knowledge management as an aid to medical decision making and education: The Explorer-1 system. MEDINFO 86. Amsterdam, Elsevier Science Publishers. 895-899
- Greenes R (1989) Desktop knowledge: A new focus for medical education and decision support. *Meth Inf Med* 28(4): 332-339
- Greenes R, Tarabar D, Cope L, Slosser E, Hersh W, Pattison-Gordon E, Abendroth T, Rathe R, Snyder-Michal J (1989a) Explorer-2: An object-oriented framework for knowledge management. MEDINFO 89. Amsterdam, Elsevier Science Publishers. 29-33
- Greenes R, Tarabar D, Krauss M, Anderson G, Wolnik W, Cope L, Slosser E, Hersh W (1989b) Knowledge management as a decision support method: A diagnostic workup strategy application. *Computers and Biomed Res* 22: 113-135
- Greenes RA (1983) Interactive microcomputer-based graphical tools for physician decision support: Aids to test selection and interpretation and use of Bayes' theorem. *Medical Decision Making* 3(1): 15-21
- Harless WG (1971) CASE: A computer-aided simulation of the clinical encounter. *Journal of Medical Education* 46: 443-448
- Hoffer E, Barnett O (1975) Computer-aided instruction in medicine. *Annual Review of Biophysical Engineering* 4: 103-118
- Humphreys BL, Lindberg DA (1989) Building the unified medical language system. Thirteenth Annual Symposium on Computer Applications in Medical Care (SCAMC) 475-480
- Kirsch AD (1963) A medical training game using a computer as a teaching aid. *Methods of Information in Medicine* 2(4): 138-143
- Komorowski H, Greenes R, Barr C, Pattison-Gordon E (1988) Browsing and authoring tools for a Unified Medical Language System. RIAO 88 Conference: User-Oriented Content-Based Text and Image Handling. Massachusetts Institute of Technology. 624-641
- Krenz W (1985) Simulation of use-dependent uptake of ion channel blocking agents by excitable membranes. *IEEE Transactions on Biomedical Engineering* 32(10): 817-825
- Mackenzie CF (1987) Tools for teaching anesthesiology and respiratory/cardiac physiology. IBM ACIS University Conference 104
- Meyer RE (1981) The psychology of how novices learn computer programming. *Computer Surveys* 13: 25-35
- Miller R, Masarie FE, Meyers JD (1986) Quick medical reference (QMR) for diagnostic assistance. *MD Computing* 3(5): 34-48
- Miller R, McNeil M, Challinor S, Masarie F, Myers J (1986) The INTERNIST-1/QUICK MEDICAL REFERENCE project Status Report. *Western J Med* 145: 816-822
- Nandedkar SD (1985) Simulation techniques in electromyography. *IEEE Transactions on Biomedical Engineering* 32(10): 775-785
- Nelson T (1987) Literary machines. South Bend, Indiana, The Distributors
- Norman D (1983) Some observations on mental models. *Mental Models*. New York, Hillsdale
- Norman D (1984) Cognitive engineering principles in the design of human-computer interfaces. *Human-Computer Interactions*. Amsterdam, Elsevier Science Publishers
- Novak JD, Glase C (1987) PC software to improve learning in a large introductory biology course. IBM ACIS University Conference 73
- Piestrup A (1984) Game sets and builders. *Byte*. June: 215-219
- Pollack M, Greenes R (1985) Computer-based educational and analytic support through linking of pictorial simulations and spreadsheets. *Proc Society for Medical Decision Making*. Seventh Annual Meeting. 18. (abstract)
- Pollack MA, Greenes RA (1986) A pictorial simulation construction kit for enhancing knowledge-based learning. MEDINFO 86. Amsterdam, Elsevier Science Publishers. 887-890
- Ross BH (1982) Remindings and their effects in learning a cognitive skill. *Cognitive and Instructional Science Series*. Palo Alto, CA, XEROX Palo Alto Research Center
- Salzsieder E (1985) Kinetic modeling of glucoregulatory systems to improve insulin therapy. *IEEE*

- Transaction on Biomedical Engineering 32(10): 846-855
- Sleeman D, Brown JS (1982) Intelligent tutoring systems. Troy, Mo, Academic Press
- Smyth-Staruch K, Littenberg B (1983) Using microcomputers to teach sensitivity analysis to medical students. *Med Decis Making* 3: 9-13
- Starmer CF, Kerr RB (1985) Simulation of use-dependent uptake of ion channel blocking agents by excitable membranes. *IEEE Transactions on Biomedical Engineering* 32(10): 770-774
- Steel AA (1978) A computer-assisted instruction (CAI) program in diseases of the thyroid gland (thyroid). *Computers and Biomedical Research* 11: 133-146
- Steinhaus BM, Spitzer KW, Isomura S (1985) Action potential collision in heart tissue – computer simulations and tissue experiments. *IEEE Transactions on Biomedical Engineering* 32(10): 731-742
- Stensaas SS, Sorenson DK (1988) Hyperbrain and Slice of Life: An interactive hypercard and videodisc core curriculum for neuroscience. Twelfth Annual Symposium on Computer Applications in Medical Care (SCAMC) 416-420
- Stofer D (1989) Hypermedia. *Popular Science*. (May): 122-160
- Stulurow KA, Cochran TM (1983) Instructional uses of computing in the health sciences: An annotated bibliography. *J Med Syst* 7(1): 61-84
- Sun Y (1985) Parameter estimation in the stenosed coronary circulatory system. *IEEE Transactions on Biomedical Engineering* 32(10): 798-805
- Suppes P, Morningstar M (1972) Computer-assisted instruction at Stanford, 1966–68: Data, models, and evaluation of the arithmetic programs. New York, Academic Press
- Tarabar D, Greenes R, Slosser E (1989) An object-oriented software architecture for the Explorer-2 knowledge management environment. Thirteenth Annual Symposium on Computer Applications in Medical Care (SCAMC) 93-97
- Tosteson DC (1990) New pathways in general medical education. *N Engl J Med* 322(4): 234-238
- Van den berg PM (1985) A computational model of the electromagnetic heating of biological tissue with application to hyperthermic cancer. *IEEE Transactions on Biomedical Engineering* 32(10): 797-805
- Weber JC, Hageman WD (1972) ATS: A new system for computer-mediated tutorials in medical education. *Journal of Medical Education* 47: 637-644
- Weneger E (1987) Artificial intelligence and tutoring systems. New York, Morgan Kaufmann Publishers
- Wexler JD (1970) Information networks in generative computer-assisted instruction. *IEEE Transactions on Man-Machine Systems* 4: 181-190
- Woolf B, McDonald DD (1984) Building a computer tutor: Design issues. *IEEE Computer* (September): 17: 61-73
- Yanlelovich N, Meyrowitz N, Dam van A (1985) Reading and writing the electronic book. *IEEE Computer* 18(10): 15-30

# Continuing education of physicians and nurses in the DBMS Area

G. INVERNIZZI<sup>1</sup>, C. COMBI<sup>1, 3</sup>, G. POZZI<sup>2, 4</sup>

<sup>1</sup> *Cardiology Department of the Ospedali Riuniti, Bergamo, Italy*

<sup>2</sup> *Dipartimento di Bioingegneria, Politecnico di Milano, Milano, Italy*

<sup>3</sup> *Fondazione Sestini, Bergamo, Italy*

<sup>4</sup> *Dottorato di Ricerca in Bioingegneria, Dipartimento di Bioingegneria, Politecnico di Milano, Milano, Italy*

## Abstract

High evolution rates of instrumentation used in clinical cardiology departments make difficult demands on personnel continuing education, both for physicians and for nurses. New instruments, providing advanced and locally-upgradable performances, call for specific skills to be acquired via appropriate training. More wide-spectrum instrumentation, typically based on personal computers and DBMS oriented, may call for different and deeper educational approaches as much as they allow the local clinical user to design the set of performances he really needs and may effectively use. We describe experiences we had in our department. In both specific training and wide-spectrum deeper education, including DBMS, we took full advantage of the full-time presence of young graduates in biomedical engineering and computer science.

## 1. Introduction

Some remarkable improvements have been observed during recent years in many branches: they can mostly be ascribed to new technological discoveries. This phenomenon has also concerned cardiology departments with an increasing intensity. Particularly, the use of digital techniques and diffusion of microprocessor technology introduced some important effects. Some of them can be summarized as follows [1]: consistent reduction of costs for numerical data processing; improvement of mathematical models and algorithms for signal and image processing; miniaturization of components and devices; improvements of the knowledge of anatomy, physiology, pathology; and computerized data storage, allowing faster retrieval and analysis [2].

Let us consider, for instance, some exams that nowadays are commonly used. One can now have images about the cardiac kinetic: this was unbelievable some years ago. Nowadays, by an echographic system, one can have images concerning the movements of the cardiac muscle. Other improvements due to technology are long-time signal recordings, pace makers, some haemodynamic invasive exams, percutaneous trans-luminal coronary-artery angioplasty



(PTCA) and many others. Acquired data, coming both from these exams and from anamnesis, diagnosis, therapies, can be managed by means of informatic tools.

The hospital environment sets up the need for continuous information about recent advances, in order to use new possibilities offered by technological improvements. The need is becoming more and more evident. If we consider that physicians started using some tools such as echographic systems or Holter monitoring as soon as they were available, then this would have to happen for informatic systems, too. Thus, it is fundamental to provide suitable training in order to have correct use of the instruments [3,4]. Training must be specific both for those who will be physicians and nurses in the future and for physicians and nurses now present in the department. In fact, it does not make any sense to wait for the new physicians and nurses (those who are already familiar with computer-science) to reach the department: it would mean giving up immediate applications for lack of personnel with the appropriate knowledge. To have computer-based systems used in everyday practice starting from the present, it is necessary to provide physicians and nurses now working in the departments with suitable training.

The correct use of informatic systems and of bio-medical engineering techniques can allow immediate improvements not only by the use of Data Base Management Systems, but also by the use of bio-signal processing, help in decision making, expert systems, and bio-medical instruments.

## **2. Materials and methods**

Our cardiology department belongs to the Ospedali Riuniti of Bergamo, Italy: the hospital works as a country hospital. The total number of admitted patients in the hospital can reach 3000 in about 50 departments.

Our department is made up of 3 main Units. The Cardiology Division (23 beds), the Pediatric Unit (newborn babies and babyhood: 12 beds), the Coronary Care Unit (CCU) (Intensive Care Unit: 7 beds, Sub-Intensive Care Unit: 3 beds). Admitted patients (1989) were: Division 1620, CCU 757, Pediatric Unit 389. Most frequently observed pathologies are AMI, arrhythmias, ischemia, and congenital diseases.

In our department, there are two operating rooms, where haemodynamic exams are performed: during 1989, examined patients were 1929 by 2584 performed exams.

The staff comprises a head of department, 12 physicians, 4 volunteer physicians, 3 head-nurses, 31 professional nurses, 14 nurses, 3 X-ray technicians, and 2 secretaries. Moreover 2 young Bio-Medical Engineers have been present for 3 years, granted by a postgraduate fellowship.

In our department we have some computers. According to their performances, they can be divided into two main groups: the CCU computerized system and the PC family systems.

The CCU computerized system performs data acquisition and real-time analysis of patients' data: these tasks are performed by bed-side intelligent monitors and by two mini computers. The PC family systems perform other tasks, like running special programs and use of personal productivity tools: this family includes two PCs running MS-DOS, an image processing workstation and a Macintosh.

### 2.1. *The CCU DBMS*

The computer system of our CCU has been present since 1983. A two-hour training course has been performed during set-up.

Particularly interested in the computer system proved a head-nurse and a couple of physicians. They learned to use the system on their own, just by trial and error techniques.

Some important tasks have to be performed by the CCU computer system: data acquisition, analysis, management and storage.

**2.1.1. *The hardware.*** The system of our CCU is specific for the analysis of both automatically acquired ECG data and via keyboard inserted data. It consists of 3 units [5].

The first unit is made up of bed-side intelligent monitors: they perform signal acquisition (ECG, pressure, temperature) and simple analysis (HR, systolic and diastolic pressures, temperature) and displaying (ECG signal, blood pressure from a catheter).

The second unit consists of an HP-1000 mini computer for the detection of arrhythmic events, their classification, and alarm-setting.

The third unit is made up of another HP-1000 mini computer, for the management of patients' data and for the automatic analysis of the ST-T segment in the ECG [6]. Both HP-1000 minis run the RTE operating system.

**2.1.2. *The software.*** The software is specific for the CCU, and it has been developed by HP and by CNR (National Research Council) of Pisa, Italy.

The software consists of three units, where some data can be shared.

***ECG analysis (I).*** The first unit (NADIA) performs real-time automatic arrhythmia detection and displays trends (HR, frequency and number of VPBs and blood pressures). As one or more beats are classified as arrhythmic, an alarm is set: an acoustic signal alerts nurses. These operations are performed within 6 seconds from the onset of the abnormal event. According to the severity of the situation, three different alarms can be set.

The system allows the user to review any beat that has been classified as arrhythmic and, consequently, stored by the program. This back-jump is allowed over an interval of 3-4 days: it can be very useful for the physicians to see in a few seconds if the patient had any particular arrhythmia during the past few hours-days. The system also allows the user to display set alarms: this way a log about the patients' history can be obtained.

*ECG analysis (II).* The second unit performs real-time ischemia detection. This takes place by an analysis of the ST-T segment: time plots of areas subtended by the signal in respect to a baseline can then be displayed. This way, it is possible to check the developments of a clinical situation. It is also possible to detect silent ischemia: without a computerized monitoring system, this phenomenon would go unobserved, since no pain is present and onset and offset occurrences cannot be foreseen. This unit has been developed in cooperation with the CNR.

*Clinical data management.* The third unit – PDMS: Patient Data Management System – performs management of data of admitted patients. Data are entered via keyboard.

PDMS can be customized according to criteria defined by final user: he can define paragraphs, subparagraphs, parameters, either inserted via keyboard and automatically acquired by instrumentation. Customization is performed by suitable programs, and it is a very delicate step. PDMS can also receive data from instrumentation (intelligent monitors) and from other programs (ECG analysis I, II): customization must consider technical data of all devices acquiring and exchanging data. PDMS has been customized during the system set-up and only very few changes have been done later on.

For each patient stored data are divided into paragraphs and, in some cases, into sub-paragraphs. Considered paragraphs are:

- Demographic data: this paragraph collects ID data (birth-date, address, phone number) and some anamnestic data (for example, allergies, reason for admission, other contemporary pathologies..). All these data are stored in free-text format;
- Lab and instrumental data: this paragraph, including some sub-paragraphs, collects data related to laboratory and instrumental exams performed during hospitalization. It includes 35 subparagraphs (e.g. blood-gas-analysis, coagulation, haematology, cytology, endocrinology, echography..). For example, in ECG data paragraph stored parameters are: heart rate, description of rhythm, PQ duration, QT duration, QT maximum duration, QRS axis, presence of VPBs, ST variations....
- Respiratory data: this paragraph collects data about respiratory activity. For example, considered parameters are: respiratory rate, ventilatory index, total current volume exchanged per breath, percentage of oxygen inspired...
- Haemodynamic data: this paragraph collects data related to haemodynamic exams that have been performed: heart rate, respiratory rhythm, systolic and diastolic blood pressure, stroke volumes, cardiac output, cardiac index, pressures in pulmonary artery, haematocritus, height, weight, oxygen partial arterious pressure, oxygen partial venous pressure...

During data input and retrieval an on-line help service is available, displaying normal and pathological values. In this way one can immediately evaluate how severe the situation of the patient is.

All data, except demographic, are stored in cards: each card considers a set

of parameters. One patient may have more cards, considering the same parameters measured in different times.

Some data are automatically computed by the system starting from previously inserted data: e.g. the body surface area (BSA) is automatically computed starting from weight and height. While using PDMS, one can also display formulas used for automatic computations.

The system allows construction of time-plots representing the evolution of some parameters. Since chosen parameters can come both from PDMS itself and from the other units (NADIA and ST-T analysis program), data exchange is needed. In fact, personal ID data (they do not change during hospitalization) are shared as they are entered. HR, number of VPBs per minute, invasive blood pressures and other parameters automatically computed by NADIA are continuously passed to PDMS. This way, data input is reduced and all data can be referenced by PDMS.

## *2.2. DBMS on personal computers*

The total number of PCs in our department is 4. Three of them can run the software commonly available under the MS-DOS operating system: they are two normal IBM-compatible PCs and a workstation for image processing. The fourth PC is a Macintosh.

*2.2.1. The hardware.* Two PCs have been present in our department for three years. They are used both by bio-medical engineers and by some members of the staff.

The workstation has been present for some months. It has an 80386-based microprocessor unit, controlling two displays: it also includes a video-tape player. Images come from X-ray haemodynamic exams and echocardiographic exams. Image acquisition mainly takes place via the video-tape player. Acquired images can then be shown on a display, and 'edited' by the use of a mouse. The workstation is mainly used by the bio-medical engineers: in the near future, as the software will have reached a final prototype release, physicians will start using it in practice.

Recently the department has been given a Macintosh. Its user-interface made it immediately very popular among physicians, nurses and secretaries: they are the main users of the Macintosh system.

### *2.2.2. The software.*

*Commercial DBMS.* PCs run some packages which, even if not specific for a cardiology department, can prove to be useful in usual tasks such as letter-writing and data storing. All these packages have been developed outside the department.

Used DBMS are dBase III plus and Paradox for MS-DOS systems, and Fourth-Dimension and MacMAX for the Mac.

These packages are widespread and are quite easy to use.

*The image processing system.* The workstation runs a special software for image processing. This software is still a prototype under development: it is being developed outside the department [7]. Having an 80386-based CPU, the workstation can also run all the software developed for the MS-DOS operating system.

The software for image processing performs data acquisition, displaying, analysis and storage. It is mainly intended for the measurement and automatic computation of parameters. Examined images come both from echocardiographic exams and haemodynamic exams.

Starting from echo exams, parameters that can be computed are, for instance, wall dimensions, stroke volume, blood speed. Starting from cath exams, parameters that can be computed are, for instance, vessel dimensions, degree of stenosis.

Typical use of the workstation relates to ventricular and coronary artery analysis. As a first step (the software still is under development) the system can evaluate stenosis in coronary arteries. By the mouse, user interactively analyzes the image and identifies position of catheter, vessel to be considered, stenosis: the system evaluates some other parameter as exact position and percentage of stenosis. In case of PTCA, the system can also evaluate, by comparing cath exams before and after dilatation, how successful angioplasty has been for that patient.

The workstation can store all these data (images, inserted and computed parameters) for subsequent use. It can also store ID data, height, weight, diagnosis, description of performed haemodynamic exam. Furthermore, all these data can be converted to ASCII format and used by most DBMS and statistical packages.

### *2.3. Adopted methods of training and education*

To perform suitable training and education in the DBMS area, the following methods can be adopted:

- DBMS theory: basic concepts on database theory are taught. Education is not directly oriented to informatic resources actually present in the department;
- DBMS practice: only very few database theory concepts are taught. Training and education are mostly oriented to the use of DBMS systems present in the department.

We chose the second solution. Many reasons led us to this choice. First of all, physicians and nurses involved had no specific knowledge about computers. None of them was specifically employed for the computerized resources management. The need was that of using the system of the department. This way the presence of DBMS in the cardiological department should have given immediate advantages.

Training and education were tailored to the systems of the department.

The DBMS of CCU can mainly be used 'as-is'. Customization, performed

during set-up, is strictly dependent from technical features of used instrumentation: only a few changes can be made. The main task thus was to train personnel in using the system at best. This kind of training was devoted to all the personnel of the department.

The DBMS on PCs can be used as development tools. Education on DBMS for PCs was thus specific about design requirements of projects. Taught topics were data types, data structures, data coding, database models. All this information is necessary when defining performances a software must have to be useful in a certain situation. This kind of education was thus devoted to only some physicians and nurses in the department, who proved to be really very interested in these topics.

Time available for training and education in the medical DBMS area was scarce for both physicians and nurses.

At the beginning, only some physicians and nurses proved to be really interested in the DBMS area. Under these circumstances, it was important to single out advantages coming from the use of the computerized data management in everyday practice.

Techniques, that can be adopted, are:

- on-site introductory courses, about DBMS used inside the department and including fundamental concepts on DBMS theory;
- personal and informal training and education on programs and packages used inside the department;
- practical instruction by means of CAI (Computer Aided Instruction) programs developed for teaching DBMS concepts.

We chose the first two techniques. By on-site courses personnel can learn basic concepts of DBMS theory and learn how to use the systems in practice. Moreover, if a problem arises while using the programs, staff are taught how to solve it.

By personal and informal training and education, any single user is assisted just during practice. This way training and education are ad-hoc for each user, physician or nurse. Learning is easier and more efficient both of basic concepts of DBMS and the way to correctly use existing DBMS.

Computer Aided Instruction techniques devoted to DBMS were not used for various reasons. In fact these techniques may be very useful for full-time students. We considered it more suitable to teach the personnel on the system they were going to use in everyday practice. Moreover, CAI packages about DBMS running on our systems were general purpose ones, and not for our specific needs.

### **3. Results**

During recent years and by the presence of the two bio-medical engineers, some results were achieved. They can be sketched out as follows: on-site organization of short introductory courses during working hours; personal and informal

training and education; development of ad-hoc special programs; participation in national standardization protocols for the medical record and for pharmacological care.

### *3.1. On-site introductory courses*

In order to provide the personnel with education about the DBMS of the department, some short on-site introductory courses have been organized. These courses were given by the bio-medical engineers during the staff's working time. They concerned both the instrumentation of the CCU and the image processing workstation.

During these courses, theory of computer science included only certain fundamental concepts. This was in order to give more space to practical lessons.

The main advantage provided by these on-site courses is the following: during the courses personnel learn how to use the system they have to use in everyday practice. Most of the typical situations that can happen, have been presented during the courses. The personnel can therefore learn how to cope with the problem.

What is important for personnel to know is that the presence of a computer system is no substitute for physicians and nurses: the system helps them to do their work better. About the image processing system, some courses have been provided for physicians in order to use the system better. In fact, some misuse of the system can lead to unpleasant situations: a wrong diagnosis of severe stenosis in coronary artery, obtained because of incorrect use of the system, can lead to a patient undergoing heart-surgery unnecessarily.

### *3.2. Personal and informal training and education*

After a first step based on training and education by means of on-site courses, we chose personal and informal education.

This kind of training and education, carried out by two bio-medical engineers present in the department, provides some advantages. First of all, the learner is personally involved and, at the same time, he is immediately led to the solution of the problems he met. Furthermore, interest in the DBMS field increases and the computer system is used more and more for the solution of problems occurring in everyday practice.

Topics dealt with during informal education are many and heterogeneous, with different levels of competence being reached. It is possible, however, to list some topics dealt with during these years:

*How useful DBMS are in clinical practice.* This subject was the first and the most recurring one. We provided explanations with practical examples both by means of PDMS, and by DBMS packages running on personal computer.

We found it necessary to focus attention on DBMS usefulness in a hospital environment, too: this was the first step to overcome the initial mistrust of some people about informatic resources.

*Data retrieval and processing through PDMS.* In order to increase the active use of PDMS, we tried, especially with personnel working in the CCU, to emphasize the PDMS functions both about patient data retrieval, and about capability of displaying through graphs the time trends of parameters.

*Standard DBMS properties and applications.* We taught some people, on DBMS packages on Personal computer. Typical features of a common DBMS (dBase III plus) were illustrated: data types, data input, sorting, retrieval. In this way some people constructed on their own some personal simple databases to manage their data.

*Requirements and properties in an ad-hoc DBMS project.* Personnel working in the Cardiology department are possible users of DBMS applications. Thus it seemed useful to explain what one can obtain from a DBMS application. At the same time, the staff can define what to expect from an application.

Training and education needed inside the department was generally devoted to users rather than programmers.

The interest about DBMS led some physicians and nurses, in cooperation with two bio-medical engineers, to define some problems and situations to be managed by DBMS applications.

### *3.3. Ad-hoc DBMS programs design and development*

As a result of on-site introductory courses, informal training and education and full-time presence of two bio-medical engineers, some physicians and some nurses could be able to define design requirements of some software. They formulated requests for development of software to solve specific needs. The software we are going to describe runs on PCs (IBM compatible or Macintosh) and has been developed inside the department. All this software is devoted to the management of data: data are related both to patients and to material used during haemodynamic exams.

The first software, running on MS-DOS PCs, stores data about the haemodynamic material in the warehouse. As some material is used during an exam, its quantity in the warehouse database has to be decreased. This way, as the total quantity of material is lower than predefined thresholds, it has to be re-ordered. Nurses appreciate this software, because it provides some automatic services like reordering letters.

A second software, also running on MS-DOS PCs, is a special log book considering special drugs like barbiturates, morphine: according to the law we have in Italy, when one of these drugs is prescribed, some data must be registered (ordering physician, patient's name, drug name, posology, date, time..). Periodically a summary has to be printed and sent to the National Health System.

A third software, running on the Macintosh, considers data from patients who underwent coronary artery angioplasty (PTCA). It is a database application developed from a spreadsheet, where data structure has rows and lines. Data structure is in agreement with the recommendation of the National



PTCA Register, and includes personal data and parameters measured both before and after angioplasty. This way, a future statistical analysis about the efficiency of the angioplasty will be possible.

The last software, also running on the Macintosh, considers data from patients who underwent aortic valve balloon dilatation. Data relate to some haemodynamic parameters (aortic valve, cardiac output and others) measured a long time before, immediately before, immediately after, and a long time after balloon dilatation of the valve. All these data have already been used by statistical packages (StatView and StatWorks), in order to evaluate the efficacy of the treatment. This program was developed using the same tools as for data from PTCA patients.

### *3.4. Participation in national standardization protocols*

Some nation-wide studies concerning applications of informatic techniques to bio-medical fields and protocol-standardization are currently taking place. Our CCU is actually involved in three of them.

A first study considers the use of a drug (streptokinase) in patients affected by AMI: patients selected for this study are those to whom the drug has been administered within 6 hours from the onset of symptoms. To evaluate how effective this therapy is, a common data-collection protocol has been adopted. Data are collected onto traditional paper flowsheets: data are then sent to a pharmacological research institute collecting them from many centers.

A second study concerns data collection of AMI patients admitted to CCU. This study involves all the 45 CCU of our region. A common data collection protocol has been adopted and implemented on a PC. A young physician collects information from traditional paper flowsheet, and puts them into the PC. Acquired data are then regularly sent to a collection center, for the final elaboration [8]. Data transmission to the collection center takes place by mailing floppy disks four times per year. Moreover, there are seven supervising physicians, whose tasks are those of solving local problems.

A third study aims to provide a common protocol for data-collection of patients admitted to cardiology departments. This protocol contains all the information traditionally put into flowsheet, including CCU data, haemodynamics, traditional ECG, Holter exams, echographic analysis, heart-surgery data, therapies, generic exams. The efforts necessary to unify this protocol are being accomplished by the Italian Hospital Cardiologist Society, ANMCO.

The common protocol is made of different sub-protocols: each of them will be specific for a single sub-set of variables, like haemodynamics, pace makers etc.. This structure is suitable since some hospitals have only certain labs (pace makers) and lack some others (e.g. haemodynamics). Recently a wide agreement about protocols has been achieved and a minimum data set has been defined.

#### 4. Conclusions

Before the arrival of the bio-medical engineers, the number of physicians and nurses that were using computer systems was small. The only system that was really used in everyday practice was the monitoring system of the CCU: the reason was that data acquisition was completely automatic, and results (time plots, histograms) could be obtained by pressing a single key on the monitor unit. On the other hand, nurses were using only the CCU system, too. Only some of the available performances were used, and the usage level was thus quite low, in respect to the possible performances offered by the system.

By the presence of bio-medical engineers, the number of physicians using the CCU system has increased, and more physicians became more familiar with computers: some of them are now using PCs in the department for report writing, discharge letters, data management (e.g. PTCA, aortic valve dilatation). Those who already used computers, are now using them better.

Some nurses are using PCs, too. They use PCs for warehouse management, letter and report writing, management of drugs present inside the Department.

More use of PCs for database management has started. It is our opinion that these changes are due both to a deeper knowledge coming from education and from the full-time presence of bio-medical engineers.

We can say that much improvement has been seen, although we believe that we are still far away from the optimal situation.

One reason is that personnel do not themselves take care of the computer instruments. The presence of bio-medical engineers is really fundamental. Without an appropriate training course and continuous on-site training and education, we cannot be sure that the instruments are used in the correct way. Obviously, any misappropriate use of the instruments could lead to potentially dangerous situations.

#### References

1. Pincirolì F, Anderson J. 1981 *Changes in Health Care Instrumentation due to Microprocessor Technology*, North Holland, Amsterdam
2. Rienhof O, Abrams M. 1981 *The Computer in the Doctor's Office*, North Holland, Amsterdam
3. Invernizzi G, Pincirolì F, Pozzi G. 1989 Continuing Medical Informatics Education in the Practice of a Clinical Cardiology Department, *Medical Informatics and Education Int Conf* May 14-19, Victoria BC, Canada, 346-349.
4. Moher J, Protti D. 1989 *Proceedings of Medical Informatics and Education Int Conf* May 14-19, Victoria BC, Canada.
5. Invernizzi G, Pizzuto R, Pozzi G. 1989 Computerized Management of a Coronary Care Unit. In *Medicina e Informatica* 6: n. 2, 142-145.
6. Marchesi C *et al.* 1984 Computer System for Analysis of ST Segment Changes on 24 Hour Holter Monitor Tapes. In *Journal of American College of Cardiology* 4: n. 2, 245-252.
7. Kontron Instruments, *The Cardio 500 User Manual*.
8. Bossi M, De Vita C. 1989 Implementation of Regional Registry of Acute Myocardial Infarction (RIMAR Study), *IEEE Annual Conference*, Seattle, WA, pag. 173-174.

## **PART THREE**

### **Subject-oriented databases**

# The ARTEMIS data and knowledge base for hypertension

PATRICE DEGOULET<sup>1</sup>, MARION LAVRIL<sup>1</sup>, PIERRE-FRANÇOIS PLOUIN<sup>2</sup>, GILLES CHATELLIER<sup>2</sup>, JOËL MÉNARD<sup>2</sup>

<sup>1</sup> *Medical Informatics Department, Broussais University Hospital, 96, rue Didot, 75014 Paris, France,*

<sup>2</sup> *Hypertension Clinic, Broussais Hospital, Paris, France*

## Abstract

In a medical care unit, computerized programs can be used to memorize patients' individual records and profiles, to facilitate patient management and follow-up, to store medical knowledge and to provide facilities for decision making at the level either of the individual patient or of the population followed up. An integrated approach, progressively implemented in the ARTEMIS system since 1975, is described for the computerized management of hypertensive patients. The methodology used integrates data and knowledge management facilities into the same software. Five hypertension clinics are presently using the system in France and more than 25000 records have been registered. A simplified version of the ARTEMIS system, called ARTEL, has been made accessible through the French videotex system and is presently tested by two groups of general practitioners. Patient database interrogation can be used to evaluate the sensitivity and specificity of various signs and symptoms for the diagnosis of secondary hypertension, and to predict, for each patient, his/her cardiovascular risk, the risk of drop-out, the risk of insufficient blood pressure control and the probable blood pressure level. An evaluation of the diagnostic performances of the expert system has been performed on 7020 patients records stored in the ARTEMIS database. Agreement between the ES proposals, the expert initial decisions and the expert final diagnosis was evaluated using the Kappa coefficient, sensitivity and specificity. The predicting values for cases of secondary hypertension were calculated for the ES proposals, separately or combined to the expert proposals. Optimization of the decision's threshold of the ES was performed on a randomized learning sample of 3510 patients and found to be robust on a test sample. The results suggest that a strategy combining data and knowledge management might help the physician to supplement theoretical knowledge derived from the academic environment, and in some cases to replace it by a more pragmatic knowledge derived from his experience stored in the computer.

## 1. Introduction

Hypertension management is characterized by the large number of patients concerned, the numerous items to be recorded for each patient and the unlimited length of the follow-up. In this domain, as in many other chronic diseases like coronary heart disease, diabetes or asthma, computer techniques proved useful [1, 2]. The computer can be used as a tool for patient care, clinical research, education or for various managerial purposes. Although the initial programs developed in the early seventies were clearly focused on a strictly limited number of objectives [3, 4], recent experience has shown the advantages of a more integrated approach [5]. In this paper we describe the overall application of ARTEMIS with special emphasis on the integration of data and knowledge management facilities and the optimization of the performances of an expert system from cases extracted of a database.

## 2. Materials and methods

### 2.1. ARTEMIS development

ARTEMIS is the generic name given in 1975 to a computerized system for the medical management and follow-up of hypertensive patients [7, 8]. The first version, operational on a DIGITAL PDP10 main-frame computer, was implemented in September 1975 in the hypertension clinic of the Saint-Joseph Hospital, a non profit-making private hospital in Paris, and extended in 1979 to the Broussais Hypertension Clinic, a public university teaching hospital, also located in Paris. In October 1983, the ARTEMIS system was transferred from the main-frame computer to a mini-computer (DIGITAL VAX) which was located in the Broussais hypertension clinic and used the LIED temporal database management system (DBMS) described in [9]. Between 1985 and 1988 ARTEMIS was also implemented in three nephrology departments in the general hospitals of Charleville-Mézières, Colmar and Macon [10].

Together with the development of the clinical database, it was decided in 1983 to develop an expert system (ES) on hypertension with its knowledge base, using the SAM expert system development tool [11, 12]. The operational version of the ES was progressively integrated into the LIED DBMS [6, 12].

Finally a simplified version of ARTEMIS was progressively designed from 1987 for use by general practitioners (GP) through the French videotex telecommunication network. This version, called ARTEL (the acronym for ARTEMIS MINITEL), is in the process of evaluation by a group of voluntary GPs of the areas of Colmar and Macon [13].

### 2.2. The ARTEMIS database

The ARTEMIS patient database contains (January 1990) 750 different items

which can be completed during the patients' visits to each hypertension or nephrology department for either inpatient or outpatient care. Information recorded includes items concerning their past medical history, initial check-up, inpatient and outpatients visits, main investigations performed, search for drug contraindications, and the complete history of drug prescriptions and side-effects. Detailed questionnaires have been designed for patients with secondary hypertension. Historical information (e.g. clinical or biological) is permanently kept in the patient's record. Unlimited free text can be added to any item, whatever its data-type (e.g. numeric, date or string) and therefore complete standardized and structured input. This facility was found particularly relevant for storing the written summary of staff discussions.

Data are interactively entered by the end-users (i.e. physicians, nurses and secretaries) via video terminals (roughly one video terminal per outpatient consultation office, one per 10 beds of inpatient care, one per head nurse and one per secretary). Depending on the clinical condition of the patient, different updating or interrogation views are proposed to the end-user. Each user (or group of users) is able to define his/her own set of views. During database updating, questions proposed by the computer depend on previous information already entered and/or stored. Depending on physicians' training with terminal, free text notes are either entered directly or dictated for secondary entry. Help screens concerning protocols and procedures are available on line. Physicians have the ability to enter default values for the most current medical prescriptions.

Except for the observations made by medical students during periods of hospitalization, the traditional hand-written medical record has been completely suppressed and replaced by the various computer reports and/or displays.

Since ARTEMIS was implemented in 1975, more than 25000 computerized patient records have been created and updated by the five hypertension clinics which presently use the system. They contain summaries of more than 80000 outpatient and 23000 inpatient visits.

### *2.3. The ARTEMIS knowledge base and the expert system*

The ARTEMIS knowledge base includes both static and dynamic forms of knowledge. Static knowledge is represented in a semantic network in which the various medical concepts (e.g. simple facts such as symptoms or biological results, syndromes, diagnoses, investigations and therapies) are connected by links such as IS-A, BELONGS-TO, DEPENDS-ON, IS-RESULT-OF or ARE-EXCLUSIVE [11].

The dynamic reasoning of the expert is expressed by means of production rules. The rule is activated if the conditional part is true or if it has a sufficient degree of certainty. The action part of a production rule either adds new elements to the factual database or modifies the degree of certainty of an existing element. Certainty factors, similar to the MYCIN ones [14] can be

linked to production rules. A more extensive description of the knowledge representation model is given in [11, 12]. The present version of the ARTEMIS knowledge base (January 1990) includes more than 700 semantic and production rules of which 200 concern the diagnostic hypotheses, 120 the investigations to be performed and 280 treatments.

#### *2.4. Data and knowledge processing*

Data are processed by the LIED DBMS, which is based on a semantic and temporal data model described in [9]. The LIED data manipulation language is used to provide intelligent data entry (e.g. constraint checking and context-dependent branching in questionnaires) and to produce different reports to be kept in the patient record, sent to the referring practitioner or given to the patient. A plain text report, combining data entered in standardized format and in free text, is produced at the end of each visit for inpatient or outpatient care. Automatic alarms (e.g. biological values out of limits or drug dosage in excess) and different tabular reports (e.g. flow charts of main clinical and biological results) and/or cumulative reports (e.g. summary of main drug contraindications and side-effects) are provided on-line to facilitate the physician decision process and adherence to protocols. To encourage patients' participation and compliance, duplicates of the records are given to the patients and personalized recall are automatically mailed before long-term appointments [8].

Knowledge management and inferences are provided by the SAM essential expert system described in [11, 12]. The inference engine uses both forward and backward chaining. It also handles two kinds of logic which make it possible to conduct two parallel modes of reasoning: a conventional propositional logic and an approximate reasoning, whereby accumulation of evidence for or against a decision permits final conclusions to be reached. For each ES decision, a Global Certainty Factor (GCF), continuously varying from 0 (false or not indicated) to 1 (true or indicated) is produced by the ES. The database and knowledge base management parts of ARTEMIS are connected through a working memory allowing bidirectional exchanges between the ES and DBMS [15].

### **3. Results**

#### *3.1. Quality of the computerized medical record*

The quality of the medical records was evaluated in 19,601 patient records selected on the following criteria: 1) Patients referred for the first time to the Broussais or Saint-Joseph between January 1st 1976 and December 31, 1987; 2) Age of at least 13 years; 3) At least one past or present blood pressure value above 140 mmHg for systolic pressure or 90 for diastolic pressure [17]. Answer

rates to 12 mandatory questions regarding past history and examination at first visit were calculated for the three four-year periods that started in 1976, 1980 and 1984. They all exceeded 90% over the three periods. The best answer rates (mean  $\pm$  SD = 99.1  $\pm$  0.8) were observed in the last period corresponding to the dissemination of terminals all over the clinic and to increasing direct data entry on terminals by the end-users.

### *3.2. Initial patients' characteristics*

Analysis of the characteristics of the patients at their initial visit shows a decreasing percentage of patients referred by general practitioners, and increasing percentages referred by specialists or other hospital departments, or came of their own volition [17]. Patients examined during the period 84-87 tended to have more severe hypertension than those seen between 76-79. This was evidenced by the higher blood pressure values and the larger percentage of patients under anti-hypertensive drugs. The percentage of tobacco smokers and alcohol drinkers at first visit tended to decrease significantly by contrast with the increasing percentage of ex-smokers [17].

### *3.3. Evaluation of the performance of the expert system*

Evaluation of the diagnostic performance of the ES has been performed on 7020 hypertension cases stored in the ARTEMIS database [18]. All patients had been examined between January 1st 1981 and December 1988 at the Broussais Hospital hypertension clinic. All patients had at least one hospitalization stay between 1981 and 1988 in which the primary or secondary nature of hypertension could be stated. They were 6573 cases of primary hypertension (93.6%), 42 of phaeochromocytoma (0.6%), 242 of renal artery stenosis (3.2%) and 99 of primary hyperaldosteronism (1.3%). The items recorded during the first visit to the clinic (history and clinical examination), the levels of glycemia, serum potassium, creatinine, cholesterol and triglycerides and the EKG results constituted the initial set of facts addressed to the ES (100 items). Specialized investigations which would have made the diagnosis obvious were excluded.

The 7020 cases were randomly divided into a learning sample of 3510 cases and a test sample of 3510 cases. The diagnoses proposed by the ES were compared to the initial diagnosis proposed by the expert at the end of the first visit (without awareness of the proposition of the ES) and to the final conclusion reached after the inpatient hospital workup. The diagnoses suspected by the expert after the first visit were also compared with the final diagnoses. Concordance between the expert and ES diagnoses were estimated by the Kappa coefficient [16]. Since the global certainty factor (GCF) produced by the ES continuously varies from 0 (rejected) to 1 (certain), different thresholds for the GCF<sup>1</sup> were used in order to optimize the concordance

<sup>1</sup> If for a diagnosis  $x$  the GCF is superior to the threshold the ES is considered to suggest  $x$ .



Table 1. Evaluation of the ES - 7020 hypertension cases

Diagnoses	$\Delta \uparrow$ proposed by the ES*			$\Delta$ suspected by the expert			$\Delta$ proposed by the ES		
	K	Se %	Sp %	K	Se %	Sp %	K	Se %	Sp %
Essential hypertension (n = 6573)	0.10	71	56	0,21	77	74	0.13	73	41
Phaeochromocytoma (n = 42)	0.04	17	96	0,07	26	97	0.26	30	97
Renal artery stenosis (n = 242)	0.25	39	95	0,39	68	94	0.43	42	97
Primary aldosteronism (n = 99)	0.18	17	99	0,27	38	98	0.10	08	99

$\Delta$ : = Diagnoses ES: = Expert system.

K = Kappa coefficient, Se = sensitivity, Sp = specificity.

between the expert and the ES. Optimization based on the learning set was found to be very robust when checked on the test sample [18] and a global estimation of the decision threshold of the ES could therefore be calculated on the overall set of 7020 cases (Table 1). After optimization, Kappa coefficients between the diagnoses suspected by the ES and the expert vary from 0.13 for essential hypertension to 0.43 for the renal artery stenosis (cf column 3 of Table 1). Concordance between the diagnoses suggested by the ES and the final diagnosis of the expert were low (cf. column 1), but also the concordances between the initial and final diagnoses of the experts (column 2).

Table 2 shows the sensitivity and the specificity of the ES diagnosis by using the final expert diagnosis as the gold standard (post hospitalization diagnosis). When considered independently, the expert has a better predictive value than the ES (cf column expert alone and ES alone of Table 2). When the expert and the ES both agree on a diagnosis of secondary hypertension, the probability to have a secondary hypertension is much higher than when each 'partner' is considered separately. Both the expert and the ES have a good negative predictive value for secondary hypertension. The expert system is susceptible to discover hypertension cases not suggested by the human expert (cf. column P/n.y of Table 2)

#### 4. Discussion

Since the ARTEMIS system was implemented in 1975, it has undergone many technical and organizational changes. Starting from a record management program in batch mode on a main-frame computer, ARTEMIS has, like many similar products, progressively evolved to a decentralized fully interactive system, implemented on a dedicated mini-computer. Knowledge management capabilities have been progressively added to more traditional data management

Table 2. Evaluation of the predictive value of the Expert system - 7020 hypertension cases

Diagnoses	Expert alone			Expert syst. alone			Expert and expert system		
	P %	P/y %	P/n %	P/y %	P/n %	P/y.y %	P/y.n %	P/n.y %	P/n.n %
Essential hypertension	93.6	97.7	81.9	95.9	88.5	98.6	95.6	86.6	75.2
Phaeochromocytoma	0.6	4.5	0.5	2.7	0.5	6.9	3.5	1.1	0.4
Renal artery stenosis	3.4	30.4	1.2	21.8	2.2	35.2	26.9	6.8	1.0
Primary aldosteronism	1.4	22.5	0.9	21.3	1.2	57.1	19.3	13.6	0.8

P = Prior probability of the diagnosis on the population of 7020 hypertension cases

P/y = Posterior probability of the diagnosis if proposed.

P/n = Posterior probability of the diagnosis if not proposed.

P/y.y = Posterior probability of the diagnosis if proposed both by the expert and the ES.

P/y.n = Posterior probability of the diagnosis if proposed by the expert only.

P/n.y = Posterior probability of the diagnosis if proposed by the ES only.

P/n.n = Posterior probability of the diagnosis if proposed neither by the expert nor by the ES.

capabilities. However data entry still constitute one of the factors limiting the acceptability of such systems by the end-users. Although it is true that most physicians agree to interact directly with ARTEMIS when entering information concerning follow-up visits in the presence of the patient, or hospital in-patient check-ups, one must admit that for the first visit of a newly referred patient, most physicians still prefer to complete a standardized input form for subsequent entry by secretaries. In ARTEMIS, as it was previously observed with self-administered questionnaires, branching techniques have been found to reduce the effort devoted to data entry drastically. On-line checking of constraints and on-line helps improve medical record consistency and database integrity. In any case, the physical appearance of present terminals (i.e. 24 lines by 80 columns) and keyboards will remain a limitation until technology which permit at a reasonable price the combination of various inputs and outputs including pointing devices, images, graphics, texts and voice [19, 20].

Electronic record quality assessments performed just after the implementation of a computerized system have shown high response rates in questionnaires completed by physicians [3, 7]. Observations from the ARTEMIS database confirm that a high standard quality can be maintained over time [17]. Immediate alarms prevent the end user from forgetting important elements during data entry. However standardization of input procedures does not completely suppress inter and intra-physician variability [7]. In addition, the high response rates we observed only apply to the subset of questions selected from the ARTEMIS system, and the reliability and subject matter of free text that physician and nurses can add to standardized questionnaires should be compared to the quality of paper medical notes. New questions have been permanently added to the ARTEMIS database since its first implementation in 1975.

Medical decisions are based on three kinds of information and knowledge: 1) Specific information on the patient; 2) General or academic knowledge which can be derived from textbooks and periodicals; and 3) and knowledge based on physician's personal experience. Patient specific information is mainly contained in or derived from the medical records. Experiences of the two last decades has shown that computer-stored medical records can solve many of the problems of availability, retrievability, legibility and organization of the paper medical record. Access to the electronic record is immediate and several users can share the same record. Physicians and/or the patients can access the medical records at home or at areas remote from their working environment [20, 21]. Information, when stored in an appropriate format, can be displayed or retrieved in many different ways according to the physician's needs (e.g. flow-sheets, graphical displays, time-oriented or problem-oriented record summaries) [22, 23]. Well designed computerized summaries can provide more information than the standard medical record and improve the clinical decision process [24]. In the ARTEMIS system the multiplication of views of the same database and the adaptation of views to the physician's environment were found useful in enhancing flexibility and physician appropriation of the software. It could however in the long run lead to non convergent use of the system with the secondary effect of increasing inter-physician variability.

A large number of simple decisions can be made easily thanks to the data manipulation language of a database management system or to the deductive capabilities of an expert system. This 'data-driven' approach has been extensively used in ARTEMIS as in other data management systems for hypertensive patients [25, 26, 27, 28, 29]. Examples of such decisions reached during the management of individual cases are the conversational checking of constraints (e.g. the verification of the range of quantitative values), the continuous recording of drug intolerance history and the provision of adequate alarms (e.g. drug allergy or contra-indication), the automatic indication of simple follow-up tests [25] and the automatic dispatching of recall letters [8, 28]. On-line helps and reminders enhance physicians' adherence to protocols and directly participate to a high quality standard of care [30, 31]. More complex protocols can be implemented for the calculation of the optimal dosage of a prescribed drug [4], or, in privileged situations, for the direct control of medication administration such as the rate of nitroprusside perfusion in malignant hypertension [32]. These applications, whose impact on the quality of care has been confirmed [3, 28, 29], should be pursued in the coming years.

Approaches integrating data and knowledge management facilities, such as the one used in ARTEMIS, might enhance end-user acceptability and overall system performance. Each item of information only needs to be entered once during the follow-up of a patient and suggestions by the computer are byproducts of monitoring or data-management facilities [33]. Intelligent data and knowledge driven questionnaires can facilitate the entry of patient history or examination [34]. Extensive interrogation of the database might help to build and/or test medical knowledge contained in the knowledge base [5, 35]. The

specificity and sensitivity of each symptom and investigation can be calculated as regards diagnostic, therapeutic or prognostic decisions, as it was done in ARTEMIS for the diagnosis of pheochromocytoma [36] and integrated in the rules of the expert system. Multivariate statistical methods can be used for the calculation of various risk indexes such as coronary risk, risk of poor compliance or risk of insufficient blood pressure control [37, 38, 39]. Predictions made from these statistical models can be more accurate than predictions of experienced clinicians made from detailed case summaries [39]. They might help the physician to focus his attention to the high risk patients [37] and should be included in the knowledge base. It seems also feasible, as in our evaluation of the ES, to optimize the thresholds used by the ES to propose a decision from the cases contained in the database. The ES should therefore be considered as another complementary exam, with the particular property that its performance might increase with time (in parallel with the increase of the number of stored cases) and its cost might decrease (i.e. reduced to the cost of the updating of the knowledge base and of some additional computer power).

However, several barriers must be recognized before more widespread use of integrated expert database systems can be proposed. The order in which investigations are suggested by an expert system should be compared to the attitudes of various specialists, as in the critiquing approach developed by Miller and Black with ATTENDING [40, 41]. The knowledge base of the ES needs to be validated in different medical environments, where the ES might be even more useful than in the present specialized hypertension clinic (e.g. in general practice for which the simplified ARTEL system was designed [13]). Explanations provided by the ARTEMIS expert system, as many other ES, are still rudimentary and considerable efforts are needed to allow the physician to determine if computer suggestions are valid or not [33, 42]. Finally, although the ARTEMIS patient database is used to evaluate and optimize the performances of the ES, the ES itself does not fully profit of all the information which is presently stored in the database. Only in this way can the theoretical knowledge derived from the academic environment be gradually replaced by the pragmatic knowledge which can be derived from the experience stored in electronic medical records.

### **Acknowledgments**

The development of LIED is supported by Assistance Publique de Paris. Developments of ARTEMIS and SAM were supported by Ministère de la Santé, Agence de l'Informatique, INSERM and CNAMTS.

## References

1. Pryor DB, Califf RM, Harrel FE *et al.* (1985) Clinical databases. Accomplishments and unrealized potential. *Medical Care* 23: 623-47
2. McDonald CJ, Tierney WM (1988) Computer-stored medical records. Their future role in medical practice. *JAMA* 259: 343-340
3. Bulpitt CJ, Beilin LJ, Coles EC *et al.* (1976) Randomised controlled trial of computer-held medical records in hypertensive patients. *Br Med J* 1: 67-79
4. Coe FL, Norton E, Oparil S *et al.* (1977) Treatment of hypertension by computer and physician – a prospective controlled study. *J Chron Dis* 30: 81-92
5. Degoulet P, Devriès C, Sauquet D (1988) Data and knowledge management integration in hospital information systems. In: Bakker A, Ball M, Scherrer JR, Willems JL (eds). *Towards New Hospital Information Systems*. Amsterdam: North-Holland, 149-156
6. Devriès C, Degoulet P, Jeunemaitre X *et al.* (1987) Integrating management and expertise in a computerized system for hypertensive patients. *Nephrol Dial Transplant* 2: 327-331
7. Degoulet P, Ménard J, Berger C *et al.* (1980) Hypertension management: the computer as a participant. *Am J Med* 68: 559-566
8. Degoulet P, Vu HA, Chatellier G *et al.* (1982) Hypertension management: the role of the computer in improving patient compliance. *Med Inform* 7: 49-55
9. Degoulet P, Devriès C, Rioux P *et al.* (1986) LIED: a temporal database management system. In: Salamon R, Blum B, Jorgensen M (eds). *MEDINFO 86 Amsterdam*: Elsevier North Holland. 53-26
10. Marichal JF, Devriès C, Aimé F *et al.* (1987) Expérience de décentralisation du système 'ARTEMIS' dans un service de Néphrologie en Hôpital Général. *Arch Mal Coeur et Vaiss* 80: 888-891
11. Gascuel O (1985) Un système expert pour la réalisation de diagnostics. *Techniques et Sciences Informatiques (TSI)*. 4: 359-72
12. Morice V, Degoulet P, Jeunemaitre X *et al.* (1986) Hypertension management: connection of an expert system to the ARTEMIS patient database. In: *Objective Decision Making. Lecture Notes in Medical Informatics*. Berlin: Springer-Verlag 28: 55-61
13. Lavril M, Chatellier G, Degoulet P *et al.* (1988) ARTEL: An expert system in hypertension for the general practitioner. *Lecture Notes in Medical Informatics*. Berlin: Springer Verlag 36: 314-321
14. Shortliffe E (1976) *Computer based medical consultations: MYCIN*. New York: American Elsevier
15. Sauquet D, Devriès C, Lavril M *et al.* (1988) Data and knowledge management with LIED. In: *Proc Medical Informatics Europe, MIE 88*. Rome: Edi-Press I: 574-579
16. Fleiss JL (1973) *Statistical Methods for Rates and Proportions*. New York: Wiley. 143-147
17. Degoulet P, Chatellier G, Devriès C, Lavril M, Ménard J (1990) Computer-assisted techniques for evaluation and treatment of hypertensive patients. *AJH* 3: 156-163
18. Lavril M, Degoulet P, Poutchnine P, Chatellier G, Jeunemaitre X, Aimé F (1989) La base de données ARTEMIS pour l'évaluation d'un système expert en hypertension artérielle. *Informatique et Santé*. Springer-Verlag. 1: 137-146
19. Lind M, Pettersson E, Sandblad B, Schneider W (1988) Computer based workstations in health care. In: Bakker A, Ball M, Scherrer JR, Willems JL (eds). *Towards New Hospital Information Systems*. Amsterdam: North-Holland 235-242
20. Smith MB, Burke KE, Torgerson JS, Stollerman JE, Kern DC, Hardy WL, Friedman RH (1988) Logical and efficient conversation between patients and the telephone linked computer system. In: *Proc 12th Annual Symposium on Computer Applications in Med Care*, Washington, DC. 463-467
21. Darnell JC, Hiner SL, Neill PJ *et al.* (1985) After-hours telephone access to physicians with access to computerized medical records. *Medical Care*. 23: 20-26
22. Hammond WE, Stead WW (1986) The evolution of a computerized medical information

- system. In: Proc 10th Annual Symposium on Computer Applications in Med Care, Washington, DC. 147-156
23. McDonald CJ, Blevins L, Glazener T *et al.* (1983) Database management, feed back control, and the Regenstrief medical record. *Med Sys* 7: 11-125
  24. Whiting-O'Keefe QE, Simborg DW, Epstein WV, Warger A (1986) A computerized summary record system can provide more information than the standard medical record. *Jama* 254: 1185-1192
  25. Evans AR, Wilkes ER, Absolon PJ *et al.* (1985) An interactive computerised protocol for the management of hypertension. *Meth Inf Med* 24: 21-6
  26. Laurent D, Mashruwala MD, Lucas CP (1980) A computerized data-handling system in hypertension management. *Arch Intern Med* 140: 345-50
  27. Wilson DH (1980) Patient registries, interviewing, scheduling, and management. Ten years of experience using computers in a specialized clinic. In: Rienhoff O, Abrams ME (eds). *The computer in the Doctor's Office*. Amsterdam: North-Holland. 205-11
  28. Barnett GO, Winickoff RN, Morgan MM, Zielstorff RD (1983) A Computer-Based monitoring system for follow-up of elevated blood pressure. *Medical Care* 21: 40-09
  29. McAlister NH, Covvey HD, Tong C *et al.* (1986) Randomised controlled trial of computer assisted management of hypertension in primary care. *Br Med J* 293: 67-04
  30. McDonald CJ, Hui SL, Smith DM *et al.* (1984) Reminders to physicians from an introspective computer medical record. *Ann Intern Med* 100: 130-138
  31. Pryor TA, Gardner RM, Clayton PD, Warner HR (1983) The Help system. *J Med Systems* 7: 87-102
  32. Meline LJ, Westenskow DR, Pace ML (1985) Computer controlled regulation of sodium nitroprusside infusion. *Anesth Analg* 64: 38-42
  33. Shortliffe EH (1987) Computer programs to support clinical decision making. *JAMA* 258: 61-66
  34. Haug PJ, Warner HR, Clayton PD *et al.* (1987) A decision-driven system to collect the patient history. *Comp Biomed Res* 20: 193-207
  35. Bouhaddou O, Haug PJ, Warner HR (1987) Use of the Help clinical database to build and test medical knowledge. In: Proc 11th Annual Symposium on Computer Applications in Med Care, Washington, DC. 64-67
  36. Plouin PF, Degoulet P, Tugayé A *et al.* (1981) Le dépistage du phéochromocytome: chez quels hypertendus?. Etude sémiologique ches 2585 hypertendus dont 11 ayant un phéochromocytome. *Nouv Press Med* 10: 869-872
  37. Degoulet P, Ménard J, Vu HA *et al.* (1983) Factors predictive of attendance at clinic and blood pressure control in hypertensive patients. *Br Med J* 287: 88-93
  38. Pryor DB, Harrell FE, Lee KL *et al.* (1983) Estimating the likelihood of significant coronary artery disease. *Am J Med* 75: 771-780
  39. Lee KL, Pryor DB, Harrell FE *et al.* (1986) Predicting outcome in coronary disease. Statistical models versus experts clinicians. *Am J Med* 80: 553-560
  40. Miller PL, Black HR (1984) Medical plan-analysis by computer: critiquing the pharmacologic management of essential hypertension. *Comp Biom Res* 17: 38-54
  41. Miller PL, Blumenfrucht SJ, Black HR (1984) An expert system which critics patient workup: modeling conflicting expertise. *Comp Biom Res* 17: 554-69
  42. Degoulet P (1987) Artificial intelligence: its use in nephrology. *Nephrol Dial Transpl* 2: 298-303

# Databases for prevention, pacemaker, and postoperative treatment

## *The Charité Experience in Cardiology*

K.H. GÜNTHER

IN COOPERATION WITH R. BOHM, V. DILBA, R.R. MOEBES, ST. PETZSCHMANN, P. PIORKOWSKI, J. WITTE

*Humboldt University Hospital Charité Berlin, Department of Internal Medicine / Cardiology and Department of Medical Informatics, Cardiovascular Disease Dispensary Cottbus-Land, Germany (DDR)*

### **Introduction**

The Charité in Berlin, founded in 1710, is the Hospital of Humboldt University. Its responsibilities are great concerning basic, emergency and specialized medical care, student education, postgraduate teaching and research in many fields. In fact, it is a leading centre in this country, cooperating with a variety of medical institutions of all kinds. A lot of computer applications, ranging from databases to highly sophisticated techniques such as computer tomography and magnetic resonance imaging, have been established. Preliminary steps were also taken in ECG processing with the establishment of three centralized ECG labs during reconstruction; the labs are usable for all surgical disciplines, all outpatient departments and the purposes of internal medicine.

Until recently experience could be gathered with databases in three quite different fields: (1) in 1963, a registry for pacemaker patients and devices was initiated which became computer-based later on. (2) In 1978, a community-oriented risk factor database was implemented that is still usable scientifically and as a basis for comparison with other preventive studies based on the same methodology and criteria. (3) Since 1987, a registry for patients after aorta-coronary bypass surgery has been run with the main aim of improving secondary prevention, following the activities of patients after myocardial infarction that are generally accepted, nationally and internationally. The common advantage of these databases is that they are the result of a close cooperation with partners outside the university.

They are aimed at controlling and steadily improving medical treatment, giving examples for other tasks to be fulfilled. And, they are in principle open for suitable national and international coordination, which is already going on in the field of pacemaker therapy.

## **The Cottbus General Practitioners Preventive Study (CGPPS)**

Prevention of chronic non-communicable diseases is generally based on risk-factor concepts. Morbidity and mortality due to the diseases can effectively be influenced only by integrated action within the public health services. Theoretically guided by epidemiological intervention studies, control programmes have to be implemented regionally to

- detect and monitoring the prevalence and incidence of risk factors and leading manifestations, as well as
- check the results of preventive activities, gender- and age-related, and evaluate the intervention methods.

Prerequisites thus are some standardization of the methods and criteria as well as uniform acquisition and documentation of the data obtained, which can only be managed by using a computerized database. This is underlined by the fact that conventional documentation of data in the general medical practice seems scarcely suitable for estimation on a higher level than individual.

### **Methods and material**

In close cooperation with the District Cardiologist of Cottbus-Land, situated between Berlin and Dresden, a data sheet (attached) was designed that was used by 24 general practitioners in their daily practice to offer a preventive examination to all people aged between 30 and 50 years who were coming for different reasons. Everybody whose data sheet indicated an elevation of risk factors, especially smoking, hypertension or hypercholesterolemia, was advised, verbally and by printed leaflets. Re-examinations were offered after 3 months and 18 months. Intervention was to be done mostly nonpharmacologically, also for blood pressure elevation. The intention was to show that coronary risk factors can be screened, managed and followed-up uniformly within a general practice.

The data originally obtained by the GPs and their nurses were checked on the sheets by the District Cardiologist and transferred to the regional computer centre in the County Hospital, and after rough evaluation and processing were given to the Department of Medical Informatics of the Charité in Berlin. Most important results were given back soon after screening to the GPs for comparison and competition. Other data processing, using more sophisticated hardware and software for scientific purposes, is still going on.<sup>1</sup>

<sup>1</sup> Software: SPSS, hardware: ES 1055, OS/VS2.



## **Experience and selected results**

The database created more than ten years ago served as an example, at least for practical purposes, for other community-based preventive studies that were in the meantime performed in Havana, Cuba, and in another district of this country, near Potsdam (Berlin). The data obtained according to the same recommendations and quality control are now stored and available at the Charité in Berlin. Data processing was done or is still going on in the following scientific areas:

- 1) After getting more precise information locally on sex- and age-related risk-factor prevalences, the efficacy of nonpharmacologic treatment could be assessed, especially concerning short- and long-term blood pressure elevation.
- 2) After finishing the studies, prevention compliance of the patients and doctors could be revealed, again differentiated necessarily according to gender and age.
- 3) Then, in Cottbus, the database was used for comparison with self-reported risk factors from physicians, dentists and pharmacists living in this country as well as with screened data from educational personnel (teachers) working in schools and kindergartens, all being responsible for the implementation of prevention, starting with their own.
- 4) Another, highly scientific aspect is followed by comparing risk-factor prevalence and levels with social aspects such as professional education level, socio-economic status, marital status and shift-work. This seems important not only from the analytical point of view but also concerning the question of whether intervention may be dependent upon those factors.

Kartenzahlen  1 Arztpraxis-Nr.   6-7  
 Lfd.-Nr.    2-5 Datum der Befragung    8-13  
 Tag Mon. Jahr

### Vorsorgeuntersuchung gegen Herz-Kreislauf-Erkrankungen<sup>1</sup>

Beantworten Sie bitte nachfolgende Fragen sorgfältig durch Ankreuzen der entsprechenden Kästchen, weil davon die Entscheidungen des Arztes abhängen, die er für Sie im Interesse Ihrer Gesundheit treffen muß. Ihre Angaben werden vertraulich behandelt.

Name: \_\_\_\_\_ Vorname: \_\_\_\_\_

Wohnort, Straße: \_\_\_\_\_

Geburtsdatum:     14-19  
 Tag Mon. Jahr.

Geschlecht:  1  2 20  
 m. w.

1. Sind Sie berufstätig?  1  2 21  
 ja nein

2. Treiben Sie in der Woche mindestens 2 Std. Sport?  1  2 22  
 ja nein

3. Betätigen Sie sich wöchentlich mindestens 2 Std. durch Gartenarbeit, Radfahren, Spaziergänge oder andere körperliche Belastungen?  1  2 23  
 ja nein

4. Wieviele Zigaretten rauchen Sie täglich? (Angabe in Stück)   24-25

5. Wieviele Zigaretten haben Sie täglich vor 1 Jahr geraucht? (Angabe in Stück)   26-27

6. Rauchen Sie regelmäßig Zigarren oder Pfeife?  1  2 28  
 ja nein

7. Wieviele Jahre rauchen Sie schon oder haben Sie geraucht?   29-30

8. Fühlen Sie sich häufig überfordert? (im Beruf, in der Familie ...)  1  2 31  
 ja nein

9. Erkrankte oder starb jemand Ihrer Eltern oder Geschwister bis zum 55. Lebensjahr an Angina pectoris, Herzinfarkt, Schlaganfall, Bluthochdruck oder Diabetes?  1  2 32  
 ja nein

10. Wurde bei Ihnen eine Zuckerkrankheit festgestellt?  1  2 33  
 ja nein

11. Hatten Sie schon selbst einen Schlaganfall?  1  2 34  
 ja nein

12. Stehen Sie wegen eines zu hohen Blutdrucks in ständiger Behandlung?  1  2 35  
 ja nein

13. Haben Sie jemals irgendwelchen Schmerz oder ein Unbehagen im Brustkorb gehabt? (Wenn nicht, gehen Sie gleich zur Frage Nr. 21 über!)  1  2 36  
 ja nein

14. Bekommen Sie diesen Schmerz oder dieses Unbehagen, wenn Sie bergauf gehen oder schneller gehen?  1  2 37  
 ja nein

15. Bekommen Sie diese Beschwerden, wenn Sie normalen Schrittes zu ebener Erde gehen?  1  2 38  
 ja nein

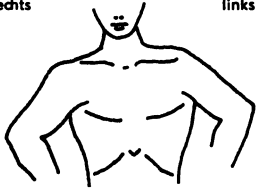
16. Wenn Sie irgendwelchen Schmerz oder ein Unbehagen in der Brust bekommen, was tun Sie dann?  
 Anhalten  1  
 langsamer gehen  2  
 weitergehen in gleichem Tempo  4 39

17. Geht es weg, wenn Sie anhalten?  1  2 40  
 ja nein

18. Wie schnell geht es weg?  
 Innerhalb von 10 Minuten  1  
 Nach mehr als 10 Minuten  2 41

19. Wo ist dieser Schmerz oder dieses Unbehagen? Machen Sie ein Kreuz an der Stelle (oder an den Stellen), wo Sie es spüren!

rechts links



20. Haben Sie jemals einen schweren Schmerz vorn in der Brust gehabt, der 1/2 Stunde oder länger anhält?  1  2 42  
 ja nein

21. Was haben Sie außerdem für Beschwerden?  43

<sup>1</sup>) Fragebogen nach K. H. Günther, P. Plorkowski und Renate Bohm, teilweise unter Anwendung des London School of Hygiene Chest Pain Questionnaire zur Selbstbeantwortung (WHO-Standard)

## Von Arzt und Schwester auszufüllen

22. Blutdruck mm/Hg	1. Messung . . . . .	systol.	<b>niedrigste Werte</b>			
	2. Messung . . . . .	diastol.	<input type="text"/>	<input type="text"/>	<input type="text"/>	44-46
23. Cholesterin mg %	1. Wert . . . . .		<input type="text"/>	<input type="text"/>	<input type="text"/>	50-52
	2. Wert . . . . .					
24. Harnsäure mg %	1. Wert . . . . .		<input type="text"/>	<input type="text"/>		53-54
	2. Wert . . . . .					
25. Körpergröße in cm			<input type="text"/>	<input type="text"/>	<input type="text"/>	55-57
26. Körpergewicht in kg			<input type="text"/>	<input type="text"/>	<input type="text"/>	58-60
<hr/>						
27. Vorstellung in der Abt. Herz-Kreislauf-Dispensare erfolgt?			<input type="text"/>	<input type="text"/>		61
			ja	nein		
28. Der Patient wurde bereits in der Diätsprechstunde betreut?			<input type="text"/>	<input type="text"/>		62
			ja	nein		
29. EKG pathologisch			<input type="text"/>			63
ja 1 nein 2 nicht erfolgt 3						
30. Belastungs-EKG pathologisch			<input type="text"/>			64
ja 1 nein 2 nicht erfolgt 3						
31. Diagnose:	<input type="text"/>	65-71				
- Ischämische Herzkrankheit	1					
- Herzinfarkt	2					
- behandlungsbedürftiger Hochdruck	3					
- Schlaganfall	4					
- Krankheiten der Arterien	5					
- Venenerkrankungen	6					
- angeborene und erworbene Herzfehler	7					
- pulmonale Herzkrankheit	8					
- funktionelle Herzkrankheit	9					
32. Qualifikation:	<input type="text"/>	72				
- Nichtfacharbeiter	1					
- Teilfacharbeiter	2					
- Facharbeiter	3					
- Meister und Techniker	4					
- Fachschulabsolvent	5					
- Hochschulabsolvent	6					
33. Tätigkeit:	<input type="text"/>	73				
vollbeschäftigt						
- ohne Schichtsystem	1					
- Schichtsystem bis 22.00 Uhr	2					
- Schichtsystem mit Nachtschicht	3					
teilbeschäftigt	4					
34. Familienstand:			<input type="text"/>			74
- ledig	1.					
- verheiratet	2					
- verwitwet	3					
- geschieden	4					
35. Sozial ökonomische Struktur	<input type="text"/>	75-76				
- Arbeiter in der materiellen Produktion	1					
- Arbeiter im nichtmateriellen Bereich	2					
- Angestellte (außer Intelligenz)	3					
- Intelligenz im Angestelltenverhältnis	4					
- Mitglieder der PGH	5					
- Genossenschaftsbauern						
a) überwiegend Mechanisatoren	6					
b) überwiegend Tierproduktion						
- in Kleinanlagen	7					
- in Großanlagen	8					
c) überwiegend Pflanzenprodukt.	9					
- Genossenschaftsbauern (außer 6-9) einschließlich GPG und FPG	10					
- Selbständige Gewerbetreibende und freiberuflich Tätige	11					
- Selbständige Handwerker	12					
- Hausfrauen	13					
- Schüler und Studenten	14					
36. Nummer des Arztes	<input type="text"/>	77-80				

Unterschrift des Arztes  
Namensstempel

### **The comprehensive country-wide pacemaker registry**

Pacemaker therapy is a widely used technical method of treatment which has proved to be very efficient in indicated cases, being however rather expensive. For creating a rational system right from the beginning compulsory registration has been demanded, including comprehensively each patient treated, the kind of arrhythmia that was the indication for implantation, the devices selected and the complications. The number of pacemaker clinics could be limited, starting with three (Berlin/Charité, Rostock and Leipzig) and there being now only 36 for the whole country, which has somewhat more than 16 million inhabitants. During the early years, the pacemakers needed were generally imported, calling for an increasing amount of foreign money, which was a problem. For more than ten years the demand has increasingly been fulfilled by an in-country factory, which has satisfied the demand, quantitatively and, recently, qualitatively as well.

### **Methods and material**

Registration had been initiated by using a data sheet especially designed in 1963. After getting an agreement within the Society of Cardiology and Angiology of the GDR in 1971 a revised version was introduced and generally accepted. Only minor changes were made in the following years, especially when computerization of the whole registry was established in 1979. The database is now compiled in dBASE III plus; the use of the more advanced dBASE IV language is in preparation. The database has been divided into two files, connected by the patient's identification number. The first file contains the fields of patient's identification, date of birth, gender and individual legal personal number, the larger second file coded fields (see sheet attached). This is because one patient may have several registrable operations or other coded follow-up events. The personal computer (IBM PS2/60) is used as central computer and works under the MS-DOS operating system with 640 kilobytes, 3,5 or 5,25 inch floppy disc, 60 megabyte disc storage and a streamer for saving the data regularly.

Each pacemaker clinic is using the same type of IBM-compatible PC. The data acquisition is carried out there via the computer keyboard. Always after three to six months the regional data are transferred on floppy disc to the central computer, being then available for comparison and self-criticism. The software for file-management of the database and data-processing is programmed in DBASE, particularly using a self-explanatory menu technique. The registration sheet serves for first implantation and follow-up events like re-operation or death.

The centralized database containing all information obtained serves

- as a tool for management control (hospital, factory, ministry);
- for checking aspects of implantation (indication, complication) regionally;

STATISTIKKARTE ZUR PERMANENTEN HERZSTIMULATION

ORZ-KUNDEN-NR.

1-3

Klinikstempel:

Unterschrift:

PATIENTEN-IDENTIFIKATION

Name

Vorname

Wohnort

Straße

PKZ - Geburtsdatum

Tag Monat Jahr 4 9

PKZ - Rest

10 15

Klinik-Nr.

8 9 4 5 16 17

lauf. HSM Pat.-Nr.

18 25

(d. Erstimpl. Kl.)

Kl. Nr. Lfd. HSM-Pat. Nr. PZ

Ereignis-Nr.

26-27

Ereignis-Datum

28-33 Tag Monat Jahr

DIAGNOSTIK

Erstimplantations-Indikation

- 0 Erstimpl. bereits erfolgt
1 Adams-Stokes-Syndrom
2 Bradykarde Herzinsuffizienz
3 Brady-Tachykardie-Syndrom
4 prophylaktisch
9 andere

34

EKG

- 0 kein Ereignis-EKG angefertigt
1 permanenter höhergradiger AV-Block
2 intermittierender höhergradiger AV-Block
3 Bradyarrhythmia absoluta
4 permanente supraventrikuläre Bradykardie
5 intermittierende supraventrikuläre Bradykardie
9 andere

35

Ereignisursache nach Erstimpl.

- 00 für I bis III bei Erstimplantation
01 für II bzw. II und III bei fehlender sek. bzw. tert. Ursache

I 36-37
II 38-39
III 40-41

kardiologisch

- 01 Parastolie
02 Reizschwellenanstieg (Exitblock)
03 Triggerungsausfall (Entranceblock)
04 Wegfall der Indikation zur permanenten Stimulation
05 Änderung der Indikation zur permanenten Stimulation
06 Fehltriggerung des HSM
19 andere kardiolog. Ereignisursachen nach Erstimpl.

chirurgisch

- 20 primäre Wundinfektion
21 beginnende Hautdrucknekrose im Bereich des HSM
22 Dekubitus im Bereich des HSM
23 Sepsis
24 beginnende Hautdrucknekrose im Bereich der Elektrode
25 Elektrodenfall an funktionierender Elektrode
26 Elektrodenfall an Elektrodenrest
27 Dislokation der Ventrikelektrode
28 Dislokation der Vorhofelektrode
29 Dislokation des HSM
30 HSM bis jetzt extern
31 Zwerchfellzucken
32 Muskelzucken
39 andere chir. Ereignisursachen nach Erstimpl.

technisch

- 40 Frequenzabfall des HSM
41 Frequenzanstieg des HSM
42 Steuerungsausfall des HSM
43 Totalausfall des HSM
44 HSM technisch überholt
45 Rißrüd des HSM durch Firma
46 Dehbüsekorrosion des HSM
47 relative Überalterung des HSM bei anderer Komplikation
48 Explant. des techn. intakten HSM bei anderer Komplikation
49 Ventrikelektrodendefekt
50 Vorhofelektrodendefekt
59 andere techn. Ereignisursachen nach Erstimpl.

organisatorisch

- 60 ins Ausland verzogen
61 an andere Impl. Klinik innerhalb der DDR gewechselt
69 andere organisat. Ereignisursachen nach Erstimpl.

Tod des Patienten

- 80 primär kardial bedingt
81 primär extrakardial bedingt
82 primär HSM - bedingt
83 primär Elektroden - bedingt
84 HSM-Operations-bedingt
85 unnatürlicher Tod
86 unbekannter Todesursache
87 unbekannter Verbleib des Patienten
88 Sektion durchgeführt
89 Teilsektion durchgeführt

OPERATION

Maßnahme I

- 0 keine Operation durchgeführt
1 kein neuer HSM/keine neue Elektrode
2 neuer HSM/neue Elektrode
3 neuer HSM/keine neue Elektrode
4 neue Elektrode/kein neuer HSM
5 nur Explantation des HSM
6 nur Erstimplantation der Elektrode

42

Implantationstechnik der Ventrikelektrode

Implantationstechnik der Vorhofelektrode

- 0 keine Ventrikel- bzw. Vorhofelektrode implantiert
1 V. cephalica ll./re.
2 V. jugularis externa ll./re.
3 V. jugularis interna ll./re.
4 V. subclavia ll./re. Punktion
5 atypische Vene
6 Thorakotomie
7 Pericardiotomie inferior
9 andere

43

44

Maßnahme II Ventrikelektrode

Maßnahme II Vorhofelektrode

- 0 bisher keine Ventrikel- bzw. Vorhofelektrode implantiert gewesen
1 keine Maßnahme
2 Lagekorrektur der Elektroden Spitze
3 Reparatur der Elektrode
4 komplette Entfernung der Elektrode
5 partielle Entfernung der Elektrode
6 Längenänderung der Elektrode
7 Steckanpassung der Elektrode
9 andere

45

46

HSM Lokalisation

- 0 kein HSM mehr im oder am Patienten
1 pectoral subkutan ll./re.
2 subpectorall ll./re.
3 abdominal subkutan ll./re.
4 abdominal subfaszial ll./re.
5 Rektusscheide ll./re.
6 extern
9 andere

47

HSM Umlagerung

- 1 nein
2 ja

48

Durchleuchtungszeit in min

- 00 DLZ nicht gemessen
01 DLZ bis 1 min
99 keine DL erfolgt

49-59

Anaesthetie

- 0 keine Anaesthetie
1 Lokalanesthetie
2 Maskenarkose
3 Intubationsarkose
9 andere

51

TECHNIK

HSM Ident-Nr.

52-64

Typ-Schlüssel-Nr. HSM-Serien-Nr. PZ

alles „0“ kein implant. HSM im Patienten weiteres siehe Hinweise zur Statistikkarte

Elektroden-Typ-Schlüssel-Nr. Ventrikel

65-67

Elektroden-Typ-Schlüssel-Nr. Vorhof

68-70

00 keine funktionsfähige (technisch und/oder biologisch) Elektrode

- 001 Elektrodenart und Hersteller unbekannt
002 Myokardelektrode ohne nähere Angaben
003 Endokardelektrode ohne nähere Angaben
weiteres siehe Hinweise zur Statistikkarte

Reizschwelle in zehntel Volt

bei 1,0 ms 71-72

bei 0,5 ms 73-74

bei 1,0 ms 75-76

Reizschwelle in zehntel mA

bei 0,5 ms 77-78

- 00 nicht gemessen
01 nicht meßbar wegen hoher Eigenfrequenz

Triggerpotential in mV

79-80

- 97 keine absolute Indikation zur Messung gegeben
98 nicht meßbar wegen fehlender Eigenaktion
99 nicht gemessen

- for comparing results with other centres abroad; and
- for calculating life expectancy and prognosticating causes of death.

### **Experience and selected results**

As of 1989 data from 43000 patients were stored in the database. Most valuable and consistent results can be provided from the period 1979-1987, the implantation rate being 3003-4273 pacemakers per year, i.e. 179.8-255.8 new implantations per 1 million inhabitants per year, for the whole country. Each pacemaker clinic implanted on average 95-126 devices per year in this period. The relative proportions of indications for pacemaker implantation remained rather constant: Adams-Stokes Syndrome in 73.0% of the cases, heart failure by bradycardia in 18.1%, Sick-Sinus Syndrome in 5.5%, prophylactic in 1.2% and miscellaneous in only 2.2%. However, in contrast to the total variation among several years, indications for implantation were rather differently managed by the individual centres although recommendations were regularly provided at meetings of the experts organized every 18 months. Concerning the confounding ECG alterations the different centres indicate an even higher variability. There may be variable use of ECG criteria rather than differences in prevalence of certain arrhythmias in the different counties of the country. The average age of all patients treated was found between 1979 and 1986 to be around 70 years, gender relation being also quite similar, from year to year. The age group above 80 is steadily increasing. Regional differences and the influence of pacemaker therapy on life expectancy can now easily be calculated by using the database.

### **Secondary prevention after aorta-coronary surgery**

Coronary heart disease can be effectively controlled only by using preventive measures. Although primary prevention has been introduced in many ways, secondary prevention seems to be much less established and needs further activities and proof. There is, however, some experience available from patients after myocardial infarction that may be made usable for patients after aorta-coronary bypass surgery. Some activities can be foreseen therefore to follow up patients postoperatively, based on databases to be created. Till now, such studies have predominantly focused however on perioperative risk factors and complications after operation, thus indicating life expectancy gain. Studies assessing risk-factor development and drug intervention in correlation with the reoccurrence of chest pain are widely missing. Although surgery plays an important role in relief of symptoms, it must be emphasized that the basic disease cannot be altered surgically; it will continue to affect both those vessels that have been operated on and those that have not if other interventions are not performed systematically. This is another reason for establishing databases for

coronary patients. On the other hand, patients who have had surgical treatment of other cardiovascular diseases, especially congenital, have to be followed up carefully.

### **Methods and material**

At the Charité there is a centralized cardiological outpatient department called Herz-Kreislauf-Dispensaire closely cooperating with the cardiological wards and cardiac surgery. A database was begun for patients after coronary bypass operation in January 1987 by using a personal computer (Robotron A5120) and dBASE II. A special sheet was designed focused on symptoms (angina pectoris) and risk factors as well as on social aspects and drug treatment. Re-examinations are performed six months after the operation and every year afterwards, the patients being continuously cared for by their home doctors, who are provided with the data ascertained at the Charité. Each re-examination is mainly used for updating the information on risk-factor development individually performed by cardiologists.

In three years, 1987-1989, the data from 363 patients were stored in the computer. In principle, the database has proved to be valuable, though the PC used is slow if more complex analyses are to be done. The data input is time-consuming as is the data processing, which has to be performed by one of the cardiologists since paramedicals are not available for this work.

### **Experience and selected results**

The evaluation of surgical intervention seems necessary on a regular basis; this evaluation may be greatly supported by organizing a database. An increasing percentage of reoccurrence of angina is found (compare P. Sergeant and E. Lesaffre, K.U. Leuven Coronary Surgery Data Base) which may be due to a progression of the disease and/or alteration of the bypass grafts. Especially after a coronary bypass operation patients and doctors may think that this intervention was the most important step of treatment, and often patients anticipate that the disease will be cured. On the other hand, we could show that the comprehensive treatment of the basic disease is thus discontinued: necessary drugs are withdrawn and risk-factor management gets neglected. In particular, blood pressure rises again (in more than 50% of the cases), frequently in combination with body weight gain, favouring the reoccurrence of angina, and cholesterol remains widely uncontrolled. Data processing has shown that there is some correlation between cholesterol level and the incidence of postoperative chest pain.

## Conclusions

Databases can be used and even be necessary for cardiology in several areas of therapy control. The necessity of standardization of methods and criteria would already provide an advantage, even without computer application. In chronic diseases treatment has to be organized and checked during long-term follow-up for therapeutic as well as preventive reasons, presuming comparability of the repeatedly obtained criteria. Simultaneously, comparison with similar databases may become reliable, which otherwise would not generally be the case.

The flexibility and compatibility of databases are highly dependent on the hardware and software. After many years of experience and development, a maximum of flexibility and compatibility may already have been achieved with the Pacemaker Registry at the Charité. The minimum may be what we are using for the coronary bypass patients. An acceptable compromise is represented by the Charité's cooperation with the Department of Medical Informatics, which is caring for the database for preventive studies, making it therefore unlimited and suitable for fast processing.

The three databases used have applications in patient care and in research. They thus give medical and technical examples for other fields of application and other institutions, inside and outside the country. The higher transparency of the data and procedures processed seem to provide an improved basis for cost reduction in medicine. Unnecessary treatment methods can be discovered and disregarded as preventive and therapeutic means are indicated. Some competition could even be organized among the pacemaker clinics of the country or among doctors of the same region.

## Bibliography

- Arntzenius AC, Epstein FH, Günther KH, Kornitzer M, Ménard J, Strasser T (1978) Preventing Coronary Heart Disease— A Guide for the Practising Physicians, European Society of Cardiology, Van Gorcum Assen, The Netherlands
- Feruglio G, Witte J, Steinbach K, Edhag O, Perez-Gomez F, Dodinot B, Rickards A (1987) Cardiac Pacing. *Eur Heart J* 8: suppl F, 21-22
- Günther KH, Piorkowski P, Bohm R, Braun H, Handreg W (1980) Integrierte Prävention der Herzkreislaufkrankheiten— Ein Modell für die allgemeine medizinische Grundbetreuung. *Dt Gesundh Wesen* 35: 1893-1907
- Günther KH, Eckoldt P, Grüber H (1981) Computer in der Kardiologie (Review in German). *Dt Gesundh Wesen* 36: 1121 and 1201
- Günther KH, Laguardia AS, Dilba V, Piorkowski P, Bohm R (1988) Hypertension and social factors in a developing country. *J Hypertension* 6, suppl 4: 608-610
- Günther KH, Piorkowski P, Laguardia AS, Dilba V, Junes P, del Campo E, Handreg W, Bohm R, Braun H (1987) Comparative studies in hypertension under different continental conditions. *Nephron* S1, B: 125-127
- Günther KH, Piorkowski P, Dilba V, Bohm R (1989) Cottbus GPs Preventive Study: intervention results correlated with professional qualification partly. *CVD Epidemiology Newsletter* 45 100-101



- Ialongo D, Barbaro V, Bartolini P, Angeloni A (1990) A pacemaker oriented DBMS of the Italian market. In *Databases for Cardiology*, Pincioli F, Meester GE (eds). Kluwer Academic Publisher Group, Dordrecht/Boston/London
- Moebes RR, Witte J, Günther KH (1989) Comprehensive country-wide pacemaker registry of the German Democratic Republic- Computer-based. *CVD Epidemiology Newsletter* 45: 97-98
- Moebes R (1989) *Herzschrittmacher-Register der DDR von 1979 bis 1986*. Dissertation B (DMSc), Humboldt University Berlin
- Sergeant P, Lesaffre E (1990) The KU Leuven coronary surgery database A clinical research database. In *Databases for Cardiology*, Pincioli F, Meester GE (eds). Kluwer Academic Publisher Group, Dordrecht/Boston/London
- Witte J, Moebes R (1982) Zum aktuellen Stand der Herzschrittmacher-Therapie in der DDR. *Dt Gesundh Wesen* 37: 1874
- Witte J, Moebes R, Zobel W, Wohlgemuth P (1989) Results of pacemaker therapy in the German Democratic Republic. *Cor Vasa* 31: 98

# Development of and experience with the coronary angiography and PTCA information systems at Leuven University

JOS L. WILLEMS, STEF RENKENS, MARCEL DE WEL,  
JAN PIESSENS

*Div. of Medical Informatics and Cardiology, University of Leuven, Belgium*

## **Abstract**

The present paper describes the objectives, implementation and experiences with the computer-assisted coronary angiography and percutaneous transluminal coronary angioplasty (PTCA) reporting and information systems, which we developed at the University Hospitals of Leuven, Belgium.

The coronary angiography system is in daily routine use since April 1, 1975. A balance was obtained between practicality, feasibility and economical aspects on the one hand, and completeness and conciseness of clinically relevant information on the other hand. The system is based on a 2-page coding document, which is coded and filled in by the physician who performed the procedure. Special care was taken to handle anatomical variability and to produce a grammatically readable report in the Dutch language. The computer-generated reports, signed by the cardiologist, are sent directly to the referring physicians. Data on over 35000 patients have been stored to date. Both in 1989 and in 1988 more than 4000 cases were entered in the database.

The PTCA reporting and information system was developed and put in daily practice early in 1987, as a supplement to the coronary angiography system. The data stored, permit a detailed description of the procedure, its results and complications. Also follow-up data are collected in the database. The PTCA database increased in size with 1050 entries in 1989, versus 850 cases in 1988.

Both systems impose systematic and standardised reporting by different cardiologists. They alleviate the steadily increasing administrative workload through an increased efficiency in acquisition, storage and retrieval of information. In addition, the systems are very useful for quality assurance testing and clinical research studies.

## **1. Introduction**

It is widely known that cardiovascular diseases especially ischemic heart disease are the leading cause of mortality and morbidity in the Western World. In the

combat of this disease we have witnessed a steady rise in the number of coronary angiographies and percutaneous transluminal coronary angioplasty (PTCA) procedures in recent years in most Western Countries. This puts a heavy burden on most cardiac catheterization laboratories.

It was Sones [1] who in May 1959 demonstrated that the human coronary arteries can be individually and selectively catheterized safely, both in health and disease, if appropriate instrumentation and methods are used. In fact Sones himself designed the first special tip-tapered catheter for clinical coronary angiography. The Sones catheter is introduced via a ‘cut-down’ approach in the brachial artery. Judkins [2] invented some years later the femoral ‘percutaneous’ approach using pre-shaped catheters. Since then the number of coronary angiographies has risen steadily and must now have reached well over 400000 procedures per year in the United States. A similar rise has been observed in many other countries. Figure 1 illustrates the progressive rise of the workload in our own institution.

In the last decade we are similarly witnessing, an extraordinary expansion in the use of PTCA as an alternative for achieving myocardial revascularisation. Such growth can be attributed to the demonstrated clinical benefit of PTCA, and to recent technical advances in balloon catheter technology and higher dilatation success rates. Coronary angioplasty was first introduced by Andreas Grüntzig in September 1977 [3]. By 1980 he had performed the procedure on 169 patients. In the United States the number of PTCA procedures rose to 32000 in 1983, 135000 in 1986 and over 185000 in 1988 [4]. At Leuven University the mark of 1000 PTCA procedures was reached in 1989 (see Figure 1).

The technique consists of advancing a balloon-catheter across a stenotic

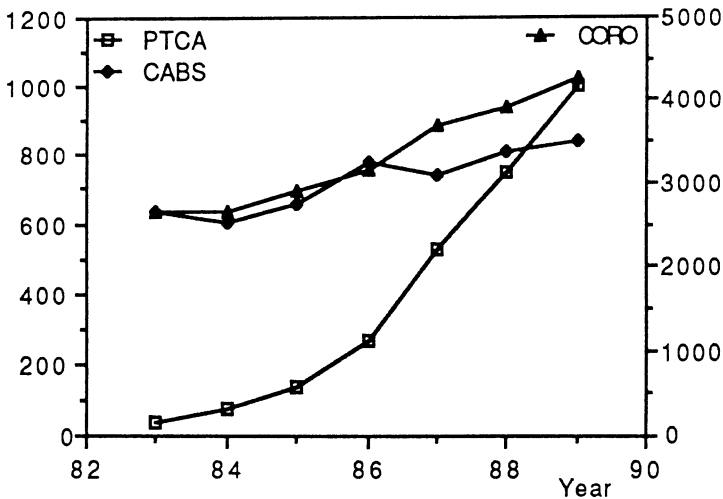


Figure 1.

Abbreviations: CABS : coronary artery bypass surgery  
CORO: coronary angiography

segment of a coronary artery and relieving the stenosis by inflating the balloon. Morphological studies have shown that this procedure, when successful, disrupts the intima and splits the atherosclerotic plaque. During the early years, PTCA was predominantly used to treat patients with discrete, proximal, non-calcified subtotal occlusive lesions in a single coronary artery. In subsequent years, thanks to advances in catheter technology and increased operator experience, the technique has been applied successfully to patients with multivessel disease, multiple stenoses in the same or another vessel, accessible complete occlusions of recent origin, as well as partial occlusions of saphenous vein or internal mammary artery grafts. Patients with total thrombotic occlusions in acute myocardial infarction, and other high risk patients with congestive heart failure and even cardiogenic shock have also been submitted to PTCA [5]. Reoccluded vessels are dilated routinely on 2, 3 or more occasions, and also distal coronary lesions are dilated more and more often.

The rate of inability to cross and dilate a lesion is well below 10 or even 5% in the best laboratories, as compared with over 20% in the early years. The mortality has dropped to a fraction of one percent and the number of complications leading to infarction or emergency coronary bypass surgery has decreased considerably to less than 3%, respectively. The major drawback of PTCA still remains a relatively high restenosis rate. Restenosis occurs within the first 6 months in approximately 30% of patients in who the procedure was initially successful. Such a recurrent stenosis is technically easy to redilate, but the restenosis rate following a second procedure is comparable to that of the first.

## **2. Objectives of the coronary angiography database and reporting system**

In the preceding paragraphs the scene of coronary angiography and PTCA was set. The large number of procedures puts a heavy burden on the cardiological, nursing and technical staff. This causes management problems and requires special attention to quality assurance testing and follow-up. In order to cope with these problems we have developed in our institution first, a coronary artery reporting system and database [6]. Subsequently, we developed the PTCA reporting and information system [7]. Our goal was to achieve the following specific objectives:

1. Achieve systematic and standardized reporting of coronary angiography and PTCA procedures by various cardiologists.
2. Alleviate the steadily increasing administrative workload.
3. Assess periodically results in a large patient population.
4. Assist future research studies by case retrieval and follow up.

The coronary angiography reporting system is in use since 1975 [6] and has been presented at the annual international conference on Computers in Cardiology in Rotterdam in 1977 [8]. This system has almost not been changed since its inception and has sofar been used to report and store data on over 35000 coronary arteriographies. In 1989 and in 1988 it was used to report 4311 and

3948 coronary angiographies, respectively, excluding PTCA procedures.

The system does not intend to give an accurate anatomical picture of all, even the most minor branches of the coronary circulation. Clinically, this would be irrelevant and practically difficult to achieve. Indeed, requirements were:

1. that the computerized system would pose no greater burden on the medical and secretarial staff than the old system with the dictaphone and typewriter, and
2. that it had to be applied in all patients, in daily routine practice. At the time of the introduction of the system the load was 800 patients per year. Since then the patient load has increased five fold to over 4300 cases in 1989.

In the following paragraphs we will first provide an overview of the coronary angiography and PTCA information systems and subsequently describe in detail various data items and their relevance. In the concluding sections comments will be given on the advantages and disadvantages of these systems and their corresponding databases. We consider it an achievement that after 15 years the same reporting system is still applied routinely in our institution, one of the busiest Cardiology Departments for adults in Western Europe.

### **3. Overview of the information systems**

#### *3.1. Overview of the coronary angiography system*

The original system was designed on a 2-page, punch card oriented document, which had to be coded and filled in by the physician who performed the coronary arteriography. The data were key-punched into 4 or, in some patients 5 punch cards by a medical secretary who had access to a multipurpose HP 2100 minicomputer system, which in 1975 was operating under RTE-II. The system had 32 kilobyte of core memory and two hard discs of 5 megabytes. The computer program was written in FORTRAN IV and consisted of 4 overlays running in the 11 K background partition.

The printouts were autographed by the cardiologist and sent directly to the referring physician together with the discharge letter. All data were stored in an IMAGE database. Reports were made in 6 copies. Eventual corrections, after review of the films or the report, could be made in an interactive way via a specific editing program. Analysis and programming required roughly 6 man/months in 1975.

The same program has in subsequent years been transferred to an HP 1000 minicomputer system, the successor of the HP 2100. The same two-page document is still being used by the cardiologists, with minor modifications, but the data are nowadays entered on a video-display terminal by the secretary. The program now runs under RTE-6, the latest operating system of the HP 1000 series, but the data are still stored in an IMAGE database, as 15 years ago. The current computer system has 1.5 megabyte of core memory and 3 hard discs of 120 megabyte each. Patient data and hard copies of the reports can be made on

an HP laserjet in the secretarial office of the Cardiology Department and the Cardiovascular Surgical Department as well.

### *3.2. Overview of the PTCA system*

The PTCA system is also based on a 2-page coding document, which consists of 3 main parts. In the first part are recorded demographic patient data, the approach (brachial or femoral) and identity of the cardiologist who performed the procedure. The second part is used to register information on the PTCA procedure itself, and the third for follow-up data. This goes one step further than the coronary angiography system, since follow-up data on patients submitted to coronary arteriography are not collected routinely but only for selected studies.

A reporting program, written in PASCAL, uses in total 87 items to print a full report on the PTCA procedure. Special care was taken to produce a grammatically readable output in our native Dutch language. The computer-generated reports are produced on the same HP laserjet printer. They are autographed by the cardiologist who performed the procedure and also sent directly to the referring physician.

Since the coronary artery reporting system was implemented on an HP 1000 system, using the IMAGE 1000 database and RTE-6 as operating system, the same basic hard- and software has been used for the PTCA reporting system. The demographic data (patient's name, address, tel., profession, referring cardiologist and general practitioner) are retrieved from the coronary artery database by linkage through the unique medical record number. This number consists of the inverted birth date (YYMMDD), sex and a 3-digit number. The first two of these digits are used to differentiate patients of the same sex, which are born on the same date. The last digit is a check digit, modulus 10. The unique patient number is attributed by the central hospital computer, to which the network of 18 closely coupled minicomputer systems of the Leuven University Hospitals is connected [9].

## **4. Details on the coronary angiography coding and reporting system**

The coding form is subdivided in 8 sections. The heading serves to register patient demographic data (patient number, ID number, angio number) and some general information, 70 bytes in total, with regard to the catheterization procedure. This includes coding of the technical quality of the films (good, moderate, bad), the referral indication, date of the procedure and whether it is a first or repeat angiography.

### *4.1. General description of the coronary anatomy (Section A)*

The reporting of the coronary arteriogram starts in the first section with a

general description of the coronary circulation (*see KUL Coronary Angiography Coding Form*). The dominance pattern is noted by indicating whether the posterior descending artery takes off from the right coronary artery and by counting the inferolateral branches coming from the right as well as from the circumflex artery. The number and relative thickness of the lateral branches of the circumflex with respect to the atrioventricular segment are noted and the length of the left anterior descending artery, ending beyond, on or before the left ventricular apex is coded. Some anatomical anomalies are recorded when present. All this information was originally punched on the first punch card, but is nowadays entered via one screen, with a lay-out which exactly follows the coding document.

#### 4.2. Coding of different coronary arterial segments (Sections B, C, D)

A detailed description on a total of 26 arterial segments can be given in the Sections 2, 3 and 4.

For coding the *right coronary artery* (RCA), the crux regio has been considered separately. Indeed, this segment is of utmost importance as site of the distal anastomosis in case of bypass surgery. The vessel proximal to the crux has been divided into 2 segments by the origin of the acute marginal branch. The vessels supplying the inferior wall of the left ventricle are coded in a separate section as *inferolateral region* (*see KUL Coronary Angiography Coding Form*). Indeed, the posterior descending artery as well as the inferolateral branches, may originate from the right as well as the circumflex artery, depending on the previously depicted dominance pattern.

No special fields have been reserved for small branches supplying, for example, the sinus node or the free wall of the right ventricle. From a surgical and functional standpoint these branches are of little significance, except when they become a source of distal filling for a stenosed artery. This is taken care of in our system under the description of the collateral circulation.

The fourth section describes the *left coronary artery* and the following segments are coded:

- first the main stem,
- then the left anterior descending artery (LAD), which is subdivided into a proximal, middle and distal third. The first major septal perforator is the delimiter between the proximal and middle third. In case of a stenosis in the proximal part of the LAD, its position with regard to the origin of the first major diagonal is noted, since this is functionally important and may have special surgical consequences.
- lesions of the circumflex artery (CX) are noted and located with respect to the origin of the different lateral branches. Up to four lateral branches can be coded and special fields are reserved for grading abnormalities of the intermediate artery when present.

The detailed description of the 4 major vascular territories (RCA, inferolateral region, LAD and CX) is preceded by a separate box allowing to indicate



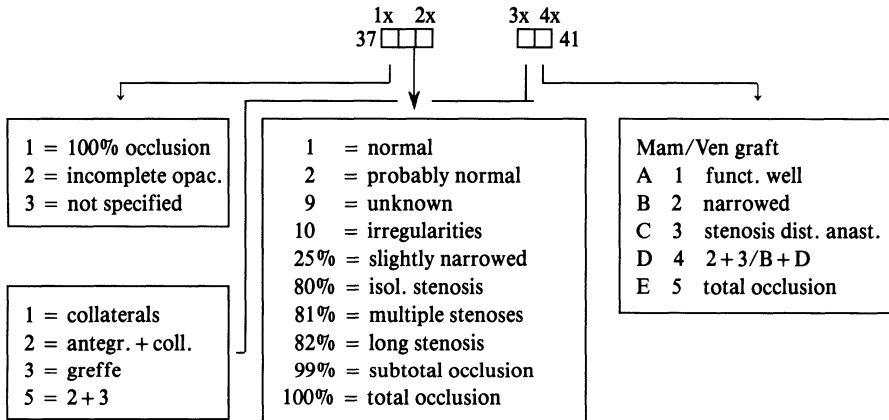


CODING DOCUMENT CORONARY ANGIOGRAPHY - KUL - PART B

<p><b>E. COLLATERAL CIRCULATION</b> <input type="checkbox"/> (1 absent; 2 present)</p> <hr/> <p>1. <u>RCA</u> → <u>LAD</u> <input type="checkbox"/>          via . margo acutus <input type="checkbox"/>              . conus branch <input type="checkbox"/>              . desc. post. apex <input type="checkbox"/>              . septal <input type="checkbox"/>              . other branches <input type="checkbox"/></p> <p>2. <u>RCA</u> → <u>Cx</u> <input type="checkbox"/>          via . atrial <input type="checkbox"/>              . terminal <input type="checkbox"/>              . other branches <input type="checkbox"/></p> <p>3. <u>LAD</u> → <u>Cx</u> <input type="checkbox"/>          via . septal <input type="checkbox"/>              . diag. and lat. <input type="checkbox"/>              . other branches <input type="checkbox"/></p> <p>4. <u>Prox. RCA</u> → <u>Dist. RCA</u> <input type="checkbox"/>          via . bridge <input type="checkbox"/>              . marginal <input type="checkbox"/>              . atrial <input type="checkbox"/>              . other branches <input type="checkbox"/></p> <p>5. <u>Prox. LAD</u> → <u>Dist. LAD</u> <input type="checkbox"/>          via . bridge <input type="checkbox"/>              . septal <input type="checkbox"/>              . diagonal <input type="checkbox"/>              . other branches <input type="checkbox"/></p> <p>6. <u>Prox. Cx</u> → <u>Dist. Cx</u> <input type="checkbox"/>          via . bridge <input type="checkbox"/>              . lateral <input type="checkbox"/>              . atrial <input type="checkbox"/>              . other branches <input type="checkbox"/></p> <hr/> <p><b>MINICODE :</b>          1 : absent          2 : →          3 : &lt;-          4 : → and &lt;-</p>	<p><b>F. VENTRICULOGRAM</b> <input type="checkbox"/> (1 : normal; 2 : abnormal)</p> <hr/> <p>Projections <input type="checkbox"/>          anterior <input type="checkbox"/>              - basal <input type="checkbox"/> <b>Minicode</b>              - lateral <input type="checkbox"/> 1 normal          inferior <input type="checkbox"/> 2 hypokinetic              - basal <input type="checkbox"/> 3 akinetic              - lateral <input type="checkbox"/> 4 paradoxical          posterior <input type="checkbox"/> pulsations          septum <input type="checkbox"/> 5 mid-systolic          apex <input type="checkbox"/> dyskinesia</p> <hr/> <p><b>G. OTHER ABNORMALITIES</b> <input type="checkbox"/> (1 : absent; 2 : present)</p> <table style="width: 100%; border: none;"> <tr> <td style="width: 15%;">Valve</td> <td style="width: 15%;">Sten (1-3)</td> <td style="width: 15%;">Insuf. (1-4)</td> <td style="width: 35%;">Hypertr. non obstructive</td> <td style="width: 20%; text-align: center;"><input type="checkbox"/></td> </tr> <tr> <td>Ao.</td> <td><input type="checkbox"/></td> <td><input type="checkbox"/></td> <td>Hypertr.obstr.</td> <td style="text-align: center;"><input type="checkbox"/></td> </tr> <tr> <td>Mitr.</td> <td><input type="checkbox"/></td> <td><input type="checkbox"/></td> <td>Congest.</td> <td style="text-align: center;"><input type="checkbox"/></td> </tr> <tr> <td>Tric.</td> <td><input type="checkbox"/></td> <td><input type="checkbox"/></td> <td>Other</td> <td style="text-align: center;"><input type="checkbox"/></td> </tr> </table> <p>(calcified +3 +4)</p> <hr/> <p><b>H. FREE TEXT</b> <input type="checkbox"/> (1 : absent; 2 : present)</p> <hr/> <p><b>I. CONCLUSION</b> <input type="checkbox"/> (1 : absent; 2 : present)</p> <table style="width: 100%; border: none;"> <tr> <td style="width: 40%;">RCA</td> <td style="width: 10%; text-align: center;"><input type="checkbox"/></td> <td style="width: 50%;"></td> </tr> <tr> <td>MAIN STEM</td> <td style="text-align: center;"><input type="checkbox"/></td> <td></td> </tr> <tr> <td>LAD</td> <td style="text-align: center;"><input type="checkbox"/></td> <td>1 normal</td> </tr> <tr> <td>CIRCUMFLEX</td> <td style="text-align: center;"><input type="checkbox"/></td> <td>2 slight</td> </tr> <tr> <td>LV FUNCTION</td> <td style="text-align: center;"><input type="checkbox"/></td> <td>3 moderate</td> </tr> <tr> <td>INTERM. ART.</td> <td style="text-align: center;"><input type="checkbox"/></td> <td>4 severe</td> </tr> <tr> <td>COR. SURGERY</td> <td style="text-align: center;"><input type="checkbox"/></td> <td></td> </tr> </table>	Valve	Sten (1-3)	Insuf. (1-4)	Hypertr. non obstructive	<input type="checkbox"/>	Ao.	<input type="checkbox"/>	<input type="checkbox"/>	Hypertr.obstr.	<input type="checkbox"/>	Mitr.	<input type="checkbox"/>	<input type="checkbox"/>	Congest.	<input type="checkbox"/>	Tric.	<input type="checkbox"/>	<input type="checkbox"/>	Other	<input type="checkbox"/>	RCA	<input type="checkbox"/>		MAIN STEM	<input type="checkbox"/>		LAD	<input type="checkbox"/>	1 normal	CIRCUMFLEX	<input type="checkbox"/>	2 slight	LV FUNCTION	<input type="checkbox"/>	3 moderate	INTERM. ART.	<input type="checkbox"/>	4 severe	COR. SURGERY	<input type="checkbox"/>	
Valve	Sten (1-3)	Insuf. (1-4)	Hypertr. non obstructive	<input type="checkbox"/>																																						
Ao.	<input type="checkbox"/>	<input type="checkbox"/>	Hypertr.obstr.	<input type="checkbox"/>																																						
Mitr.	<input type="checkbox"/>	<input type="checkbox"/>	Congest.	<input type="checkbox"/>																																						
Tric.	<input type="checkbox"/>	<input type="checkbox"/>	Other	<input type="checkbox"/>																																						
RCA	<input type="checkbox"/>																																									
MAIN STEM	<input type="checkbox"/>																																									
LAD	<input type="checkbox"/>	1 normal																																								
CIRCUMFLEX	<input type="checkbox"/>	2 slight																																								
LV FUNCTION	<input type="checkbox"/>	3 moderate																																								
INTERM. ART.	<input type="checkbox"/>	4 severe																																								
COR. SURGERY	<input type="checkbox"/>																																									

normality versus abnormality. When the LAD for example is found to be entirely normal, a *one* is noted in the corresponding box on the data sheet, and the interpreter can skip to the next major artery. A two is entered when the segment presents abnormalities. This option greatly simplifies and speeds up the coding procedure, especially for normal angiograms.

The 5 boxes for coding each of the vessel segments constitute 4 different scoring fields, as specified in *Figure 2*. The second field is always filled in for the main description of the coronary artery segment, whereas zones 1, 3 and 4 are used as suffix descriptors. Code 1 is filled in the second field when the segment appears completely normal. When only slight wall abnormalities are present we use code 10. Stenoses are graded and can be filled in with a maximal resolution of 5%, by using figures from 30 to 95. Through the use of the end digits 0, 1 and 2 or respectively 5, 6 or 7 we differentiate isolated from multiple stenoses, and narrowing over a short from narrowing over a long segment, respectively. Finally, 99% stands for subtotal and 100% for a total vessel occlusion. The most severe lesion is coded in case multiple stenoses are present in a single arterial segment.



*Figure 2.*

The third field is used to indicate if the vessel is being opacified by collateral channels, by both collateral and antegrade filling, or by a bypass graft.

The fourth field serves as a descriptor zone for the graftable status of the coronary arterial segment or, in case of a postoperative study, as a description of the bypass patency. This zone is only filled in for the segment where the graft has to be or has been inserted. Numeric codes are used for venous bypass grafts and alphabetic codes for internal mammary artery implants.

#### 4.3. Description of collateral coronary circulation (Section E)

If no collateral filling exists, again a 1 is written in the separate box heading this section and the interpreter can skip to the next section of the form for coding the

ventriculogram (see KUL Coronary Angiography Coding Form).

As specified in the Coding Form collateral circulation can be described between 6 different arterial segments. The direction of the blood flow is indicated by using the codes 2, 3 and 4, and the branches through which the retrograde filling occurs are checked. Code 2 is used if the flow runs in the direction as indicated on the document; 3 is used for the opposite way; and 4 is filled in when collateral circulation is present in both directions.

#### *4.4. Coding of the left ventriculogram, valvular and other abnormalities (Sections F, G, H)*

Wall motion of 7 segments of the left ventricle are described similarly as in the AHA protocol [10] using minicodes for hypokinesia, akinesia, paradoxical motion and midsystolic dyskinesia. The left ventricular wall has been divided into an anterobasal, anterolateral, apical, inferolateral, inferobasal, septal and posterolateral segment. Left ventricular and aortic pressures are noted as well.

Abnormalities other than coronary lesions, especially valvular, congenital and primary disease are coded when present (see Coding Form).

In case the interpreter wants free formatted text to be printed on the report, he fills in a 2 in the appropriate box. The text, up to 80 characters long is then transferred and stored in the computer files. This route of entering specific, non-codable information, was used in 10% of the first 2000 cases and has in recent years increased to about 15%.

#### *4.5. Conclusions and final judgment of the reader (Section I)*

A summary of the coronary arteriography interpretation is made at the end. Each main coronary artery is coded as normal, or as having slight, moderate or severe coronary lesions. A 50% or greater narrowing of the vessel diameter has been defined as a severe lesion. A subjective angiographic evaluation of global left ventricular contractility is noted as well, i.e., normal, slightly, moderate or severely depressed (see Coding Form).

The cardiologist finally makes a judgment on the operable status of the patient strictly based on angiographic findings. For example, the following judgments can be made: the patient is a candidate for coronary artery surgery; only partial revascularization is possible; there is an increased risk due to impaired LV function; coronary artery surgery is not indicated in view of the diffuse disease or in view of severely depressed left ventricular (LV) function. One of 15 such comments can be made. The judgment on the operable status of the coronary artery tree is summarized in the last line of the report.

#### *4.6. Overall functional coronary severity score*

An overall score of severity of the coronary lesions is calculated, using the graded narrowing and a weighting of the relative importance of the different arterial segments.

Based on mortality figures published by Brusckhe *et al.* [11] and Reeves *et al.* [12], the main stem was given a basis weighting factor of 10, the right coronary artery 2, the LAD 4 and the circumflex 3. The weighting factors of the RAC and CX are adjusted in accordance with the dominance pattern of the coronary circulation. Distal arterial segments receive a fraction of the weight of the main coronary arteries. The final score is calculated by summing up the products of the squared graded stenoses times the basic weighting factor multiplied by a modifying factor for each of the respective arterial segments. The modifying factor takes into account the length and number of the different stenoses, and the presence of collateral circulation.

The severity score which we developed at our institution is similar to the one developed by Gensini and coworkers [14]. Although not suitable for calculating by hand, the severity score provides more useful information than the simple division of patients into single-, double- and triple- vessel disease. It allows a detailed stratification of patients according to the functional significance of their disease. In addition to this index, which grades the overall hemodynamic and prognostic severity of the coronary lesions in a somewhat complex manner, the computer is also programmed to derive the Friesinger index [13] which is a more simple angiographic score. The Friesinger score varies between 0 and 15.

The severity score devised by us varied between 0 (normal) and 224 in the first 2000 cases, with an average score of 46 (SD 31) excluding 561 patients with normal coronary arteries. The corresponding mean Friesinger index was 7.7 (SD 3.6). A Pearson correlation coefficient of 0.67 was found between both indices in this series presented at *Computers in Cardiology* in 1977 [8]. For the 3091 most recent cases with a score  $> 0$  a similar correlation was obtained (Pearson  $r = 0.73$ ; Spearman  $r = 0.79$ ). In this series, the average severity score was 42 (SD = 33; range 1 to 220) after excluding patients with a score of 0. The corresponding Friesinger score was 6.74 (SD 3.7).

## 5. Detailed data on the PTCA information system

### 5.1. PTCA coded data items

Per dilated stenosis are coded:

- The affected arterial segment, i.e. the left anterior descending (LAD), proximal, middle or distal third; the 2 LAD diagonals; the circumflex (CX) proximal or distal of the margo obtusus; the 1st and 2nd lateral and inferolateral branches of the CX; the intermediate branch; the right coronary artery (RCA), proximal or distal half; the posterior descendens or inferolateral branch of the RCA; one or more venous or mammary artery bypass grafts and their location, and finally the main stem.
- The measured or estimated diameter stenosis, pre- and post-dilatation, in percent.
- When measured, the densitometric and geometric area stenosis in percent.

- Qualitative aspects of the stenosis: its length, single versus multiple, type (eccentric or concentric) and calcifications.
- The number, duration, and applied pressure of the balloon inflations.
- Type and size of the balloon catheter used.
- Appreciation of the final result of the procedure per dilated vessel, including the possibility to indicate the cause of an eventual technical failure (inability to reach or to cross the lesions with either the guidewire or the balloon).

Are also coded:

- The presence or absence of ischemia (pain and/or ECG changes).
- Eventual complications i.e. thrombosis of a main vessel or occlusion of a side branch, its timing with respect to the procedure as well as its relation to heparin infusion; distal embolisation; occurrence of arrhythmias, conduction disturbances, a cerebrovascular accident, major dissection of a vessel or myocardial infarction.
- Mortality and/or eventual surgery (elective or urgent).
- The possibility exists to register the global ejection fraction after each vessel dilatation and whether special techniques were applied, such as kissing-wire technique. Also some anatomical particularities may be noted such as whether the stenosis was located proximal or distal to a bypass graft.

There is no limit in reporting the number of vessel dilatations. In addition, significant stenoses for which no dilatation attempt was made can be coded.

### 5.2. Coded information of follow-up.

At each follow-up visit or written interview, the following data can be recorded:

- Date of follow-up.
- Activity status: back at work, sick leave, retired,...
- Coronary risk factors: serum cholesterol, smoking habits, arterial hypertension, diabetes; also weight and height are stored.
- Angina pectoris, classified according to the Canadian Heart Association.
- Myocardial infarction, when present: timing, location and extent.
- Quantitative exercise ECG and scintigraphic thallium test results.
- Therapy: medical (type of medication), coronary surgery or redilatation.

### 5.3. Example of a PTCA report for a single lesion dilatation

A copy of a translated report is presented below for a single vessel dilatation, which in 1988 accounted for 69.6% of all cases in our institution.

NAME OF PATIENT

DATE 11/05/88

PTCA NR: 1286      FILM NR: 34218

PATIENT NR: 410630M122

*Description of lesion before PTCA.* A stenosis over a long segment, estimated as 80%, was present in the middle third of the left anterior descending artery. The lesion was eccentrically located and not calcified.

*Procedure and results.* Four balloon inflations were performed over a total of 395 seconds. The maximally applied pressure was 5 atmosphere. A Schneider-Monorail catheter was used with a balloon size of 2.5 mm. The angiographic result was satisfactory in so far that only a 30% stenosis remained after dilatation. During inflation the patient experienced chest pain and showed ischemic ECG changes.

*Complications.* An acute rethrombosis occurred during the procedure, which was successfully treated by redilatation.

Procedure and results of this redilatation:

Two inflations were given over a total time span of 385 seconds, with a maximally inflation pressure of 10 atmosphere. A Schneider-Monorail catheter was used with a balloon of 2.5 mm. The angiographic result was excellent since there was only a mild residual stenosis. Routine medical treatment was continued.

Name and Signature  
of the Cardiologist

## **6. Physicians' acceptance of the coronary angiography and PTCA information systems**

When we started the implementation of our coronary artery reporting and information system, stimulated by the work of Gensini and coworkers [14], we felt that the system had to be used in daily routine practice for all patients submitted to selective coronary arteriography and that it had to replace entirely the conventional dictated reports. Our goals were aimed at an increased efficiency in the acquisition, storage and retrieval of information for routine patient care management as well as for research purposes. Fifteen years of experience have demonstrated that these goals can be reached in a cost-effective manner.

Special care was taken to handle anatomical and pathological variability and to produce a grammatically acceptable report. When consecutive artery segments, for example, are normal or present identical abnormalities, the description is combined into one syntactical sentence. Redundant stereotyped expressions are avoided as much as possible, which in our opinion is an asset for physicians' acceptance of computer-generated reports.

The results of these efforts have been very rewarding. All angiograms are coded. Data on over 35000 patients have been handled to date. The computer-generated reports signed by the cardiologist are sent directly to the referring physician and the data are stored in a database. Eventual corrections, to be made after reviewing the films, can be made in an interactive manner via an editing program. In this way accuracy and completeness of stored information is taken

care of as far as possible, taking into account the known inter- and intra-observer variability of visual interpretation of coronary arteriograms [15]. Nowadays, digital quantitative coronary angiography [16-18] is being applied more and more, but it is still difficult to forecast when this new technology will replace classic analog cine-angiography completely in routine practice.

The length of an average computer report is one page (50 lines). The computer reports are more complete and systematic than the previously dictated reports. This has been achieved without expansion in personnel or undue strain on the medical and secretarial staff. After a couple of weeks of training, an arteriogram can be coded by the cardiologist in less time than dictating a report. Data entry of the forms requires less secretarial time than typing a dictated report. Error checking is done for each item at data entry.

Alternative ways of entering the data have been considered, but have as yet not been implemented. The possibility of entering specific non-codable information will always be desired by physicians even with the most complete and complex coding system. Direct entry of the information on a video-display terminal by the physician, as done in some radiology reporting systems, bypasses the secretary. However, this is not always well-accepted by clinicians, especially not by those who do not like to use a key-board in a darkened viewing room.

Also the PTCA information system is well accepted by the cardiologists and referring physicians. This system is in daily use since May 1987 for the reporting of all PTCA procedures in our institution. In 1989, the number of procedures totaled 1051 versus 749 in 1988. Last year the number of PTCA procedures surpassed the number of coronary bypass operations in our hospital (see Figure 1). About 49% of the PTCA procedures were performed via the transfemoral approach. This approach was only used in 12.4% of the coronary diagnostic angiographies.

## **7. Main advantages of the systems**

The main advantages of the computer assisted coronary artery and PTCA reporting systems, which we developed in Leuven can be summarized as follows

1. The system forces upon the cardiologist a more systematic and standardized reading of the coronary arteriograms, and this leads to a more complete and better report for the referring physician.
2. Overall administrative work, especially typing work, has decreased. Computerized reporting of all coronary angiography and PTCA procedures at Leuven University is estimated to save yearly more than 1 man/year of typing work. Also, the ease of report retrieval and reproduction starting from the database is a net advantage. The database is also almost daily used by the secretarial staff of the Department of Cardiovascular Surgery.
3. The information systems have also been very useful from a managerial and clinical point of view. The databases proved to be very valuable to get quick insight in the rate of success and various complications. Dependent on the

selection criteria various statistics can be obtained on the results obtained with various sizes and types of balloons, different duration and pressure of inflation, by a particular operator, the brachial or femoral approach, etc. The coronary angiography system allows easy derivation of an overall severity score of coronary lesions, which might be of prognostic significance in the stratification of patients for follow-up.

4. The systems have added enormous memory extension to physicians trying to recall discrete patient data. Problems of case retrieval and summary reviews have been eliminated with the computerized database. The computer-based system of angiographic data, linked with other clinical laboratory, ECG and surgical information, is a basic tool for certain research projects. The system has been used for case retrieval in several studies in our institution on clinical, electrocardiographic and angiographic topics, as early as 1976 [19-22]. Statistical analysis is performed by the SAS package, after transfer of the data from the IMAGE database to a flat file.

## **8. Summary data derived from the coronary angiography and PTCA information systems in our institution**

### *8.1. Results derived from the coronary angiography database*

Some summary data on the first series of cases handled by the computerized reporting system were presented at Computers in Cardiology in September 1977 [8]. General statistics on patients examined since then, and available in the database (N = 32050) are presented in Tables I, II, III. Of the total population 73% were male and 27% female. Severe ( $\geq 50\%$ ) coronary stenosis was present in the RCA in 43.4%, in the LAD in 49.3% and in the CX in 37.2%. Severe main stem disease was found in 3.8%. This is very similar to figures reported previously [8]. Indeed, during the period before September 1977, representing 2006 procedures, severe narrowing of the RAC, LAD, CX and left main stem was found in 39.7, 51.3, 38.1 and 3.9% of the patients, respectively [8].

Severe single-, double- and triple-vessel disease was present in respectively 25.7, 21.3 and 21.7% of the total population, whereas 31.3 had no major narrowing in any of the coronary vessels, including patients studied for valvular, congenital or primary heart disease. Changes over time in the distribution of severe vessel disease was highly significant (see Tables 1 and 2) but it is beyond the scope of this paper to comment on these observations. Changes can be due to differences in the technical quality of the ciné-films, in reporting, patient selection, sex distribution, and other factors. However, it may be noted that differences were mainly found for cases catheterized before 1980. In the period before September 1977 the total percentage with normal vessels was 28.2% and with major narrowing (i.e. any stenosis  $\geq 50\%$ ) 48.9% (see [8]).

The distribution of left ventricular wall abnormalities as defined by AHA recommendations [10] is given in Table 3. Of the total (N = 31477) 45.9% were



*Table 1.* Coronary angiographic findings over time in patients studied at the Leuven University Hospitals\*

Period		No severe narrowing	Single vessel	Double vessel	Triple vessel	Total group
1978-1979**	N	1290	570	636	878	3374
	%	38.2	16.9	18.9	26.0	
1980-1984	N	3149	2478	2217	2812	10656
	%	29.6	23.3	20.8	26.4	
1985-1989	N	5589	5193	3970	3268	18020
	%	31.0	28.8	22.0	18.1	
Total	N	10028	8241	6823	6958	32050
	%	31.3	25.7	21.3	21.7	

Significance of difference in distribution: Chi-square = 515 ( $P < 0.0001$ ).

\* Numbers include patients studied for valvular, congenital and primary heart disease.

\*\* Including the last 4 months of 1977.

*Table 2.* Distribution of abnormal and normal coronary arteriography findings in patients examined at Leuven University\*

Time Period		Degree of vessel disease			
		None	Slight**	Moderate 25% - <50%	Severe ≥ 50%
1978-1979**	N	673	287	276	2138
	%	20.0	8.5	8.2	63.4
1980-1984	N	1231	933	837	7655
	%	11.6	8.8	7.9	71.8
1985-1989	N	1982	1894	1469	12675
	%	11.0	10.5	8.2	70.3
Total	N	3886	3114	2582	22468
	%	12.1	9.7	8.1	70.1

Significance of difference in distribution: Chi-square = 246 ( $P < 0.0001$ ).

\* Including the last 4 months of 1977.

\*\* Only vessel wall irregularities.

Table 3. Percentage of patients with left ventricular akinesia or dyskinesia

Wall Segment	1978-1979*	1980-1984	1985-1989	Total
Antero-basal	0.71	0.77	1.15	0.98
Antero-lateral	10.1	11.1	10.8	10.8
Infero-basal	10.1	13.4	14.5	13.7
Infero-lateral	11.9	12.9	12.3	12.6
Posterior	1.9	2.2	3.0	2.6
Septum	4.7	4.8	3.8	4.2
Apex	12.0	14.0	14.1	13.9
Any wall	23.3	27.4	28.7	27.7
Total number of cases	3374	10656	18020	32050

\* Including the last 4 months of 1977.

judged to have normal, 29.1% slight, 16.6% moderate and 8.4% severe depressed LV function. Of the total group of 32050 cases 15.3% presented valvular disease. Aortic stenosis was reported in 5.4%, aortic insufficiency in 7.6%, mitral stenosis in 3.3% and mitral insufficiency in 7.3%.

In our institution the mortality rate for the procedures performed in 1988 and 1989 was 0.04% (3 cases out of 8300). A cerebrovascular accident occurred in 6 cases, 31 received electrical defibrillation or a temporary pacemaker for arrhythmia's and 29 of these 8300 cases developed a myocardial infarction or had to undergo urgent coronary surgery. Anaphylatic shock as a result of the procedure was reported in 4 cases and local complications requiring treatment with a Fogarty catheter or other intervention in 78 cases.

## 8.2. Results derived from the PTCA database

Figures 3, as well as Tables IV and V, illustrate the experience with PTCA in the University Hospitals of Leuven. In the first 6 months of 1989 a significant decrease in thrombotic complications was observed probably because of the use of non-ionic instead of ionic contrast material. The incidence of immediate rethrombosis fell from a total of 11.5% in 1988 to 4% in 1989, and the need for immediate surgery from 2.2% to 1.9%. The ease with which such figures can be obtained is an important asset of the system. It may be important to note that most cases of the acute rethromboses in 1988 (11.5%) could be solved by redilatation and only 2.2% needed surgery. In 1988 the in-hospital mortality of all PTCA procedures was zero. These figures compare favorably with those reported in the literature.

PTCA EXPERIENCE 1988  
ANATOMICAL INDICATIONS (N = 716)

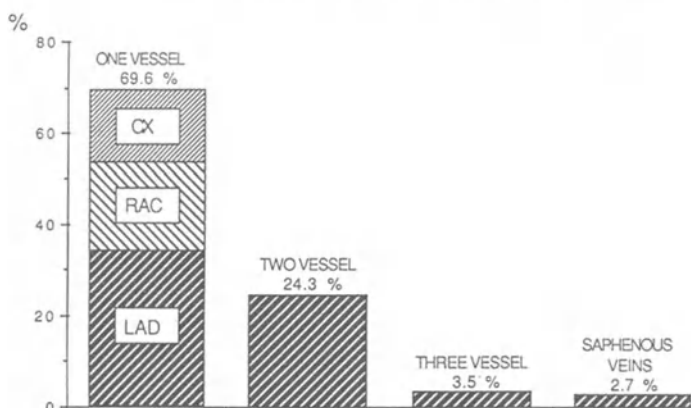


Figure 3.

Table 4. PTCA experience in 1988. Immediate results (900 lesions)

	Stenosis		Chronic occlusion		All lesions	
	N	(%)	N	(%)	N	(%)
1. <50% residual stenosis	796	(96)	41	(57.7)	837	(93)
2. Lesion not reached	1	(0.1)	0		1	(0.1)
3. Lesion not crossed	15	(1.8)	29	(40.8)	44	(4.9)
4. >50% residual stenosis or dissection	17	(2.1)	1	(1.4)	18	(2.0)
<b>Total</b>	<b>829</b>		<b>71</b>		<b>900</b>	

Table 5. PTCA experience in 1988 at Univ. Leuven: Complications of PTCA (N = 716)

	N	%
1. Acute rethrombosis	82	11.5
2. Occlusion side branch	10	1.4
3. Distal embolisation	2	0.3
4. Ventricular arrhythmias	20	2.8
5. Local complications	11	1.5
6. Death, MI, CVA	0	
<b>Total</b>	<b>125</b>	<b>17.5</b>

Abbreviations: MI : myocardial infarction  
CVA : cerebrovascular accident

## 9. Comparison with data reported in the literature

### 9.1. Coronary angiography information systems

Gensini *et al.* [14] probably developed the first coronary angiography information system which was later marketed by the Mennen and Greatbach Company. Since then, many have followed his example, including us at the Univ. of Leuven in Belgium [6, 8, 23-28].

However, overall summary data such as described in the present paper have to our knowledge not been reported in recent years, except for data on the Registry of the Society for Cardiac Angiography and Interventions [29]. This registry focuses on the complications of the procedure, but includes data on the extent of coronary artery disease. A vessel with more than 50% diameter reduction is considered significantly narrowed, similarly, as in the present study. Results on 222 553 patients were recently reported by Johnson *et al.* [29]. Of this population 66% were male and 34% female. Respectively 21, 20 and 26% had single-, double- and triple-vessel disease. Twenty percent were normal and 7% had less than 50% narrowing, which is very similar to our findings since 1980 (see Tables 1 and 2). Death was reported in 0.1% (218 cases), myocardial infarction in 0.06% (N = 144) and a cerebrovascular accident in 0.07% (N = 156) in this large series of over 220000 cases. The incidence of these complications rises significantly in cases with left main disease as well as in the older and sicker patients i.e. those with lower ejection fractions [29]. The laboratories contributing to this series represent both university and community facilities. However, they probably are not completely representative of the USA as a whole since only experienced centers with a considerable case load are participating. Neither is the present series representative for the whole of Belgium, although it may be noted that approximately one third of the total number of approximately 13000 coronary angiographies in Flanders i.e. the Flemish speaking, Northern part of Belgium (population  $\pm$  6000000) were performed in our institution in 1989.

### 9.2. Comparison with PTCA results obtained elsewhere

An initial angiographic success rate of 96% (796 of 829) was achieved for the dilated lesions in our institution in 1988. This compares favorably with the 90% (2342 of 2596) success rate in 1985-86 obtained in the NIH registry of selected hospitals with long-standing PCTA-experience in the States. Of the totally occluded lesions, in which angioplasty was attempted, 67% (197 of 296) were dilated in the NIH series, against 58% (41 of 71) in the Leuven data. Safian *et al.* [30] reported in 1988 a success rate of 90% in 711 patients with conventional (70 to 95%) stenoses, and of 63% in 169 pts with total occlusions and 78% in 102 cases with subtotal stenoses, respectively. Baim *et al.* [31] reported at the same symposium an overall success rate of 91%. In 3% emergency bypass surgery was required, as compared to 2.2% in the cases treated in Leuven in 1988, and 1.9% in the first half of 1989.

Nowadays, more PTCA procedures are performed than coronary bypass surgery. Baim [31] estimated that PTCA accounts for 58% of revascularisation in the United States. In 1989, this ratio was very similar in our institution. PTCA has proved its value as an alternative therapeutic modality for coronary disease, a modality that offers a lower initial cost, a shorter hospital stay, and even reduced morbidity contingent upon a successful procedure. However, the major disadvantage of the method remains a restenosis rate of 25 to 35% in the first 3 to 6 months after dilatation in all centers across the world.

## 10. Conclusion

The coronary angiography and PTCA reporting information systems which we developed in our institution at the University of Leuven meet the objectives specified initially. Through an increased efficiency in acquisition, storage and retrieval of information, the systems offer clear advantages both for routine patient care and research purposes.

Although we have attained our original goal, several possibilities for improvement remain. The network database structure of IMAGE imposes several limitations, which could be solved by conversion to a relational database such as SYBASE, INGRES or ORACLE. Except for routine retrieval jobs, all queries for more complicated case retrieval have to be made through the medical informatics department, nowadays. SQL facilities or a Tree-base structure such as described by Jung *et al.* [32] could possibly solve this problem by making the interactions more simple. However, cardiology fellows and full time staff have limited time in a very busy clinical department. They are still hesitant towards computer technology. It remains to be seen whether SQL will help to overcome this reluctance. The data contents of the coronary angiography and PTCA information system has been very stable over years. However, temporal linkage and serial comparison of sequential procedures in the same patient have not yet been implemented. Since 25 to 35% of PTCA procedures result in a restenosis, and a high percentage of these patients are undergoing a second procedure, serial comparison with previous findings is definitely desirable. Automatic linkage is possible in our system by means of the unique patient identification number (cfr. supra) but comparative evaluation has not yet received enough priority in the programming back-log.

Although over the past 15 years there has been no great demand for graphics from our cardiologists, we feel that such output would be a useful adjunct. With current laser printer technology, integration of text and graphic representation of the coronary artery tree can be solved quite easily. As a result, an extension in this direction is at present investigated.

No pictures are stored as yet, but storage of selected frames e.g. in a format of 256/256 or 512/512 pixels is in our opinion technically feasible for digitally recorded coronary angiographies.

Finally, further integration with other databases of the Department of

Cardiology and the Hospital Information System e.g. the Medical Discharge Summary Subsystem would be highly desirable. Such linkages have been made in the past for research studies. However, they have insufficiently been implemented in routine practice. Nevertheless, our experience of over 15 years with the coronary angiography information system, and more recently with the PTCA database, have convinced us that Wallace and Rosati [33] were indeed right when they forecasted in 1973 that: 'Computers can change cardiology'. Computer technology necessary to develop and run information systems, such as described in this paper, is readily available and has become inexpensive for widespread use.

### Acknowledgments

We are grateful to all cardiologists, nurses and technicians of the Catheterization Laboratory for their support and collaboration. We also want to thank Ms. I. Tassens and Mrs. V. Dillemans for typing the manuscript.

### References

1. Sones FM Jr (1959) Acquired heart disease: Symposium on present and future of cine-angiography. *Am J Cardiol* 3: 710
2. Judkins MP (1967) Selective coronary arteriography. A percutaneous transfemoral technic. *Radiology* 89: 815
3. Grüntzig AR, Senning A, Seigenthaler WE (1979) Non-operative dilatation of coronary-artery stenosis: percutaneous transluminal coronary angioplasty. *N Engl J Med* 301: 61-68
4. Detre K, Holubkov R, Kelsey S *et al.* (1988) Percutaneous transluminal coronary angioplasty in 1985-1986 and 1977-1981. The National Heart, Lung, and Blood Institute Registry. *N Engl J Med* 318: 265-270
5. Ryan TJ *et al.* (1988) Subcommittee on PTCA. Guidelines for percutaneous transluminal angioplasty. *JACC* 12: 529-545
6. Willems JL, Piessens J (1976) Codering en rapportering van het coronarogram met behulp van de computer. *Hart Bulletin* 7: 173-178
7. Willems JL, Piessens J, De Wel M, Renkens S (1990) Computerized reporting of percutaneous transluminal coronary angioplasty (PTCA) procedures. In *Proceedings Computers in Cardiology 1989*. Ripley KL (ed). IEEE Computer Society, Long Beach (in press)
8. Willems JL, Piessens J (1977) Implementation and experience with a computer based coronary angiography reporting and information system. In *Computers in Cardiology 1977* Ripley KL *et al.* (eds) IEEE Computer Society, Long Beach, 465-470
9. Willems JL (1988) Hospital information systems. The case of Leuven. In *Towards New Hospital Information Systems*. Bakker AR, Ball M, Scherrer JR, Willems JL (eds) North Holland Publ., Amsterdam, 77-82
10. AHA Committee Report (1975) A reporting system on patients evaluated for coronary artery disease. *Circulation* 51: Suppl 1-40
11. Brusckhe AVG, Proudfit WL, Sones FM (1973) Progress study of 590 consecutive nonsurgical cases of coronary disease followed 5-9 years. Part I and II. *Circulation* 47: 1147-1163
12. Reeves TJ, Oberman A, Jones WB, Sheffield LT (1974) Natural history of angina pectoris. *Am J Cardiol* 33: 423-430

13. Friesinger GC, Page EE, Ross RS (1975) Prognostic significance of coronary arteriograms. *Circulation* 52: 979-986
14. Gensini GG, Kelly AE, Esente P, Szency C (1975) Acquisition and management of a clinical and angiographic database in patients with ischemic heart disease. In *Proceedings Computers in Cardiology*. Ripley KL (ed). IEEE Computer Society, Long Beach, 65-68
15. Derouen TA, Murray JA, Owen W (1977) Variability in the analysis of coronary arteriograms. *Circulation* 55: 324-328
16. Reiber JHC, Serruys PW, Slager CJ (1986) *Quantitative Coronary and Left Ventricular Angiography*. Martinus Nijhof, Kluwer Acad Publ., Dordrecht
17. Sanders WJ, Alderman EL, Harrison DC (1975) Coronary artery quantitation using digital image processing techniques. In *Proceedings Computers in Cardiology*. Kempner KM (ed). IEEE Computer Society, Long Beach, 15-20
18. Bolson EL, Brown BG, Dodge HT (1975) Quantitative coronary arteriography. Use of a general DBMS for reporting results of a lesion progression study. In *Proceedings Computers in Cardiology*. Kempner KM (ed). IEEE Computer Society, Long Beach, 141-144
19. Willems JL, Ector H, Piessens J, Kesteloot H, De Geest H (1976) Comparison of computer and visual ECG analysis results: correlation with independent cine-angiographic data. *Proceedings Computers in Cardiology*. Ripley KL, Ostrow H (eds). IEEE Computer Society, Long Beach, 89
20. Vandorpe A, Piessens J, Willems JL, De Geest H (1987) Unexplained chest pain with normal coronary arteriograms. A follow-up study. *Cardiology* 74: 436-443
21. Timmermans Ph, Willems JL, Piessens J, De Geest H (1988) Angina pectoris and coronary artery disease in severe aortic regurgitation. *Am J Cardiol* 61: 826-829
22. Van de Plas A, Willems JL, Piessens J, De Geest H (1988) Frequency of angina pectoris and coronary artery disease in severe isolated aortic stenosis. *Am J Cardiol* 62: 117-120
23. Brandt PWT, Partridge JB, Wattie WJ (1977) Coronary arteriography: Method of presentation of the arteriogram report and a scoring system. *Clin Radiol* 28: 361-365
24. Hamilton KK, Alderman EL, Silverman JF, Sanders WJ (1979) Computer reporting of coronary arteriograms using interactive graphics. In *Proceedings Computers in Cardiology*. Ripley KL (ed). IEEE Computer Society, Long Beach, 111-116
25. Krone RJ, Igielnik S, Miller P (1979) A computerized cardiology report generating/data management system. In *Proceedings Computers in Cardiology*. Ripley KL (ed). IEEE Computer Society, Long Beach, 469-472
26. Smith DN, Colfer H, Bryer JF, Pitt B, Kliman SH (1981) A semiautomatic computer technique for processing coronary angiograms. In *Proceedings Computers in Cardiology*. Ripley KL (ed). IEEE Computer Society, Long Beach, 325-328
27. Angelhed JE, Falk KJ (1984) A system for reporting and analysis of data from coronary angiography. In *Proceedings Computers in Cardiology*. Ripley KL (ed). IEEE Computer Society, Long Beach, 297-300
28. Bozzi G, Casolo F, Ascari A, Pastini I, Ferrante M, Difestano A, Verna E (1987) Descrizione interattiva e archiviazione di coronarografie e ventriculografie. In *Banche di Dati Integrate in Cardiologica*. Milazzotto F, Martiny W, Matta F, Pincirolì F (eds). Publ Centro Scientifico Torinese, 203-214
29. Johnson LW, Lozner EC, Johnson S, Krone R, Pichard AD, Vetrovec GW, Noto TJ, the Registry Committee of the Society for Cardiac Angiography and Interventions (1989) *Coronary Arteriography 1984-1987: A report of the Registry of the Society for Cardiac Angiography and Interventions. I. Results and Complications*. *Cathet Cardiovasc Diagn* 17: 5-10
30. Safian RD, McCabe CH, Sipperly ME *et al.* (1988) Initial success and long-term follow-up of PTCA in chronic total occlusions versus conventional stenoses. *Am J Cardiol* 61: 23G-28G
31. Baim DS, Ignatius EJ (1988) Use of percutaneous transluminal coronary angioplasty: results of a current survey. *Am J Cardiol* 61: 3G-8G
32. Jung D, Brennecke R, Kottmeyer M, Hering R, Clas W, Erbel R, Meyer J (1986) Tree-base: a query language for a time-oriented cardiology database system. In *Computers in Cardiology*. Ripley KL (ed). IEEE Computer Society, Long Beach, 381-384
33. Wallace AG, Rosati RA (1973) Computers can change cardiology. *Circulation* 47: 439

# A database for the follow-up of heart transplant patients

PH. DOUBLET

*INSA, Dept. Informatique, 69621 Lyon Cedex 03, France*

P. RUBEL

*INSERM U121, Hôpital Cardiologique, 69394 Lyon-Montchat, France*

G. DUREAU

*Hôpital Cardiologique, PAM de Transplantations Cardiaques, 69394 Lyon-Montchat, France*

P. VUILLARD

*Serveur Allô 26, Pommerol, 26470 La Motte Chalancón, France*

## Abstract

This chapter presents a description of a system which, by means of informatics, is aimed at improving the quality of the follow-up of cardiac transplant patients of the Lyon Cardiological Hospital.

Two databases have been set up: one, in the hospital, concerns the computerization of the medical file of cardiac transplant out-patients; the other, physically distinct from the first, is managed by an independent videotex server center. In this way, it is accessible from any minitel, which almost all physicians have and can be found in one out of four homes. This approach allows a 'bicephal' follow-up, i.e. close collaboration between the Lyon medical team and attending physicians who are geographically close to their patients, thereby ensuring greater patient comfort.

Developments in progress concern the creation of aid tools for decision making so as to better apprehend the content of medical files, the aim being to obtain a personalized follow-up that should help improve the overall efficacy of the system.

## 1. Introduction

Although the beginnings of cardiac transplants go back as far as 1967, thanks to C. Barnard in Cape Town, the real start of such operations occurred only in the early 80's with the arrival of a new immunodepressor agent: Ciclosporine [1]. The therapeutic effectiveness of cardiac transplants has proved to be convincing and surgical operations have multiplied exponentially, doubling each year between 1981 and 1986, to reach a limit which is fixed by the number of organs available [2]. The surgical technique is now well mastered, this



operation being relatively 'easy' as compared to other cardiac operations such as coronary bypass.

The main problem is in fact the follow-up, which presents two completely different aspects from a medical point of view: the follow-up during hospitalization, which immediately follows the transplant, and the follow-up in consultation, once the patient has left the hospital.

The follow-up in consultation presents, from an informatics point of view, the interest of a certain simplicity (the number of parameters is relatively limited and their value only evolves each month on average) and a great originality since it needs to take into account, over a very long period – from transplant to the possible death of the patient – a new parameter, time, without which the other parameters have no meaning.

The follow-up during hospitalization also needs to take time into account, but only over a few weeks, which restricts it to more classical problems that are essentially a matter for the state of the art, of intensive follow-up over a short period of patients who have undergone serious operations.

These reasons, added to priorities defined by the medical team, prompted us to initially favour the follow-up in consultation, while esteeming that the informatic solution of problems linked to this kind of follow-up should be of help, in the future, in the solution of problems linked to hospitalization [3].

Therefore, this chapter will only deal with the follow-up of cardiac transplant patients in consultation. We will first present the method of follow-up adopted by the Lyon team and the specific needs that ensue. Secondly, we will propose some solutions, in terms of methods that use informatic techniques, and we will particularly go into the question of analysis with regard to the conception of a data base for cardiac transplant follow-up in consultation. Finally, we will set forth the results that have been achieved, evaluate the acceptability of the system proposed by the medical team and envisage some improvements, mainly intended to make the present system much more flexible.

## **2. Material and methods**

### *2.1. Analysis of the medical team's needs*

*2.1.1. Follow-up method adopted by the Lyon team.* The follow-up in consultation of a cardiac transplant patient is difficult insofar as the risk of rejection can never be definitively dismissed, even for a patient who seems to be healthy several years after the transplant. Everywhere, until 1984, this follow-up was exclusively based on biopsy, a highly invasive test and the only one to be recognized by the international community [1].

Breaking with this philosophy, essentially in the concern for the well-being of patients, the team of Professor Dureau, in Lyon, which realises 5% of transplants in the world, has adopted a radically different method based on classical examinations of cardiology, such as electrocardiography or echo-

graphy [1]. This method has now been validated on patients of the Cardiological Hospital in Lyon. It gives very good results, in terms of survival rate, even often better than those obtained by teams that systematically perform a biopsy [1]. Moreover, it has been adopted by other medical teams, including that of Professor Cabrol in Paris.

It needs, however, a rigorous follow-up, based on two principles:

- consultations must be frequent, their periodicity ranging from one week to two months, depending on the state of the patient.
- one must take into account a great number of parameters, clinical, paraclinical, biological and therapeutic, and their evolution over time, from the beginning of the follow-up. A follow-up of quality can in fact only be carried out by the comparison of successive values of parameters.

*2.1.2. Needs of the medical team.* The Lyon team, which at present follows more than 200 patients, is encountering more and more difficulties in facing up to the requirements linked to the follow-up method chosen, and now has a double need:

- find a method that allows a better apprehension of the medical file's contents, which is not easy when we know that a consultation means more than a hundred parameters and that a follow-up of quality needs to consider the evolution of these parameters over time.
- decentralize the follow-up to attending cardiologists, which has become possibly by abandoning systematic biopsy, a procedure which is only workable in a hospital.

Such a goal, which it has become imperative to reach because of the increasing number of patients to follow-up, is also fundamental for the comfort of patients. These patients live in fact often very far from Lyon, and the frequency of consultations can be tiring and also handicapping, since cardiac transplant patients may otherwise lead a quasi-normal life and, in particular, may work.

However, decentralization comes up against the problem of the insufficient knowledge that most town cardiologists have of the specificities of the cardiac transplant patient's follow-up. That is why it is not possible to entrust, overnight, the entire follow-up of a patient to an attending cardiologist.

The method to envisage is a 'bicephal' follow-up, that is to say a follow-up that is jointly run by the Lyon medical team and the local cardiologist. For instance, patients may go alternately to Lyon and to their attending cardiologist.

Such a method nevertheless is faced with the problem of shared access to the medical file, a problem that is difficult to solve without the help of computers, because of the considerable distances between certain attending cardiologists and the Lyon medical team.

## 2.2. *Functional analysis in view of the setting-up of a data base for cardiac transplant patient follow-up during consultation*

The increasing difficulty in storing and retrieving data leads us to envisage first the setting up of a data base within the Cardiological Hospital, with the aim of computerizing the medical file. Such a computerization must, in order to avoid repeating failures often encountered in medical computer science [4, 5], become integrated into working and decision processes that are usually followed by the medical team. This is why the setting-up of the data base first needs an analysis of the thought processes followed by the physician during consultations and of the way he records, on paper, the results of consultations.

### 2.2.1. *Analysis of the thought processes of the physician during a consultation.*

In fact, a consultation is realized in two steps:

Step 1, the physician sees the patient, searches for clinical signs, asks questions, and consults laboratory results that are already available (pulmonary radiography, vectorcardiogram, etc.) He also takes some notes: a small summary and some questions about unresolved problems. In addition, the physician may prescribe some examinations that will have to be made, either just after the consultation, or before the next consultation. These prescriptions are determined either by the results of the current consultation or by invariant constraints such as 'such an examination must be done every X consultations'.

Documents stemming from this first step are as follows:

- results of the medical examination (auscultation, blood pressure, etc.), which correspond to several perfectly defined parameters, the protocol of clinical examination being invariant from one patient to another;
- answers to questions asked by the physician, which are on the contrary very ill-defined, insofar as each patient generally presents a particular problem;
- several quantitative values and/or several synthetic qualitative values, obtained from results of tests already available;
- a synthesis: general impression, unresolved problems.

Step 2, when results not previously available arrive from laboratories, the physician considers the entire consultation (clinical examination, first results from laboratories, report of the first step and new results from laboratories), compares it with previous consultations, and establishes a therapeutic protocol.

Documents resulting from this second step are:

- complementary list of quantitative and/or qualitative values resulting from test findings;
- list of therapeutics (type, posology).

2.2.2. *Typology of data.* Information to be stored in the data base may be divided into two broad groups, linked to two types of data: data that are invariant over time and time-varying data.

## 1. Invariant data

We find in this group:

- administrative data of patients (identification number, family name, date of birth, address, phone number etc.). We consider that these data are 'invariant' insofar as they are not a function of time: they may evolve (e.g. address, phone number) but, in this case, it is the more recent information that is stored in the data base, and previous information is overwritten;
- several invariant medical data, such as blood group;
- administrative data of physicians (name, address, etc.);
- administrative data of care centers (care-units, laboratories, etc.).

## 2. Time-varying data

Each time the value of such a parameter is obtained it is recorded in the data base's memory and the date of this input is also stored, but values previously recorded are not deleted. Thus, each parameter should be considered as a time series, thereby allowing the notion of evolution to be taken into account.

This group may be divided into two sub-groups, respectively corresponding to:

- consultation data, which are collected and recorded by the physician during either the first or second step of the consultation;
- supplementary data about the transplant(s) and the hospitalization(s), which should be considered as an initial working basis with regard to follow-up in consultation. These parameters will afterwards be the link between the 'consultation' data base and the 'hospitalization' data base.

We find in this second sub-group:

- parameters indicating the cause of hospitalization, the complications occurring during hospitalization, and the patient's transfer to various care-units (this last point is a good indication of the rate of the patient's convalescence);
- parameters concerning the transplant(s):
  - information about the donor: age, sex, weight, cause of death, blood group, HLA group. These parameters are important as they may help to account for a complication occurring during the follow-up in consultation.
  - pre-transplant checkup: cardiac pathology that occasioned the transplant, linked pathologies, pulmonary resistance, cross match, renal function, respiratory functional test, cerebral scanner, virology, pre-transfusion protocol. These parameters are all, except pulmonary resistance, qualitative, because our purpose is not for the time being to make an exhaustive check-up, which is the function of a data base for the follow-up during hospitalization. Only a few data, which permit a synthesis with the follow-up in consultation in mind, are of interest here.
  - course of surgery: date and time of operation, name of the surgeon,

name of the anaesthetist, organs transplanted other than the heart (some heart/lung transplants have already been done at the Cardiological Hospital), time of extra-corporeal circulation, lowering of the extra-corporeal circulation temperature, total time of suture, duration of ischemia, time of warming-up, hemodynamics after stabilization, cardiotoxic administered, possible complications.

- parameters corresponding to the admission and discharge check-ups of each hospitalization, which are almost identical to the parameters documented during the consultations.

*2.2.3. Identification of consultation parameters requiring storage in the data base.* The procedures presented below attempt to answer two imperatives, unfortunately antagonistic: the necessity of storing a great amount of data, in order to obtain a follow-up of quality, and the obligation of building a system that is not too cumbersome, otherwise it may be refused by the medical team.

The identification made here, which remains at a conceptual level, proposes the taking into account of a fairly large number of parameters, nevertheless strictly selected among all the parameters it would have been possible to consider. We will see in the discussion that not all these parameters were retained at the time of physical implementation, the idea being to begin with a small data base that will progressively build up, in parallel with the growing need of the physicians and with the computer techniques that we are developing to better apprehend the content of the medical files.

Parameters to be stored in the data base were chosen according to the following principles:

- all the parameters corresponding to the usual protocol of clinical examination are stocked. Some of these parameters are qualitative, others are quantitative. We find in this group
  - functional state: coded from 1 to 4, according to the NYHA classification;
  - functional test: number of floors the patient is able to climb before dyspnea;
  - weight;
  - height, which is sometimes variable because cardiac transplant patients may have problems of compression of the spinal column;
  - blood pressure;
  - qualitative parameters indicating whether there are any oedemas.
- we have reserved a few fields devoted to the storage of patient problem-oriented diagnosis statements of peculiar interest (from a practical point of view, up to three statements of this type may be considered). Input is done by the physician, who 'enters' a key-word. A generic search algorithm then tries to find standardized key-words that are nearest to the key-word entered by the physician - the search is made in an already recorded list - and proposes a sub-list among which the physician has to fix his final choice.

The standard list of key-words must be able to evolve as the physician wants. However, any set of statements must necessarily include the following clinical

signs which are most frequently observable among transplant patients:

- anorexia
  - hirsutism
  - icterus
  - cyanosis
  - jugular pressure
  - hepatomegaly
  - hepatojugular reflux
  - ascites
  - neurological problems
  - digestive problems
  - respiratory problems
  - muscular problems
  - articular problems
  - bone problems
  - cutaneous problems
- for each type of biological or functional investigation, selection of a subset of quantitative parameters that are the most significant for cardiac transplant follow-up and creation of several qualitative parameters which in fact correspond to an interpretation, by the physician, of all the quantitative parameters.
  - for every drug usually administered to cardiac transplant patients, selection of parameters corresponding to the posology and, if needed, to the determination of the amount of drug in the blood, to secondary effects and to drugs used to reduce these secondary effects.  
For instance, the parameters selected for Ciclosporine are
    - posology;
    - determination of monoclonal and polyclonal plasma rate and polyclonal and monoclonal blood rate;
    - secondary effects: creatinemia and uricemia rate;
    - drugs associated with secondary effects: posologies of Zyloric, Desuric and Colchicine.
  - the addition of a limited number of free items, essential because it is not possible to foresee everything, but limited because they would be difficult to utilize automatically.

### 2.3. Structuring of data: *C(onceptual) D(ata) M(odel)*

The Conceptual Data Model which, for the essential, is the consequence of the analysis presented above, is given in Figure 1. The diagram, although classical, requires some comments:

- the formalism used is that of Merise, very widespread in France [6, 7].  
Rectangles represent entities and ovals represent associations. The entities are linked together by means of the associations. Numbers zero to  $n$  indicate the cardinality of the relation.  
On the Conceptual Data Model presented, we may read for instance that a patient has been to zero, one or up to  $n$  consultations and that any consultation is linked to one and only one patient.
- check-up of admission and discharge is regarded as a particular consultation (cf. the associations between the entity 'hospitalization' and the entity 'consultation'). Such a choice, directly inspired from the method used in an

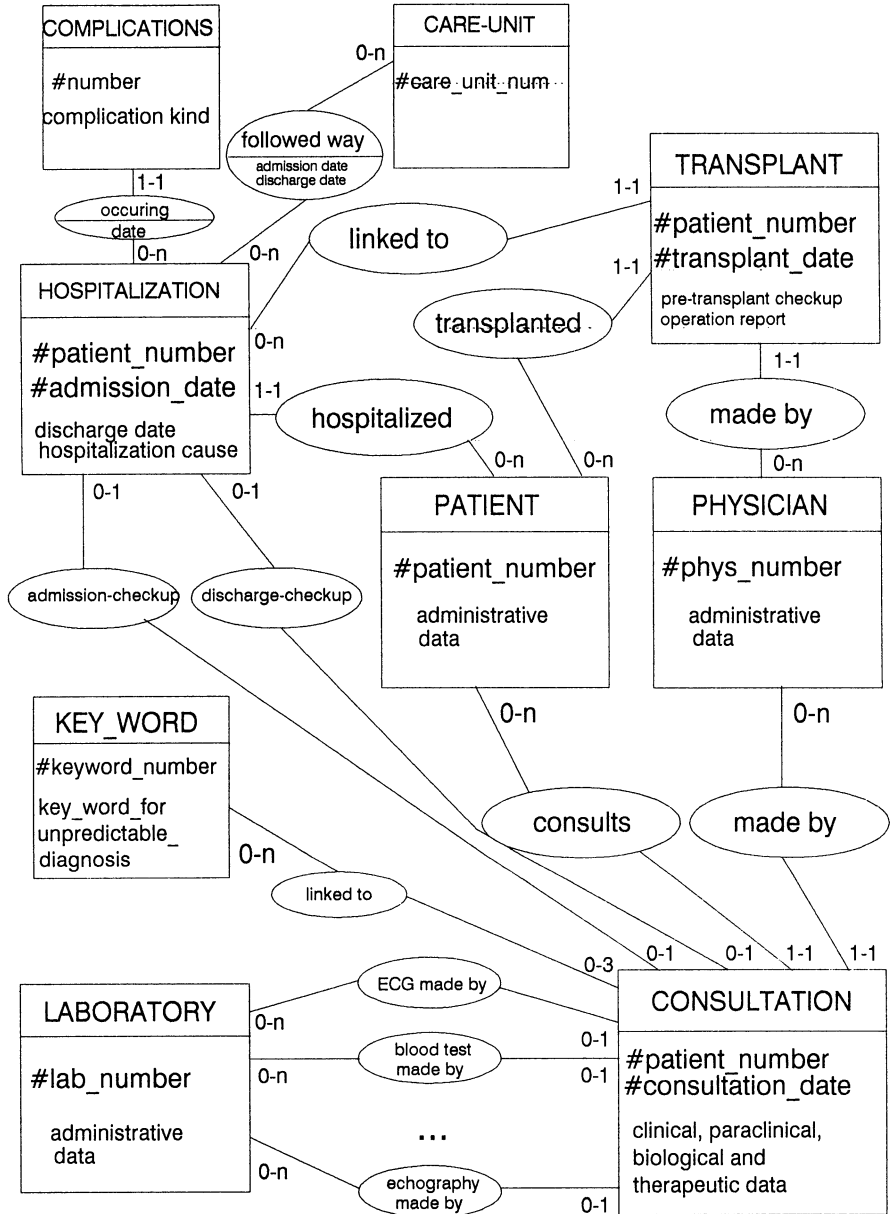


Fig. 1. Conceptual Data Model.

American project for the design of a data base for cystic fibrosis [8], avoids useless complication of the Conceptual Data Model and so of the man/machine interface. In addition, the concept of hospitalization can, with this method, be easily integrated into the follow-up in consultation, without changing too much the usual decisional process.

- the entity 'laboratory' identifies where the tests have been realized and thus

may permit, through conversion tables, a correction of results in order to render comparable the results supplied by different laboratories.

#### *2.4. Decentralization of the follow-up to attending cardiologists: use of a French videotex-based communication tool, the minitel*

Although our first objective is the design of a data base internal to the Lyon Cardiological Hospital that would permit the computerization of the medical file, another objective, very important in terms of public health, is the bicephal follow-up of cardiac transplant patients. The implementation of such a strategy requires a shared access to the medical file so that attending cardiologists can visualize and update the data base. Such a goal, from a computer-science point of view, is feasible: it would be sufficient that attending cardiologists have a personal computer or a terminal and that they be connected, by a public network (TRANSPAC, in France), to the computer at the Cardiological Hospital in Lyon.

Such a solution will probably be realistic in a few years, but now it seems rather premature, as most cardiologists do not yet have a personal computer and, because they generally follow only a very few, often only one, cardiac transplant patients, it is not feasible to ask them to buy a computer for the follow-up of so few patients. Furthermore, it would be also highly desirable if all physicians, including general practitioners, could have access to the data base (if they have, of course, the patient's authorization), given that the profile of a transplant patient is very particular and that every one of his health problems must be carefully treated. Finally it is of course not possible to connect to our data base all the physicians who might one day treat a cardiac transplant patient!

For these various reasons, the use of a new computer technique, videotex – in France, the minitel – seems to be most interesting.

*2.4.1. The French videotex standard: TELETEL.* Since the middle of the 80's, France has been equipped with a very powerful tool: the TELETEL, the name of the French videotex standard and, with it, the minitel. The minitel is a terminal of the alphanumeric type, designed to be connected to the French telephone network and to permit access, through this network, to numerous server centers that maintain data bases on varied subjects and develop videotex applications.

The minitel is today very widespread in France. The TELECOM (French telecommunications service) installs it free of charge, and there is no rent to pay. Other countries have also developed videotex systems [9, 10, 11], but at the present time only France offers such a wide diffusion of terminals. More than five million minitels, one in five telephone lines, had been installed by the end of 1989.

New videotex applications become available every day. Minitel owners, provided they are duly authorized, have access to them.



There are several different systems of rate fixing, generally based on a charge depending on the time of connection, to which can be added a forfeit fee to the server center. Each server center can choose its system of charges and may even offer its services completely free of charge.

To implement an application on minitel, there are two solutions:

- create one's own server center, which may be a simple personal computer with a suitable interface card;
- ask an existing server center to develop the application according to one's own specifications, and, possibly, connect one's own computer to the server center in order to permit data exchange.

Some medical applications have already been developed on minitel. Most of them have an informative or even a teaching mission, often in the form of computerized medical encyclopaedias or expert systems [12].

There have been, on the other hand, very few experiments concerning implementing medical files on minitel, because of doubts concerning the confidentiality of medical data. The French legislation is very strict on this subject and many projects have been refused by the CNIL (Commission Nationale Informatique et Libertés) or by the 'Conseil de l'Ordre', two bodies responsible for the safeguarding of individual rights.

*2.4.2. Use of the minitel for the bicephal follow-up.* In the present state of the technology, the minitel seems, if we overcome the problems of confidentiality, to be a method that is very well adapted to the problem we have to solve, i.e. making medical files accessible from any place, at any time and by any physician who has valid authorizations.

Furthermore, the minitel may also become a tool of information and even perhaps of teaching of physicians. It is indeed possible to envisage the implementation of specific aid tools, for instance in order to prevent any drug incompatibility with patent medicines used in the treatment of cardiac transplant patients.

Finally, the minitel may also be used for electronic mail exchanges: the attending physician drops a question in a 'mailbox' and the Lyon medical team answers it in the same way. Such a process can only be used if there is no urgency. In spite of this restriction, it can be of great usefulness insofar as it might dispense with incessant phone calls, which often needlessly trouble the medical team.

There is practically no doubt about the ready acceptance of the minitel by cardiologists, given that the use of the minitel does not involve any particular investment. The minitel is lent by TELECOM and, in view of the small number of patients followed up by each attending physician, it is of course out of the question to ask them to pay a rental to the server center that holds the data base. The only financial constraint is the payment of the rates corresponding to the connection times, but this is far from being exorbitant and is quite comparable to the amount that physicians pay when they have to phone a colleague.

The only problem therefore is guaranteeing the confidentiality of infor-

mation, a *sine qua non* condition for the acceptance of the project by ethics committees. An examination of the criteria of these committees allows us to understand that the only process that we can envisage is the following:

1. It is the patient, and only the patient, who may ask for his medical file to be put on minitel and who gives, and whenever he wishes, withdraws, the right of access, thanks to a code, to his medical file. It means that the patient is the 'creator' and so the owner of his file.

The situation is very similar to the one where we use paper files which, from a medico-legal point of view, also belong to the patient. Risks of violation of confidentiality admittedly still exist because a computer system cannot be completely protected against all attacks. But these risks are not greater than in the case of a non-computerized file, to which it is always possible to envisage that an ill-intentioned person may have illegal access.

2. The data base of the minitel application, which from a conceptual point of view is a sub-set of the Cardiological Hospitals internal data base, must be physically separated from the internal data base. In fact, ethics committees cannot tolerate the systematic scientific exploitation of data that are acquired on the minitel, and so these data cannot be mixed with data recorded at the hospital and which can be scientifically exploited.

Thus, we will have two data bases, physically separated and nevertheless partly made up with the same parameters. The inconvenience of such a system is twofold:

- the Lyon medical team must visualize the data that have been entered by the attending physicians on the minitel, which has fewer ergonomic possibilities than the local system;
- the keyboarding of consultations made at the Cardiological Hospital must be done in duplicate, on the minitel in order to permit the bicephal follow-up and on the local system in order to collect more parameters than on the minitel and to use the ergonomic possibilities of this system.

To overcome these drawbacks, we have envisaged the coupling of the server's data base with the data base in the Cardiological Hospital, the information transfer between the two data bases being of course subject to the authorization of the patient. Such a solution is not in contradiction with ethics criteria: the separation between the two data bases still really exists insofar as the patient may at any time withdraw the authorization of transfer and so of scientific exploitation of data collected on minitel.

The patient, the only one allowed to make decisions in this matter, may thus give authorizations at different levels:

- no authorization;
- authorization to put the file on minitel but banning the transfer of data to the Cardiological Hospital;
- authorization to put the file on minitel and to transfer data in the direction 'hospital towards minitel server';

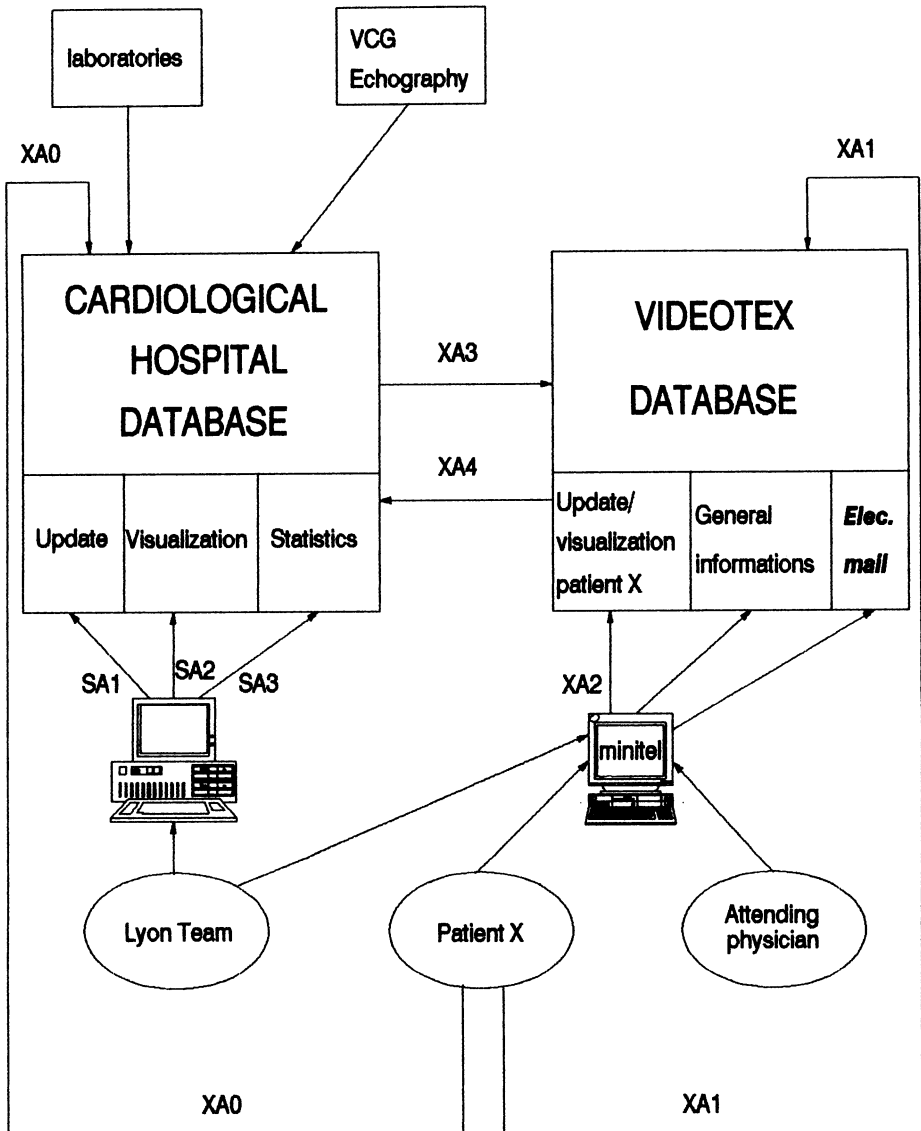


Fig. 2. System architecture for the follow-up of heart transplant out-patients.

- authorization to put the file on minitel and to transfer data in the direction 'minitel server towards hospital';
- authorization to put the file on minitel and to transfer data in both directions.

The overall architecture of the system for the follow-up of heart transplant patients may thus be represented as in Figure 2. Abbreviations are as follows: XA0: patient X gives the authorization to put his file on the data base in the Cardiological Hospital.

XA1: patient X gives the authorization to put his file on the minitel data base.

- XA2: authorization of patient X to a given physician for the access to his medical file through the minitel (in practice, the patient gives his secret code to the physician or enters himself the code in order to give a selective access right to the physician).
- XA3: authorization of patient X for the transfer of data in the direction 'internal data base towards minitel data base'.
- XA4: authorization of patient X for the transfer of data in the direction 'minitel data base towards internal data base'.
- SA1: access right, given by the administrator of the data base in the Cardiological Hospital, for the updating of the medical file.
- SA2: access right, given by the administrator of the data base in the Cardiological Hospital, for the visualization of the medical file.
- SA3: access right, given by the administrator of the data base in the Cardiological Hospital, for the realization of statistics on data contained in the medical file.

One should however note that the SA-type access rights may change from one user to another. A nurse from a given laboratory, for instance, may have an SA1 limited to data measured by that laboratory.

### 3. Results, evaluation and discussion

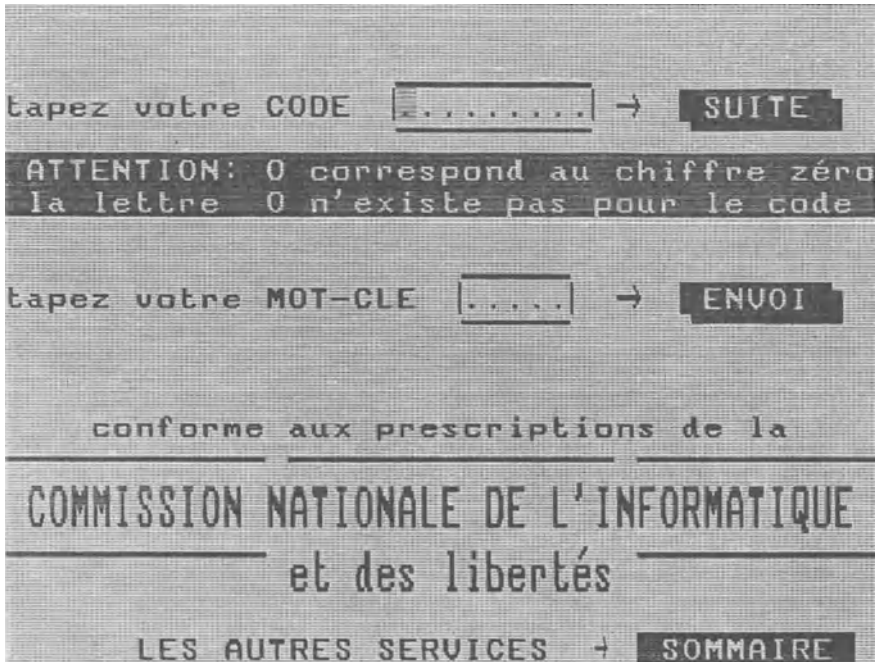
#### 3.1. *Implementing the minitel application*

Because the decentralization of the follow-up had been defined as a priority by the medical team, it was decided to implement first the minitel application. This has been designed by a server center specializing in medical applications and managed by a surgeon, Professor Vuillard, who implemented patients' medical files following the concepts described in an analysis realized in 1988 [3].

The items that have been chosen for the data base held by the server center are only a sub-set of the items that had been retained for the data base in the Cardiological Hospital. It is in fact not possible, given the limited ergonomics offered by the minitel, to ask a physician to work on more than two or three screens of data acquisition. Thus, only the most important parameters that are likely to be measured outside a hospital have been chosen.

In its present version, the application is composed of

- a main menu;
- a patient-identification screen display (Figure 3). This identification is made both by means of a secret code, automatically allocated by the minitel server at the first connexion, and by a confidential keyword, chosen, also at the first connexion, by the patient;
- two screens for the updating of the medical file (Figure 4);
- a screen for the tabular display of the evolution, during previous consultations, of the most important parameters (Figure 5);
- several aid screens, in particular concerning drug incompatibilities and vaccinations.



*Fig. 3.* Minitel screen-display for patient identification. To access a patient file, the system requires first entering the numerical identification code that had been automatically allocated at the first connexion time, then entering the patient-password which had been chosen by the patient himself at the first system entry.

The realization of the application itself was achieved relatively quickly. On the other hand, Professor Vuillard's team had to give much thought to problems of confidentiality and spent a lot of time in meetings with ethics committees. These steps were all the better received as they were undertaken by physicians and not by computer scientists.

The project was finally accepted and the minitel application has been running since the beginning of 1989.

This application has been well accepted by the several attending physicians who follow up cardiac transplant patients, but it is too early to evaluate the system, insofar as the minitel application will really develop only when the server data base is connected up with the data base in the Cardiological Hospital. The coupling between these two data bases will have to be done as soon as possible, perhaps before the end of 1990, the main problem being again the acceptance of such a project by ethics committees.

<b>GREFFE COEUR 1</b>		NOM:	
fiche clinique		DATE EXAMEN: 13/03/90	
		DATE GREFFE: 26/01/88	
MEDECIN EXAMINATEUR: DUREAU			
VA BIEN (O/N): 0	ASTHENIQUE (O/N): N		
ESSOUFFLE (O/N): N	HEPATOMEGALIE (O/N):		
GALOP (O/N):	INSUF. TRICUS. (O/N):		
	INSUF. MITRALE (O/N):		
T.A.: 160/100 mmHg	RAP. CARDIO-THOR:		
IND. SHUMWAY: mm	FREQUENCE CARD.:		
<b>ECHO</b> OEDEME (O/N):	DILAT. CaUR DT (O/N):		
TRIUG: msec	ALTER. CONTRAC (O/N):		
<b>BIOPSIE</b> STADES 0, 1-, 1+, 2, 3 ou 4:			
PROCHAIN CONTROLE: 22/05/90 CONSEIL: <b>GUIDE</b>			
		Page suivante +	<b>SUITE</b>
Autre fiche ↑↓		RETOUR ou	<b>SUITE</b>
1/2		Pour modifier tapez MOD +	<b>ENVOI</b>
		Ligne suivante +	<b>SUITE</b>
<b>GREFFE COEUR 1</b>		CONDUITE A TENIR + <b>GUIDE</b>	
fiche de soins			
CYCLO pos.	350 mg	AC. URIQUE	mimo
Taux pl.*	71 ng/ml	DESURIC	comp
CREATININE	121 mimol	CHOLESTER	mmol
IMUREL	2,5 comp.	HYPOLIP.	comp
GL. BLANCS	4.56G/l	LASILIX	40X1comp
GL. ROUGES	4.20T/l	Ca BLOQU.	20X2comp
BIL. TOT.	mimol	IN. ENZ. CO	25X2comp
PHOSP. ALC.	U/l	VASO-DIL.	comp
SOLUPRED	17,5mg	ASPIRINE	mg
CORTISONE	mg		
GLYCEMIE	mmol		
*DOSAGE CICLO (1) monoclonal (2) polyclonal (3) sang tot. (4) plasma (5) HPLC :			
		Page précédente +	<b>RETOUR</b>
Autre fiche ↑↓		<b>SUITE</b> ou	<b>RETOUR</b>
2/2		EN MODIF.: Ligne suivante +	<b>SUITE</b>
		Pour enregistrer +	<b>ENVOI</b>

Fig. 4. Minitel screen-displays for patient consultation data entry. The items presented in screen 1/2 are clinical or paraclinical ones: functional state, blood pressure, cardio-thoracic ratio etc. Screen 2/2 is dedicated to the updating of biological and therapeutic parameters, such as posology of Ciclosporine, level of Ciclosporine in blood, secondary effects of Ciclosporine, etc.

G.C.1						
		1	2	3	4	5
		VA BIEN	CREATIN	GL. BLANCS		
		2	4	6		
		CYCL.dos	IMUREL	CORTISONE		
N°	DATE	1	2	3	4	5
001	17/05/88					
002	31/05/88					
003	21/06/88					
004	12/07/88					
005	09/08/88					
006	13/09/88					
007	18/10/88					
008	22/11/88					
009	23/12/88					
010	20/01/89	0	250	123	3	5.65
011	10/03/89	0	400	103	3	7.25
012	11/05/89	0	400	116	2,5	4.08
013	11/01/90	0	350	111	2,5	5.12
014	13/03/90	0	350	121	2,5	4.56

1 VA BIEN: OUI/NON 23456 unités cf fiche  
suite de la liste des fiches + SUITE  
fiche complète N° ... votre choix + ENVOI

Fig. 5. Tabular display of the evolution of some parameters on the minitel. Items which have been selected are

1. patient's general state;
2. level of Ciclosporine;
3. creatinemia;
4. Imurel posology;
5. white corpuscle count;
6. cortisone posology.

### 3.2. Creation of a data base in the Cardiological Hospital of Lyon

The realization of a data base in the hospital was begun, following the principles explained in 'Material and Methods'.

The development was made using the data-base management system ORACLE, which presents the major advantage of very good portability, and under the operating system MS-DOS, which will soon be replaced by UNIX in order to permit multiple access to the internal data base.

A first version, completed in the summer of 1989, was developed according to specifications made in collaboration with the medical team. However, this version has never run in routine because physicians, seeing what a system that requires the entering of 100 items at each consultation means in practice, have judged it too unwieldy. For this reason, we are now trying to draw up a new, very simplified version which is a kind of compromise between the first version and the minitel application, and which we will complete as physicians require.

T H E R A P E U T I Q U E   C Y C L O S P O R I N E		1/1
<b>POSOLOGIE ET DOSAGE</b>	<b>EFFETS TOXIQUES ET THERAPEUTIQUE</b>	
Posologie 100 x 3 (mg/prise x prises/jour)	<b>ASSOCIEE</b>	
Taux plasmatique : - polyclonal 187 (mg/mole) - monoclonal 400 (mg/mole)	Créatinémie 200 (mmol/l) Uricémie 215 (mmol/l)	
Taux sanguin : - polyclonal 950 (mg/mole) - monoclonal 234 (mg/mole)	Posologie Désuric 2 (comp/jour) Zyloric (comp/jour) Colchicine (comp/jour)	
No matricule patient 881200001A Nom patient Prénom patient Philippe_	Date consultation 14/12/89 Né(e) le 13/02/65 Greffé(e) le 08/05/89	

Fig. 6. A partial view of the Cardiological Hospital data base for the update of the items related to Ciclosporine therapy. In the left half of the screen, posology and level of Ciclosporine, in the right, secondary effects and associated therapeutics.

In Figure 6 we present the most recent version of the screen for the updating of the therapeutic protocol for Ciclosporine. This screen is only a part of the set of screens for the updating of a consultation, which are themselves a part of the set of screens for the updating of the entire database, which also contains some screens for the updating of entities such as administrative data and hospitalizations. The box at the bottom of the screen, which is also found on all updating screens, allows a recall of the principle characteristics of the patient being examined.

At any time during the updating of a consultation the user may call

- aid screens, also available during the updating of entities other than the consultation;
- information screens: display of physicians' administrative data, essential for the decentralization of the follow-up (Figure 7); tabular display of the evolution of some of the parameters of the patient on whom the physician is working (Figure 8); display of the patient's administrative data.

The new version will run on a personal computer during the second quarter of 1990 but the system will present a really attractive tool for the follow-up of all the cardiac transplant patients only when the internal data base has been connected with the minitel data base.

Further, the discarding of certain parameters, which had nevertheless been judged very important in the analysis, is not very satisfying. That is why it has appeared necessary to search for methods that can take into account many more parameters, but without making the system any more difficult to exploit.

The following section sets out the state of our thought in two fields: data



COORDONNEES ADMINISTRATIVES DES MEDECINS				
Nom	Prénom	Spécialité	Adresse	Téléphone
Chuzel	André	Chirurgien card	28 av doyen Lép	72 35 72 35 (
Dureau	Georges	Chirurgien card	28 av doyen Lép	72 35 72 35 (

APPUYEZ SUR LA TOUCHE <ESC> POUR RETOUR A L'ECRAN DE CONSULTATION.

Fig. 7. A selected view of the items contained in the Cardiological Hospital data base for the display of the physicians' administrative data (family name, first name, address, speciality, phone number).

TABLEAU D'EVOLUTION DE QUELQUES PARAMETRES				
DATE	NOM MEDECIN	ETAT FCT	TENSION	CYCLOSPORINE
22-JAN-90	Chuzel	1	120 / 80	400
15-JAN-90	Dureau	3	170 150	350
12-JAN-90	Dureau	1	120 70	350
15-DEC-89	Dureau	1	135 110	300
15-NOV-89	Dureau	1	120 70	300
15-JAN-89	Dureau	4	190 130	

Fig. 8. Tabular display of the evolution of some parameters contained in the Cardiological Hospital data base.

acquisition, which represents too much work if there are a lot of parameters to keyboard, and the visualization of data: how to find the most important facts in an enormous file?

### 3.3. *Current developments for improving the ergonomics of the system for the follow-up of cardiac transplant patients*

#### 3.3.1. *How to lighten the work of data acquisition?*

*Linking of the internal data base with laboratories.* The analysis realized at the time of the conception of the data base made it apparent that most data are supplied by hospital laboratories, most of which are computerized. That is why the main way to improve the present system of data acquisition seems to be the coupling of the data base with each laboratory in order to directly retrieve these results. A system of the 'file transfer' kind will probably be chosen. We will, moreover, have to envisage the creation of a 'filter', insofar as data from laboratories do not all necessarily have any interest for the follow-up of cardiac transplant patients.

*Increase in the number of data-acquisition screens.* In most traditional database management applications, the user only disposes of one series of screens, which permits the updating of the whole data base. In the best case, if the man/machine interface is well designed, the user can proceed relatively quickly to the screen where the information he needs appears.

The first idea, already amply implanted in the version to be routinely used during the second quarter of 1990, is to increase the number of 'angles' from which it is possible to view the data base, i.e. to increase the number of data-acquisition screens so that the user may choose those that are nearest to how he wishes to see the data base. We may thus envisage screens enabling either a complete updating (main case) or only the updating of

- parameters of the first step of consultation;
- parameters of the second step of consultation;
- parameters from a given laboratory;
- 'minitel' parameters, i.e. of the items from the server data base, which represents a kind of minitel emulation;
- etc.

This first method is dependent on the state of the art and thus not very difficult to implement. On the contrary, there is a risk of drastically complicating the whole process if we wish that the angles of view of the data base should not be arbitrarily fixed, but should be dynamically defined, in real time, according to the patient's profile. Thus, the physician might, for each patient, have access to one or more updating screen(s), containing only the items judged most important for this patient at the time of the consultation.

The problem lies in the definition of what is important for a patient. It is of course not possible nor desirable that each patient's profile be definitively defined, since this profile is of course evolutive. The definition of the profile and the generation of the corresponding data-acquisition screens linked to it must therefore be automatic. This requires solving problems which for the moment lie in the field of research. However, this aspect will probably not yet

be looked into deeply, insofar as data visualization problems seem to have priority for improvement of the follow-up quality.

### *3.3.2. How to help the physician to better apprehend the information content of very detailed files*

*Development of computer tools allowing a better visualization of data.* With the system to be routinely used in the second quarter of 1990, we will of course be able to display data, but only on data-acquisition screens, which do not allow any overall vision of the information content of the medical files nor permit any but a poor perception of evolution over time. That is why it appears essential to improve this system by proposing tools enabling a more suitable visualization of file contents. Two ideas will be implemented:

- creation of a program for the visualization of quantitative parameter evolution curves, which would allow a good representation of the concept of time [3];
- creation of a program allowing automatic generation and printing of a pre-consultation synthesis report for each patient having an appointment that day. This report would contain facts judged abnormal or ambiguous over the last two or three consultations [3].

The problem is how to define abnormalities. Indeed, we could only implement one-to-one, threshold-based decision trees, but it soon appeared that the notion of time must be taken into account. This led us to take an interest in finding solutions to complex problems, still in the field of research, which are evoked below.

*Concept of personalized representation of knowledge and data.* The concept of personalized representation of knowledge and data is certainly one of the most interesting lines of research that we hope will result in a high-performance decision tool for cardiac transplant follow-up. The idea is the following: in addition to the screen displays that we have already described in the previous section, we should also be able to propose to the physician any new kind of visualization, of any new parameter or its trends, their display being automatically configured according to rule-defined criteria especially based on either

- the medical profile of the patient, which will be automatically obtained by taking into account the notion of time and possibly an analogy with previously encountered cases;
- the interest given by the physician to the patient's displays that had been automatically proposed during previous consultations or to the display of another patient who had a profile similar to the one of the current patient (Did the physician carefully look at what was proposed? Did he ask for other kinds of visualization or for the visualization of other parameters than those proposed?).

Such a system would probably meet the first requirement of the medical

team, set out at the beginning of this chapter: 'find a method permitting a better apprehension of the information content of the medical file'.

#### 4. Conclusion

We have presented a system which aims at the improvement of the quality of the follow-up in consultation of cardiac transplant patients, both in terms of comfort for patients and perhaps in terms of survival rate, mainly because the notion of bicephal follow-up, which is possible thanks to the minitel, will allow the decentralization of the follow-up towards attending cardiologists who are geographically near their patients, and also because the computerization of the medical file should result in a better apprehension of its information content.

The follow-up by minitel, already operational, will soon be backed up by a more complete system, representing a major turning point. Supplementary developments are envisaged in order to make this system more flexible, the final goal being to be free of the usual rigidity of medical software and to end up with a personalized follow-up of patients.

#### Acknowledgments

The authors are indebted to Dr. J. Carew for reviewing the original manuscript.

#### References

1. Dureau G, Chuzel M, Termet H *et al.* (1987) Bilan de l'activité de transplantation cardiaque de l'Hôpital Cardiologique Louis Pradel de Lyon. *Lyon Chirurgical* 83: 198-203
2. Fragomeni LS, Kaye MP (1988) The registry of the international Society for Heart Transplantation: Fifth Official Report-1988. *The Journal of Heart Transplantation*, 7: 249-253
3. Doublet P (1988) Conception d'un système d'information pour le suivi des transplantés cardiaques. Rapport de DEA: Institut National des Sciences Appliquées de Lyon, 152
4. Lacheze-Pasquet P (1984) Informatisation des hôpitaux. *Techniques hospitalières*, no. 461: 39-55
5. Van Brunt E (1984) Some problems of the 1970's: a clinical perspective. In *Information systems for patient care*. Bruce I. Blum (ed) New York: Springer-Verlag, 27-39
6. Tardieu H, Rochfeld A, Coletti R (1989) *La méthode Merise*, tome 1, Principes et outils. Paris: les éditions d'organisation, 328
7. Tardieu H, Rochfeld A, Coletti R (1989) *La méthode Merise*, tome 2, Démarche et pratiques. Paris: les éditions d'organisation, 461
8. Ceri S, Warwick W (1986) Design and development of CFDB: a standard database for cystic fibrosis. In *Medinfo 86, Proceedings of the fifth conference on medical informatics*, Washington, October 26-30, 1986. R Salamon, B Blum, M Jorgensen (eds). Amsterdam: North Holland, 522-526.
9. Nishimoto K, Inoue H, Murata Y, Mogami T (1986) Comprehensive medical information service by videotex in Japan. In *Medinfo 86, Proceedings of the fifth conference on medical Informatics*, Washington, October 26-30, 1986. R Salamon, B Blum, M Jorgensen (eds).

- . Amsterdam: North Holland, 1022-1023
10. Suss R, Malter M (1986) Videotex as a new medium for a 'cancer encyclopedia'. In Medinfo 86, Proceedings of the fifth conference on medical informatics, Washington, October 26-30, 1986. R Salamon, B Blum, M Jorgensen (eds). Amsterdam: North Holland, 1028-1029
  11. McDonald MP (1986) The videotex-based personal health information system. In Medinfo 86, Proceedings of the fifth conference on medical informatics, Washington, October 26-30, 1986. R Salamon, B Blum, M Jorgensen (eds) Amsterdam: North Holland, 1024-1027
  12. Palmer G, Morlat P, Prague Y, *et al.* (1986) A videotex system to help the general practitioner for prescribing antibiotics. In Medinfo 86, Proceedings of the fifth conference on medical informatics, Washington, October 26-30, 1986. R Salamon, B Blum, M Jorgensen (eds) Amsterdam: North Holland, 1030-1032.

# The K.U. Leuven Coronary Surgery Data Base: a clinical research data base

P. SERGEANT and E. LESAFFRE

*Department of Cardiac Surgery, Department of Epidemiology, Gasthuisberg University Hospital, Katholieke Universiteit Leuven, 3000 Leuven, Belgium*

## 1. Introduction

The K.U. Leuven Coronary Surgery Data Base (KULCSDB) [1] was started as a clinical research data base in 1980. It structures the preoperative, operative, postoperative and follow-up data of all the patients operated for isolated coronary pathology since 1970 at the K.U. Leuven University Hospital of Gasthuisberg in Leuven Belgium. The current number of records in the data base is 8200 patients with more than 1000 variables for each record. The sum of the real follow-up structured in the data base is 35 000 years. The technology of such a large data base and the correct study of time-related events was not available in Europe at the time and therefore many problems and pitfalls were encountered. The purpose of this chapter is to analyze these problems and to communicate them so that the information becomes available to those interested.

## 2. Purpose of the data base

Many data bases are started and most are halted after some months, because the purpose of the data base was never very clear from the beginning. The final purpose should be very well defined before even structuring the datafiles and fields. It is possible to identify administrative medical and scientific medical data bases.

### 2.1. Administrative medical data bases

Administrative medical data bases are necessary for hospital management purposes or for national registry purposes. They can help the National Health Ministers in obtaining an insight into the workloads and if possible efficiencies of the hospitals. They might, as in Belgium, base the daily cost price of the hospital bed or of the daily hospital laboratory cost on packages systems, defined by pathologies treated in that institution. The early experience in

Belgium has proven that the data are not analyzed in a scientific way and that the conclusions have not been better than the ones taken without any 'pseudo-scientific' basis. Administrative data bases can also be used by hospital management to study the efficiency of the different departments. The experience at the K.U. Leuven is that their analyses are usually only on a very crude and univariate basis, without bringing in any scientific aspects in the analysis and therefore prone to a lot of error. The occupancy of a department can be very high but with a very slow turnover of patients and therefore not very efficient. A couple of chronic patients can reduce the turnover of a whole department which is still running very efficiently, so the median value of hospital stay should be used instead of the mean hospital stay. Very few administrative databases are using the essential multivariate techniques in their analyses.

## *2.2. Scientific medical data bases*

Scientific medical data bases, structuring medical or surgical treatments and their follow-up, can be divided in quality control data bases and clinical research data bases.

*2.2.1. Quality control data bases.* The quality control data base should have a very easy data structure with a limited number of variables (absolute maximum 100). The variables should be very simple and easily definable, because quality control data bases only come to their real significance when they are used by as many centers as possible (e.g. The SMS data base [2], or the German Pilot Study Data Base) [3]. There should therefore be no individualized changes to the data structure and variables. The events studied should be very simple e.g. revision for bleeding, hospital stay and survival. A variable as intensive care stay is subjected to a lot of individual and local situations and is not a parameter defining a 'quality of medical treatment'. Survival is already a difficult variable, it can be presented as 30-day or hospital mortality but it should certainly not be analyzed on this basis (see data analysis). Death, as well as return of symptoms, should be analyzed as a time-related variable with the date of the event and the mode of the event, independent of the fact if the patient was still hospitalized or not. The variables studied should be selected from the incremental risk factors obtained in clinical research data bases and are known for most subdivisions of cardiology or cardiac surgery. Events should be structured in flat files. The results of these quality control data bases should not be available to the hospital administration (therefore outside funding) but maybe to professional (medical) associations or to a national bureau of medical evaluation as created in 1989 by the French Minister of Solidarity and Health Care Claude Evin. An essential element to make these data structures available to as many centers as possible, is their cost price and a few very well structured commercially available data base applications (e.g. SMS data base) are on the shelf.

*2.2.2. Clinical research data bases.* Clinical Research Data Bases have very sophisticated data structures with a considerable number of general variables without any known importance but also a very large number of variables centered around the purpose of the study. The data base should not be available to the hospital administration and needs therefore outside funding. Events are not structured in flat files but chronological files record the follow-up so that events can be restructured and redefined during the analysis.

The KULCSDB is an observational clinical research data base with the purpose of studying the influence of some operative variables on the occurrence of late events [4, 5, 6, 7, 8, 9] after isolated coronary surgery. The study group is very well defined as are the study variables.

### **3. Userfriendly and powerful**

The work on a data base has many different aspects, each performed by different people with different levels of education or different interests. The work of each has to be made as userfriendly as possible but power remains important.

#### *3.1. Data base hardware*

The original plan for the KULCSDB was that all work from data entry to data analysis should be performed on the same computer. This was found to be an error for a clinical research data base but is certainly perfect and maybe essential for a quality control data base. The cost of a relational data base package on mainframes and its adaptation to the clinical research requirements will be exorbitant, a secretary will have trouble learning the sophisticated codes for data entry, the clinician will have trouble creating reports without outside help, but the statistician doing sophisticated descriptive, comparative or actuarial runs will be delighted by the batch work capacities of the large mainframes. Since 1986 the execution of the work was divided between two computersystems not connected on-line: an Apple II with 5 megabyte random access memory, an 80 megabyte hard disk and streaming tape back-up, and a very large mainframe system consisting of several mainframes allowing even supercomputing (Universitair Rekencentrum KU Leuven). Table 1 lists the personal computer functions and Table 2 lists the mainframe functions.



*Table 1. The personal computer functions*

- 
- data base software creation and software management
  - data entry and data update
  - data consistency checking
  - missing data management
  - follow-up management (e.g. mail merge)
  - follow-up evaluation
  - data base reporting
  - data compression for data transfert
  - text processing and graphics
  - back-up
- 

*Table 2. The mainframe functions*

- 
- field labelling
  - data entry from compressed data transfert
  - outlier analysis and data reserach
  - standardized extensive descriptive and actuarial reporting
  - stepwise logistic regression
  - survival analysis using the method of Cox
  - survival analysis using the method of Blackstone
    - hazard function creation
    - parametric multivariate incremental risk factor analysis
    - parametric simulations
  - international data transfert
- 

### *3.2. Data base software*

The personal computer is equipped with a commercially available data base package (Omnis III plus, currently updating to Omnis V). A 100 page document was created by the author describing the required file and field structures and the screen presentations. For each field the consistency checking and the eventual calculation requirements were defined. The data base package was then adapted, based on these suggestions, by a professional expert. The personal computer is also equipped with a very large spreadsheet, a text processing package, a Webster dictionnary with added medical extensions, a statistical package, graphical display packages, a slide maker and backup software. The data base stays on-line on the personal computer.

The mainframes are using the SAS statistical package [10], initially upgraded with the experimental procedures of Blackstone but these procedures are currently included in SAS from version 5.18. The data base stays on-line on the mainframe but there is no continuous link between the personal computer and the mainframe.

#### 4. The personnel of the data base

The software is the core of the data base, similarly, the quality of the data is the core of the analysis. It is well known, that the more people are involved in a project, the less control there is and also the more risk of inter-observer variability. Since it was also the intention to build and maintain the data base without any outside influence, the data base funding was based on personal funds of the author. The cost-efficiency of the data base was therefore essential and the number of persons involved was reduced to its basic essentials, a data base manager, a data base secretary, a statistical expert and a software specialist.

##### 4.1. The data base manager

The data base manager is a cardiac surgeon, responsible for the coronary surgery program at the K.U. Leuven (1000 operations per year). Table 3 lists his responsibilities in the data base structure. He masters all aspects of the work for the data base. His weekly work, under normal circumstances, for the data base is around 20 hours. He works mostly on the personal computer level but also on the mainframe level for the parametric simulations.

*Table 3.* The responsibilities of the data base manager

- 
- the financial funding of the data base
  - creation of the file and field structure
  - definition of the consistency checking
  - development of new coding systems in the area of interest
  - the quality of the data
  - coding of all the variables from the patient files into a paper copy of the data base
  - the completing of the missing data
  - creation of reports from the data base
  - the management of the follow-up
  - the analysis of the results
  - the interpretation of the results
- 

##### 4.2. The data base secretary

Table 4 lists the responsibilities of the data base secretary. She is full-time employed for the data base. She was chosen after extensive testing for the high percentile score in speed and accuracy of numerical data entry. The ease of use of the Apple allows her to create herself new reports or to adapt existing reports, necessary in the secretarial work. She is only active on the personal computer level.

*Table 4. The responsibilities of the data base secretary*

- 
- management of the data flow
  - data entry from the paper code sheet into the computer
  - query and archiving of medical files
- 

#### *4.3. The data base statistician*

The data base statistician is a biostatistician, active in survival analysis. He has, as should be, been actively involved from the beginning in the creation of the data base, in every aspect of the analysis and in the building of the inferences. His responsibilities are listed in table 5, he is only active on the mainframe level. The statistical analysis, performed on the personal computer using the Stratview program, is done by the data base manager. Additional outside statistical support has been, and is still given at the University of Alabama by the statistical department of Prof. Dr. Blackstone. A mainframe copy of our data base (as of 1 January, 1988) is on-line available at their institution. The total worktime of the data base statistician has been around one year fulltime. The standardized descriptive and actuarial reporting systems are very powerful and easy systems, available on the mainframe to the data base manager. By just changing the search format (on line of software), a descriptive report is made of several hundred pages in a readable form with graphics included, similarly more than sixhundred different actuarial curves are created for five different events (total survival, cardiac survival, return of angina, reangio and reoperation) each time for a whole series of classifications (total defined group, by gender, by time frame of surgery, by internal mammary use, by anginal class, by NYHA class, by ventricular function, by surgeon...). For each of these classifications a statistical comparison is made between the curves using Wilcoxon and Logrank tests. Inferences from these analyses are limited by the univariate character.

*Table 5. The responsibilities of the data base statistician*

- 
- advice in the creation and coding of the fields on the personal computer
  - the transfert of the data to the mainframe into SAS
  - structuring and labelling the SAS fields and files
  - outlier analysis on SAS
  - structuring a standardized descriptive reporting system on SAS
  - structuring a standardized actuarial reporting system on SAS
  - survival analysis using Cox and Blackstone on SAS
  - structuring a standardized parametric simulation system
- 

#### *4.4. The data base software expert*

A data base software expert has been called in from an outside company. He has realized the fileformats, the screens and windows, as suggested by the data

base manager. He has only been active on the personal computer level. All software problems in SAS were solved by the data base statistician. A very sophisticated window was created showing in explicit views at data entry: the theoretical maximum follow-up duration, the real follow-up, the missing follow-up, the total mortality, the cardiac mortality, the return of angina in two different definitions..., each of these values is expressed in days.

## **5. The data**

### *5.1. The quality of the data*

Good quality of the data is the core of a good analysis. The utmost effort has been put into this problem. The dataflow was studied and standardized. The patient files were transformed into numerical values by only one person, the data base manager, in an effort to reduce the interobserver variability. The total number of records structured in the data base the 1st of January 1990 is 8200 records or patients. To reduce errors in data entry, a paper code sheet stores all the data before they are entered into the computer. This paper code sheet is an exact replica of the computerwindows. This increases the input speed and increases the accuracy. There is no coding directly from the patient file into the computer. This paper data base was run for five years to streamline dataflow and to guarantee the survival of the data base. The current paper version is the third version and all patients were first transferred to this final version before insertion into the computer.

Consistency checking is automatically performed at data entry and additional outlier analysis has been performed for several months and is done before each new analysis project.

New coding systems were developed for the preoperative and postoperative anginal and functional classes. Two hundred and forty variables define the operative report. This is divided in a structure of the operation and a description of the distal anastomoses, their size, quality, outflow area... The coding system used for the operative report is also used for the reangiograms. The total number of preoperative, operative and postoperative variables is more than 1000. They were grouped in some 400 fields and four files. For each group of variables free text room is available to comment specific situations.

### *5.2. The quality of the follow-up*

It is possible to structure the follow-up in a flat file. This would then be a variable day of death and a variable flagging the death, the same can then be done for a series of events after surgery. This is certainly the method of choice for a quality control data base, because of its apparent ease of use. For a clinical research data base it is essential to know the evolution of the changes over time and some comment about each follow-up entry under the form of free text. This

allows for recoding some events or creating combined events. In this last method the data base stays a living structure, changing every day. In the KULCSDB an unlimited number of follow-up entries can be entered for each record (patient). Each of these entries has 40 fields with several free text segments. Each entry represents data from one day in the life of a patient.

### *5.3. The methodology of follow-up*

The methodology of follow-up is also an essential element in an analysis. Most studies present actuarial curves without giving any information about the completeness of the follow-up. The KULCSDB method has a little bias as the common closing date method and gives the same completeness of information as the date-of-last-report method [11]. The data base is continuously updated through a regular, at least six-monthly, stream of follow-up reports from the referring cardiologists. In addition, after each event or suspicion of event, an additional report is sent by these specialists. Due to the historical importance of our institution in the treatment of cardiovascular pathology, and due to the relative short distances in the densely populated Belgian country (90% of our patients are living less than 100 km from the hospital), events and recatheterizations are often updated the same week. Ninety-five percent of all recatheterizations and reoperations of our patients were performed at our institution and most of the follow-up infarcts were treated here.

### *5.4. The completeness of follow-up*

A high number of patients untraced after hospital discharge, introduces a bias since untraced patients have a higher probability of having died than the traced patients [12]. In the last follow-up study ( $n = 6034$  patients) in November 1987, 99.6% of the patients were followed after discharge. Only two Belgian patients could not be traced. Both are known to be alive but they could not be contacted. Some of the missing foreign patients have been traced only after the closure date and were therefore not included in the analysis.

In the KULCSDB alle calculations of dates are performed using algorithms which generate intervals between Julian dates [13], the intervals are therefore obtained in days, The completeness of follow-up in the KUL Coronary Surgery Data Base is calculated from the real follow-up duration and from the missing follow-up duration.

The real follow-up is the interval between the day of surgery (time of study entry) and the day of the last information included in the study. The sum of the real follow-up periods studied in November 1987 was 10201236 days. The mean real follow-up of the censored survivors was  $1722 \pm 1156$  days.

The missing follow-up is the interval between the day of the last included information and the last day (10 January 1988) of the follow-up study (in the deceased patients the missing follow-up is 0 days). The completeness of follow-up, calculated as  $(\text{the real follow-up} / 100) / (\text{real follow-up} + \text{missing follow-up})$ , was 93.4%.

## 6. The analysis

### 6.1. Risk factor analysis

*6.1.1. The event is not time-related.* A postoperative event like 30-day or hospital mortality after coronary surgery can be studied comparing the distributions of discrete variables among clinical subgroups, using chi-square statistics, and the distributions of continuous variables, by one-way analysis of variance.

This usual presentation of the early perioperative risk has the limitation that it ignores events happening the 31st day or just after hospital discharge. These events could be caused by the same phenomena as the events in the first 30 days, like early graft occlusion or pulmonary embolism. It will analyze only the relation of one variable to an event without taking into consideration differences in distribution in other variables and more complex interactions between parameters.

Several variables are usually related to the occurrence of the event, they are therefore called incremental risk factors. However they need not be the cause of the event. For instance, left main disease can be a highly significant and relevant risk factor for, but it is not the cause of, a low output syndrome after surgery. The real cause could have been inadequate myocardial protection.

To evaluate the joint effect of several risk factors, appropriate statistical techniques are needed. The multivariate method of discriminant analysis determines the probability of having a certain disease or event based on explanatory factors. Several discrimination techniques exist, the KULCSDB has chosen the logistic methodology.

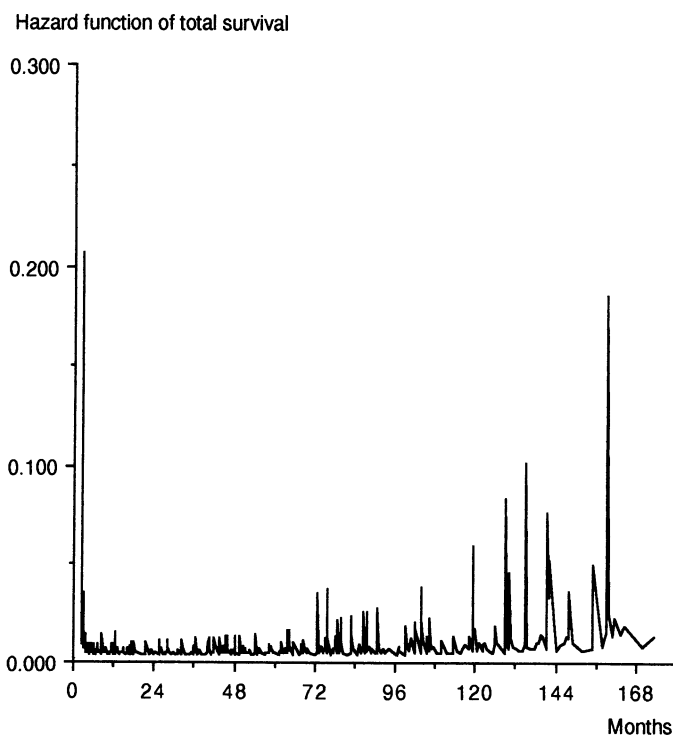
Different iterative steps of the computation include or exclude variables and alters their coefficients. When a variable is not retained in a model it means that it can be significant but not relevant, not significant and not relevant or that the information retained in this variable is already better presented in another variable.

The selection of variables presented for computation is a very delicate matter which requires extensive knowledge of statistics, of the data and of the pathology. The same parameter can and should sometimes be presented in a slightly different way. This process should be done for most of the important variables. It is called data research. One endarterectomy might be a risk factor, but if two were performed, the effect might not be additive. It could be that the parameter 'number of endarterectomies' is not selected but 'one or more endarterectomies' could well be included in a model. If cardiopulmonary resuscitation after surgery would be presented as a parameter for analysis of the early mortality after coronary surgery, it would probably turn out to be highly significant, but not relevant, and must therefore be omitted or analyzed with extreme precaution.

These discrimination methods combine the effect of several risk factors but are still not appropriate for the analysis of time-related events. In the

KULCSDB stepwise logistic regression is used in the analysis of non-time related events as mediastinitis after surgery.

*6.1.2. The event is time-related.* In the KULCSDB analysis of a time-related event after surgery a series of functions are created. The survivorship function expresses the probability that an individual survives at least to time  $t$ . The hazard function (Figure 1) expresses the fixed or changing instantaneous rate of prevalence of a time-related event across time for those patients surviving up to time  $t$ . The cumulative hazard function, expressed as the prevalence of the event over time, is the time integral of the hazard function.



*Fig. 1.* Hazard function of total mortality after coronary surgery in a population of 5880 consecutive patients.

In the simplest case of a constant hazard function, the risk of an event is not really related to time. The cumulative hazard function is then increasing linearly with time. The survival function is decreasing exponentially. The calculation of the constant hazard rate is very easy and can be estimated by the number of events divided by the sum of the total real follow-up or as events per patient year. But the calculated constant hazard rate is usually meaningless because the rate of events after surgery is usually not constant but changing in time. Therefore a presentation in events per patient-year at risk should never be used unless one has

verified that a constant hazard phase is adequate.

The most interesting events after coronary surgery are total mortality, cardiac mortality, the return of anginal symptomatology, reangiography and reoperation. The distribution of these events is usually not uniform over time, a time-varying hazard function must therefore be analyzed.

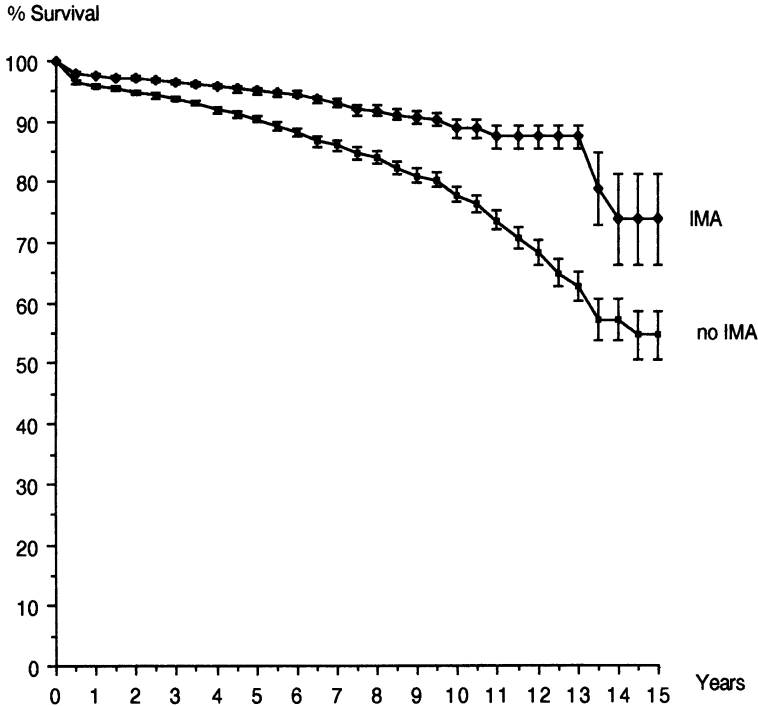
In a crude life table, each estimate of survival consists of the proportion of patients alive who have been followed from study entry to a specified point in time. For a ten-year survival estimate one has therefore to exclude all patients operated in the last ten years, the information about survival of these patients is not analyzed. Patients lost to follow-up during the studied interval create computational difficulties. An additional problem of this method is the fact that since the denominator is different in each calculation 5-year survival can be lower than 10-year survival due to the variability in small samples. This method does not use all available information and has therefore been abandoned by the KULCSDB for more accurate methods.

In the actuarial method the interval estimates of survival are dependent on the estimates of the previous interval so that the curve decreases in time, it also includes all available information. Assumptions are made relating to censored patients. Censored patients are patients not finishing the complete interval without having the event. Actuarial curves should be presented with numerical or graphical information of the standard deviation. The representativeness of the actuarial curve will increase with the decrease in interval duration. Intervals longer than one year dilute the time-relatedness of the event. In the KULCSDB the interval used is six months. The product-limit method by Kaplan and Meier [14] uses as interval the time period between consecutive events. The intervals are therefore unequal in length. The censored patients were simply not counted as at risk within the interval. The Kaplan-Meier method has therefore as many intervals as events.

Univariate analysis is done in the KULCSDB by constructing event-free curves (Figure 2) for different clinical subgroups and using a Wilcoxon or logrank test in the comparison of these curves. The Wilcoxon test compares the initial segments of the curve, whilst the logrank test is more specific for the whole curve. Comparing actuarial curves using these univariate methods will again only analyze the relation of one variable to an event without taking into consideration differences in distribution in other variables and possible interactions between parameters. It is therefore also very arbitrary and carries a similar bias but it is an essential step in the analysis of follow-up data.

Cox [15] proposed in 1972 a frequently employed regression model for estimating risk factors in survival or event analysis (the Cox Proportional Hazards Model). By this form of equation the risk factor causes a proportionate and not additive change in survival, and the underlying hazard functions structure is not specified. This method is therefore parametric with respect of the risk factors but non-parametric with respect of the hazard function structure. Risk factors have very often an effect on different time intervals in the follow-up. The anesthetist could have a positive or negative effect in the early phase





*Fig. 2.* Actuarial curves, using the method of fixed intervals, describing the survival in a population of 5880 patients from the KULCSDB. The population is divided in two groups, the patients with and those without an internal mammary artery graft. The logranktest P-value was  $<0001$  and the Wilcoxon test P-value was  $<0001$ .

after surgery but it is very difficult to understand how this effect could still be present and equally strong after fifteen years. Cox's regression method does not give the possibility of identifying relations between risk factors and phases, nor does it give different coefficients for the same risk factor for different phases.

These very critical limitations are avoided by using a method for analysing multiple phases of time-varying hazard and their phase-related incremental risk factors (Blackstone's methodology). This is the final method used in the KULCSDB in the study of time-related events.

Blackstone, Naftel and Turner [16] supposed that the complex distribution of events over time can be decomposed in a combination of a small number of phases, additive and overlapping. These phases should be medically meaningful. One can understand, in survival analysis of coronary bypass patients, that there is an initial period of high risk the first days which is then rapidly decreasing, but probably not 0 the 30th day as in the arbitrary method of a contingency table of hospital or 30-day mortality. In addition to this initial period one knows that all patients will die one day and that there will be a late rising phase of mortality. It is possible that the presence of a certain risk factor, for example diabetes, carries a constant risk throughout the follow-up period and can therefore be responsible for a constant phase.

The calculation and presentation of the cumulative hazard curve is an essential step in the parametric analysis. The number of phases in the hazard distribution over time can be deduced and some ideas can be formulated about the shape of these phases. Following Blackstone's method the shaping parameters can be guessed initially by the empirical cumulative hazard function before refinement by computation. Once the parametric estimates of cumulative hazard function and survival probability are computed it is very easy to present a similar estimate of the hazard function and its 70% confidence limit.

The number of phases, their scaling parameters as well as the number and magnitude of their shaping parameters are estimated statistically from the data using the maximum likelihood method.

### *6.2. Risk prediction for the individual*

Using the available equations, formulated by Blackstone, and adding in these the selected variables with their estimates and their standard deviation, it is now possible using the database to construct, for individuals or for groups of patients, time-related parametric curves with a degree of uncertainty predicting the occurrence of an event after surgery. The curves can be computed, holding the value of all other risk factors constant, with one parameter omitted (e.g. the use of a mammary artery graft) and with the parameter included (Figure 3). This will give, for this particular group or patient, a qualification and quantification of the effect of this parameter on the occurrence of the event. This can simulate randomized trials and it can show the effect of treating a particular patient with treatment A and treatment B, without actually submitting this particular patient to the risk of treatment A or B. Data bases become also a tool and not an entity in themselves. The American Heart-American College of Cardiology Task Force on Coronary Artery Disease has chosen the KU Leuven Coronary Surgery Data Base to obtain an insight in the late events after coronary artery surgery and to obtain equations that can be used in the comparisons with the treatment of percutaneous coronary angioplasty (PTCA). The data base can and has also been used in the analysis of centers with below average results. The multivariate predictions using the information from these other centers, but the equations developed from the KU Leuven Coronary Surgery Data Base, gave clear evaluations of the local results.

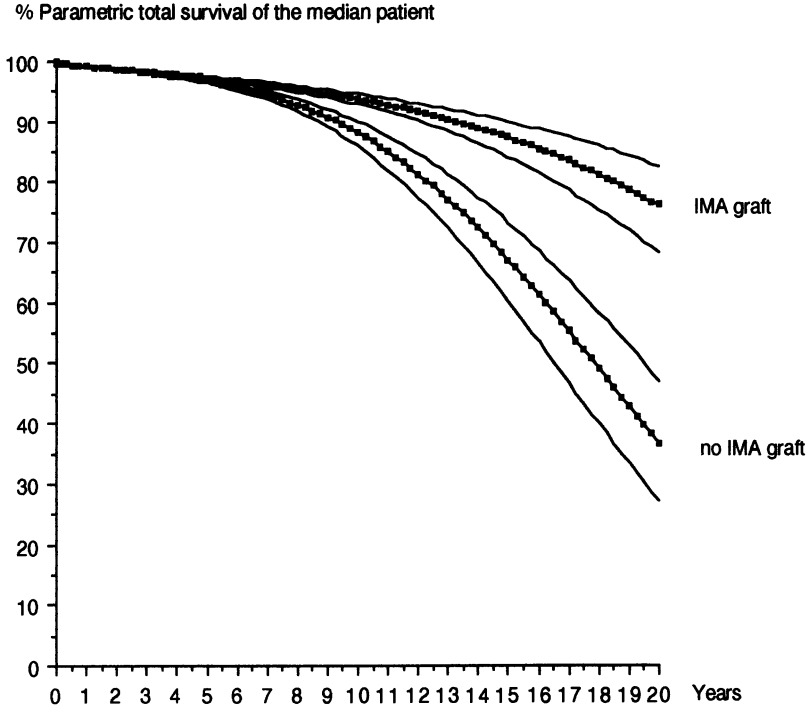


Fig. 3. Prediction of the parametric results of total survival for a hypothetical patient having the median values of the K.U.L.C.S.D.B. for his incremental risk factors, once calculated with and once without the use of an internal mammary artery graft.

**7. Conclusions**

A data base is a living structure, changing every day. To analyze it in a scientific way, we need to create a specific data set, make a correct follow-up study, analyze the follow-up study, analyze the results and then construct the inferences. The construction of a clinical research data base is a laborious project, where challenges are rising in the even smallest details due to the large number of fields or records involved. Once a data base has been realized and a methodology of analysis has been structured the benefits are considerable [17] sine it is the only way a message can be found in the experience of the past. The continuity in nature is an indication that this message should not be lost.

**References**

1. Sergeant P (1988) Hazard analysis in coronary artery bypass surgery. Ph D Thesis, Acco Leuven, Belgium ISBN 90-900-2572-3
2. Summit Medical Systems, One Carlson Parkway, Mineapolis, Minnesota 55447

3. Wilde E, Christof K, Struck E (1988) Pilot Study on Quality Assurance in Cardiac Surgery. *Thorac Cardiovasc Surgeon* 36: 174-179
4. Sergeant P, Lesaffre E, Flameng W, Suy R, (1985) Predictability of the postoperative workresumption after aortocoronary bypass surgery. Springer Verlag Editor: Walter 108-111
5. Sergeant P, Flameng W, Vanhaecke J, Suy R (1987) Time constraints in the emergency coronary bypass surgery for acute involving myocardial infarction. *Journal of Cardiovascular Surgery* 28: 68-74
6. Sergeant P, Lesaffre E, Flameng W, Suy R (1986) How predictable is the postoperative workresumption after aortocoronary bypass surgery. *Acta Cardiologica* XLI:41-52
7. Sergeant P, Wouters L, Dekeyser L, Flameng W, Suy R (1988) Is the outcome of coronary artery surgery predictable in patients with severe ventricular function impairment? *Journal of Cardiovascular Surgery* 27: 618-621
8. Sergeant P, Flameng W, Suy R (1988) The sequential mammary artery graft. *Journal of Cardiovascular Surgery* 29: 596-600
9. Sergeant P, Lesaffre E, Flameng W, Suy R (1990) The effect of the use and the method of use of the internal mammary artery on survival after coronary bypass surgery. *European Journal of Cardiothoracic Surgery* 4: 72-78
10. SAS Institute Inc., Box 800, Cary, NC 27511-8000, USA
11. Elveback L (1958) Estimation of survivorship in chronic disease: the 'actuarial' method. *J Am Stat Assoc* 53: 420-40
12. Austin MA, Berreyesa S, Elliott JL, Wallace RB, Barrett-Connor E, Criqui MH (1979) Methods for determining long-term survival in a population based study. *Am J Epidemiol* 110: 747-52
13. Fliegel HF, Van Flandern TC (1968) A machine algorithm for processing calendar data. *Communications of the ACM* 11: 657
14. Kaplan EL, Meier P (1958) Non parametric estimation from incomplete observations. *Am Stat Assoc J* 53: 457-81
15. Cox DR (1972) Regression models and life tables. *J R Stat Soc* 34: 187-220
16. Blackstone EH, Naftel DC, Turner ME Jr (1986) The decomposition of time-varying hazard into phases, each incorporating a separate stream of concomittant information. *J Am Stat Assoc* 81: 615-24
17. Kirklin JW, Naftel DC, Blackstone EH, Pohost G (1989) Summary of a consensus concerning death and ischemic events after coronary artery bypass grafting. *Circulation Suppl.* Vol. 798, 6, 81-91

# Clinical results with computer support of the decisions (in cardiosurgical intensive care unit)

V.A. LISHCHOUK

*Cybernetic Department, Bakulev Institute of Cardiovascular Surgery, Leninsky prosp., 8,  
117049, Moscow, USSR*

## Introduction

Mathematical and algorithmic methods of diagnosis and treatment have been discussed, developed and introduced for the last third of the century. An analysis of this scientific trend and of clinical practice leaves no doubt that they are developing as if parallel, sometimes without intersecting. In connection with this E.N. Shortliff noted that '... beyond a range of scientific investigations any system, in essence, even that one providing high-quality work hasn't been effectively used' [17].

Here I would like to discuss principal and technical decisions contributing to unification of clinical practice and up-to-date means of intellectual support of decisions. I will begin with short fragments illustrating the urgency of the goals set forth before illustrating computer support of a physician's decisions.

*Software.* Monitors connected to a patient provide a control of the data and alarm signals. Up-to-date monitors are functioning in a strikingly perfect mode (Figure 1). And still a physician chooses treatment using no means of intellectual support. How can we make computer systems help a physician in decision making? First of all a physician needs help not only in gaining and presenting information but in establishing the diagnosis and in choosing the treatment [3, 12, 13, 15, 17].

*Methodic support.* Usually a physician uses images and generalizations. For example, J. Kirklin [10, 16] proposed an algorithm, the essence of which lies in strictly defined rules of the type 'if during 3 postoperative hours the left atrial pressure is over 14 mmHG, but less than 18 mmHG, cardiac index is less than or equal to  $2\text{ l}/(\text{min m}^2)$  and mean arterial pressure is over 100 mmHG, then the diagnosis is cardiac failure and peripheral vascular spasm and in this case epinephrine, nitroprusside, etc. are recommended.' Here, all mentioned values are in essence statistical or intuitive generalizations of clinical experience.

Another approach is images and intuition. As an example we may present the clever diagrams composed by M. Braimbridge [2], in which a patient's image is



*Fig. 1.* The system supporting decision making installed at the Bakulev Institute of Cardiovascular Surgery.

created in case a patient has a low cardiac output syndrome caused by various situations. A diagnosis is based on a set of signs corresponding to this disorder. For example, at the advanced stage of left ventricular failure a patient has cyanosis, delirium, lung oedema, arrhythmia, elevated left atrial pressure, markedly lowered cardiac index, etc. In this case 'lowered' or 'elevated' is estimated in accordance with the gained experience.

In both cases it is difficult to take into account the patient's individual features. In neither case is fundamental biological knowledge taken into account. A careful analysis of such situation shows that in order to make methodical and technical means helpful in the choice of treatment it is necessary to unite clinical experience, current information, fundamental knowledge and the physician's skill into general methodology (technology).

## **1. Concept and principles**

Let's discuss the above-mentioned problem of necessity, constructivity and unification of fundamental knowledge, experience, current information and the physician's skill in detail.

### *1.1. Fundamental knowledge*

Out of a number of problems we must dwell on the system of regularities and on structure.

*1.1.1. Regularities.* An idea of the system of regularities can be illustrated by means of circulation. Below we will mention only those regularities and characteristics that have a marked effect (at least, higher than the total error of measuring) on the closed blood circulation in the human cardiovascular system. As examples we may refer to Starling's law, Hill's three-element scheme, used for myocardial tissue, Hook's law, used for cardiac fibrils, dynamics and statics of homeometric self-regulation, Laplace's law for a thick-walled or thin-walled sphere [1], Frank's 'windkessel' for cardiac cavities in diastole [9], Poiseuille's law for blood flow [8], cardiac rate/phase ratio (usually directly measured), valvular function, pressure, flow, annular displacement, stenosis and insufficiency ratio [14] as well as self-regulatory and pharmacological effects, which are, as a rule, additionally defined in relation to certain situations [5]. Within the frame of this presentation systems of cardiac and vascular regulations cannot be discussed in detail. If necessary, one can read other publications [11-14]. It must be emphasised that each of such sets must be chosen in the manner that really reflects the essence of the process and that meanwhile can be used to organize a complete and noncontradictory system. Our staff has developed the system of regularities in relation to the heart, circulation, respiration and their regulation.

*1.1.2. Structure.* Let's discuss another component of fundamental knowledge, that is, functional structure. We will discuss an example with circulation (see Figure 2). As a rule, a structure consists of elements and relationships. As elements we have the left heart, arterial system, microcirculatory bed, venous system, right heart, pulmonary artery, pulmonary capillaries, pulmonary veins and cardiovascular center. Relationships are shown with arrows. The arrows directed from element to element circumferentially represent blood flow. Radial arrows designate regulation (see Figure 2). Any of the mentioned laws corresponds to every element. In this way laws and structures are united. Detailed schemes must agree with the representation of all elements on the basis of a real time clinical and monitor control.

### *1.2. Experience*

Nowadays, experience is for convenience organized with the help of data bases. To do this we make a thorough analysis and load clinical material (from the publications). In addition, we have personal clinical experience organized as data bases. As a whole it is thoroughly evaluated by statistical and other methods and compared from year to year, from group to group, from patient to patient and from status to status. The general characteristics of our clinical

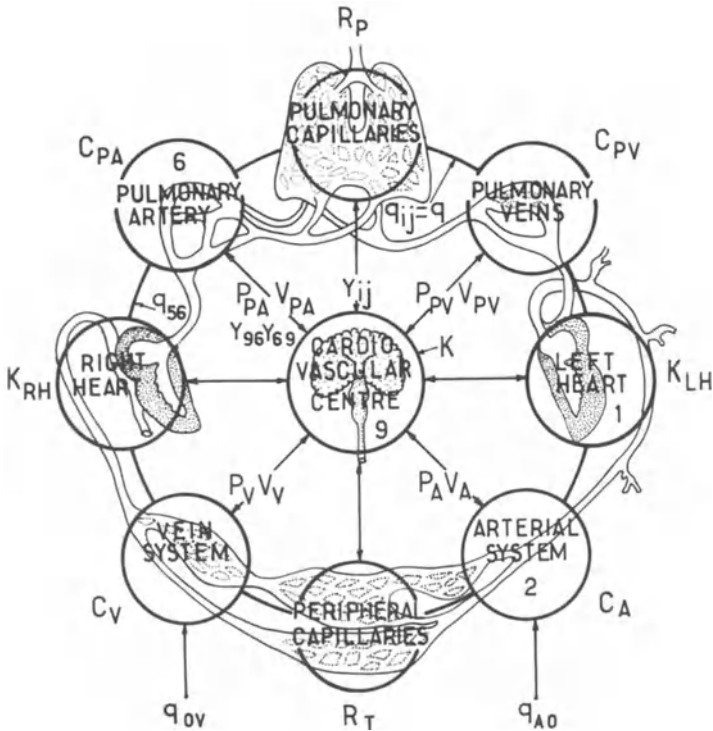


Fig. 2. A structure scheme of cardiovascular system.

Values: P – pressure; q – flow; V – volume; C – elasticity; Y – control; K – bulbar center activity;  $R_T$  – total peripheral resistance;  $R_p$  – pulmonary resistance;  $K_{LH}$ ,  $K_{RH}$  – pump coefficients of the left and right ventricles. Indices: A – arterial; V – venous; PA – pulmonary arterial; PV – pulmonary venous; i, j – number of reservoir; 0 – environment.

material, gained during evaluation and treatment and realized by mathematical models, are presented in Table 1.

The differentiation of evaluations of functions and properties is a matter of principle. As for circulation, to estimations of a function we may refer blood flow or cardiac index (CI; measured by RPG, thermodilution, Fik's method), arterial pressure (mean,  $P_A$ ; minimal, maximal, instant), central venous pressure ( $P_V$ ), pulmonary arterial pressure (mean,  $P_{PA}$ ; instant, systolic, diastolic), left atrial pressure ( $P_{LA}$ ; or pulmonary venous pressure,  $P_{PV}$ ), and control (which is estimated by means of test effect). Estimations, when necessary, are shown on a display at corresponding points of the scheme (see Figures 2-5). To the estimated properties we refer a pump coefficient of the left ( $K_{LH}$ ) and right ( $K_{RH}$ ) ventricle, elasticity of arterial ( $C_A$ ), venous ( $C_V$ ), pulmonary arterial ( $C_{PA}$ ) and pulmonary venous ( $C_{PV}$ ) reservoirs, total peripheral resistance ( $R_T$ ), pulmonary resistance ( $R_p$ ), circulating blood volume (V) and a regulation level (K).



Table 1. Type of pathology and number of patients examined by a monitor-computer system supporting decisions (1988).

	Type of pathology	Number of patients	Total
Acquired heart disease	Replacement mitral valve	228	534
	Replacement aortic valve	106	
	Multivalvular replacement	200	
	Radical correction of tetralogy Fallot	236	
Congenital heart disease	Closure of ventricular septal defect	111	509
	Closure of atrial septal defect	66	
	Fontain operation	40	
	Other operations	73	
	Coronary bypass grafting	171	
Ischemic heart disease	Aneurysm resection + coronary bypass grafting	28	226
	Aneurysm resection	27	35
	Other operations	35	
	Total number of cardiovascular patients		—

These estimates may be displayed on the screen. If a clinical control value is elevated, then a structural scheme may be detailed. For example, with a control of intraventricular pressure the following estimates for the heart may be defined: contractile ability (separately for the right and left ventricle), diastolic compliance, tonus (systolic and diastolic), heart rate and phase, etc. [11].

### 1.3. Mathematical models

A mathematical model provides organic unification of the discussed systems of regularities, structure and experience.

*1.3.1. General description.* Sequential realization of such an approach results in a general description presented in the concentrated parameters. For circulation such a description unites changes in volumes at every selected site ( $\dot{V}$ ) with their filling ( $\dot{V}$ ), rigidity (E), tonus ( $U^{-1}$ ), resistances (R), gravity (G), regulation (Y), tissue pressure (T) and blood loss (Q). By means of volumes, pressures and flows are estimated:

$$\begin{aligned} \dot{V} &= R^T \{E [V(t) - U] + T(t) + G(t)\} + Q_0(T), \\ P(t) &= E [V(t) - U] + T(t), \\ Q(t) &= D(t) \cdot R - R^T \cdot D(t), \\ F^* [E(y), U(y), R(y)] &\xrightarrow{E(y), U(y), R(y)} \min \end{aligned}$$

V, P, U, T, G, Q<sub>0</sub> – n-estimated vector-columns; Q, D, R and E – matrixes n \* n, D = diag [D<sub>11</sub>, ... , D<sub>nn</sub>], E = diag [E<sub>11</sub>, ... , E<sub>nn</sub>], F\* – a target function.

*1.3.2. Concretization.* Models used for concrete attachments are derived from a general description at the given member of reservoirs (n), interrelations (r), as well as of specificity. Specificity loading is not a formal task and often requires continuous thorough studies just up to experiments.

In case of orientation to the present-day clinical control a four-reservoir model of proper circulation (without regulation) and a 14-reservoir model of circulation with the beating heart and regulation for the sake of simulation are extremely useful for identification. Detailed descriptions of clinically oriented models of the heart, circulation and respiration are given in other publications [5, 14].

*1.3.3. Quantitative estimates.* Initial conditions and coefficient values of the models are defined by clinical experience. Examples of circulation estimates done during uncomplicated periods after open-heart operations and united by the model are shown in Tables 2 and 3.

*Table 2.* A summary table of the estimates of circulatory functions united by the model seen in patients undergoing open-heart operations and having an uncomplicated early postoperative period.

Operations	CI min <sup>-1</sup> m <sup>2</sup>	P <sub>A</sub> mmHg	P <sub>PA</sub> mmHg	P <sub>V</sub> mmHg	P <sub>LA</sub> mmHg	HR min <sup>-1</sup>
Mistral valve replacement	3.13	89	26	11	14	102
Aortic valve replacement	3.37	93	19	10	14	96
Mitral and aortic valve replacement	3.06	94	26	9	13	93
Surgical correction of Fallot tetralogy	3.61	75	24	10	10	113
Surgical correction of ventricular septal defect	3.62	76	22	4	120	
Coronary bypass grafting	3.45	75	18	10	11	83
Normal state	3.00	92	17	5	9	70

Table 3. A summary table of the estimates of circulatory properties seen in the patients undergoing open-heart operations during early smooth postoperative period.

Operations	$\frac{K_{LH}}{mL}$ s m <sup>2</sup> mmHg	$\frac{K_{RH}}{mL}$ s m <sup>2</sup> mmHg	$\frac{R_T}{dyn\ m^2c}$ mL <sup>s</sup>	$\frac{R_P}{dyn\ m^2c}$ mL <sup>s</sup>	$\frac{C_A}{mL}$ mmHg m <sup>2</sup>	$\frac{C_V}{mL}$ mmHg m <sup>2</sup>	$\frac{C_{PA}}{mL}$ mmHg m <sup>2</sup>	$\frac{C_{PV}}{mL}$ mmHg m <sup>2</sup>	V mL	S m <sup>2</sup>
Mitral valve replacement	3.70	4.97	2000	293	0.71	56.0	1.55	8.61	1400	1.67
Aortic valve replacement	4.12	5.65	1971	177	0.86	55.4	1.78	8.01	1318	1.69
Mitral and aortic valve replacement	3.92	5.44	2221	348	0.76	60.8	1.08	8.82	1311	1.67
Correction of Fallot tetralogy	6.04	5.89	1439	333	0.73	53.6	1.26	7.77	782	1.14
Correction of ventricular septal defect	6.39	7.07	1479	269	0.76	66.4	9.62	0.63	473	
Coronary bypass grafting	5.33	6.00	1518	173	0.98	58.3	2.06	8.43	1431	1.88
Normal state	5.90	6.00	2266	227	0.88	118.0	1.76	17.70	1500	1.70

#### 1.4. Individualization

The described models are based on the laws, that is, on general, repeated and stable interrelationships. Their quantitative estimates are taken from averaged experience with a monitor control (see Tables 2, 3); therefore they do not reflect individuality.

To take into account individuality, the computer used as a monitor control selects, within dialogue with a physician, model properties in a mode eliminating the difference between functions and indices. To do this we use various methods, for example, Kalman's filters and functions of sensitivity of the second kind [14]. At present we believe that those methods which substantially take into account specific features of the object are most effective. In accordance with it a procedure of individualization consists of preliminary adjustment of a model to the object, that is, adjustment of a model structure, choice of initial values of functions and coefficients, choice of limitations, their clear and simple image on the display obtained during studies; only then parameter (properties) are measured according to the measurable indices. Thus, we must take into account general, repeated and substantially individualized relationships.

#### 1.5. Physician's skill

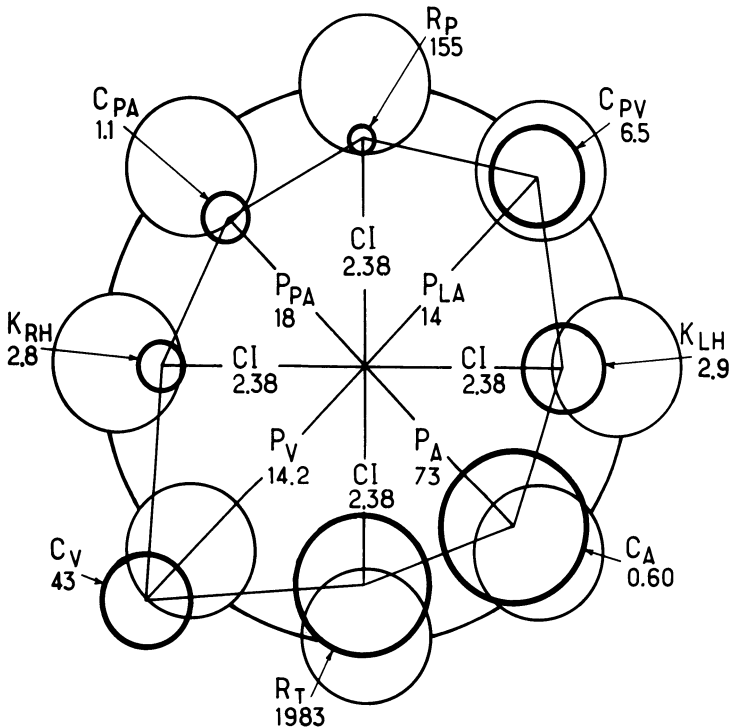
*1.5.1. Dialogue.* No matter how good the model is (it cannot be said about the described ones; they must be considered not as an adequate analogue but as a constructive instrument) it covers only a part of a situation on which a physician must make decisions. The models poorly reflect a medium state, the general condition of a person making a decision, economic possibilities, ethical prohibitions, etc. Therefore, models must be used within a dialogue regimen. A dialogue must be constructed in such a way that formal methods are effectively united in relation to the experience and skill of the persons making decisions. Within the systems of decision support organized by means of a dialogue the tasks of curve interpretation, determination of errors in measurements and calculations, determination of physiologically valuable thresholds of changes in properties, choice and correction of the due and normal values of quality criteria, explanation of value of pathological shift, etc. are solved.

*1.5.2. Quality criteria.* Propounding and realization of the criteria of an optimal control is a common and natural wish. In practice, it is very difficult to find out quality criteria and set adequate limitations before medical effects are known. Definition of criteria is often related to expert estimates and systems. Evidently, such an approach may improve the general level of correct decisions, but, as a rule, does not result in basically new horizons, since it is limited by an existing level of knowledge and skill.

*1.5.3. A weak link.* While solving this problem our staff has developed a concept of a weak link. A weak (or basal) link is the property highly contributing to a

dominating harmful (or useful) process. Under certain conditions this property is such that it defines the working ability of the system. A method of determination of a weak link is based on a theory of sensitivity. A weak link is a natural criterion of governing, in essence the opposite of a quality criterion, but just as effective [6, 13, 14].

*1.5.4. Substantial (explaining) analysis.* We use simultaneous-image representation of the cardiovascular system and the mathematical model. For this purpose we add graphic representation and numerical estimates of the patient treated in the intensive care unit to this scheme (Figure 3). A scheme of the patient with a favourable effect is shown by a thin line. The distance between elements shows function estimates (see Figure 3). The radius of circles shows property estimates. Relative estimates of functions and properties are shown on the display.



*Fig. 3.* A structure scheme of circulation of the patient subjected to treatment represented against a structure scheme of circulations of the averaged patient who is doing well.

Now let us discuss values of changes in the formation of the pathological process. Pulmonary resistance is markedly changed (by 1.9 times). However, the computer, which provides an imitation study on the individualized model, determines that the changed cardiovascular system is highly influenced by the lowered pump ability of the right ventricle.

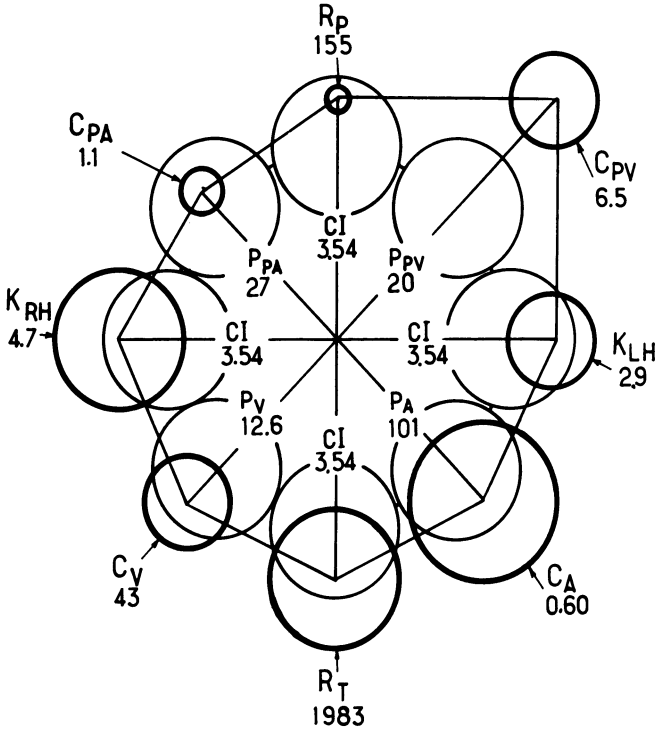


Fig. 4. A scheme of circulation reflecting an individualized model in a patient with a normalized right heart.

To evaluate the role of the right heart in pathological changes, we put the results of the imitation study on the display as a scheme showing on the individualized model of the patient, with normalized right ventricle (see Figure 4). It is seen that cardiac index and left atrial pressure are elevated 1.5 times. Here, left atrial pressure exceeds the normal one. Arterial pressure is elevated by one third, cardiac load – even higher. Therefore the causes of pathological changes are probably not in the right ventricle. Probably, the lowered pump ability is of a regulatory, overload-preventing character. This version may be tested on the model. A physician-computer dialogue may present several variants of such studies. How can we choose the next step of such an analysis? A corrected right ventricle normalizes pulmonary vein pressure. The latter is most sensitive to changes in a heart's status. Therefore, let's try to correct a left ventricular state. In Figure 5 you can see a result of simultaneously normalized right and left ventricles. There is general hyperfunction. The latter is highly sensitive to the status of the right heart, blood volume and vein elasticity. Vein elasticity is lowered in this patient. It probably has a compensatory character and prevents reduction of the cardiac index. When the right heart is normalized, such a compensatory shift becomes senseless and results in hyperfunction. In this case to cause a proper pathological state we must normalize the right heart and venous tonus.

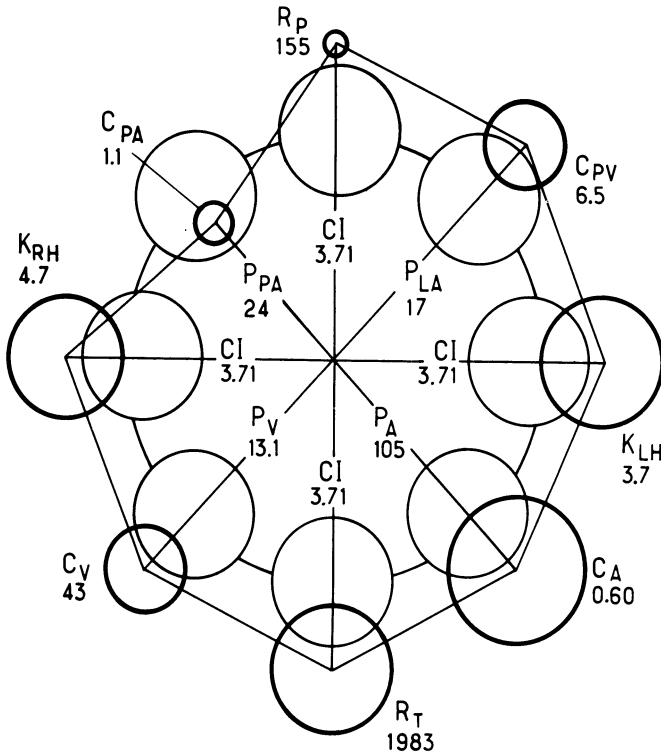


Fig. 5. A scheme of circulation reflecting an individualized model in a patient with normalized left and right hearts.

Thus, a model allows us to attain understanding of a course of pathological process and enables us to separate pathological shifts into pathological and purposeful. Here our experience showed that at one and the same pathological change the treatment is different if there are different compensatory processes. Hence, it follows that estimation of compensation must be an integral part of diagnosis and must be included in formulation and content of treatment.

## 2. Realization and use

*Technical realization.* The described concept has been realized (since 1974) by means of the monitors produced by the companies Hewlett-Packard, Hellige and Siemens and by means of mini- and personal computers (Figure 1).

*The program is realized* in FORTRAN, ALGOL, BASIC and C languages. Several parts of the program are written in the ASSEMBLER.

*Technology of use.* A scheme of the system providing reanimatologist's decisions is shown in Figure 6. Data of the patient's status are loaded into

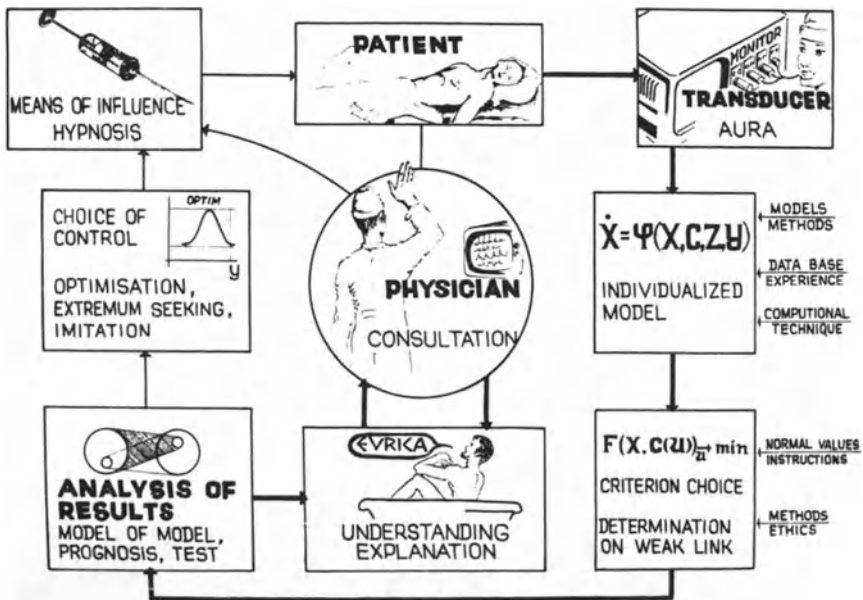


Fig. 6. A structural scheme of the computer-monitor system providing reanimatologist's decisions.

computer memory for 10 seconds every 2-10 minutes (or on the physician's demand), approximately 50-150 times during the night. All results are represented on ward displays. If a physician makes no corrections then a process of measurements, treatment, remembering and governing the data is automatic. From 1975 to 1989 over 1500 patients were treated according to this scheme. During the last years an analogous system has operated in the operating room and during stress veloergometric tests.

### 3. Results

For many years (since 1974) practice of algorithmic support of decisions in the postoperative department of the Bakulev Institute of Cardiovascular Surgery has contributed to an early diagnosis of disorders, to a lowered rate of complications and mortality relative to similar populations of patients not connected to the automatized system (more than twice in certain groups), to more economical drug administration and to a shorter stay in the intensive care unit [7]. What is the cause of such results? One of the main causes is haemodynamic unloading of the heart controlled by a weakest-link method.

#### 3.1. Heart unloading

In the late sixties and early seventies vasodilators were rarely administered. And



when used, they gave unstable results. Consequently, the motto 'don't do any harm' restrained optimism. In those years a method of postoperative treatment reflected a natural tendency towards normalization of the body's functional systems and of circulation in particular. A widely used tactic offered support of cardiac output within the normal physiological limits, or exceeding them. This tactic satisfied two other commonly used recommendations: (a) support of a high volume of circulating blood; (b) saturating glycoside administration.

Studies done on the automatized system have convinced us that treatment measures based on principles of circulation normalization (catecholamine or derivative administration, hypervolemia, saturating glycoside administration, etc.) have not always resulted in positive results [4]. Haemodynamics improved by mesaton, alupent, blood transfusion and blood substitutes was of temporary character, since it was achieved at the expense of overload of the operated heart.

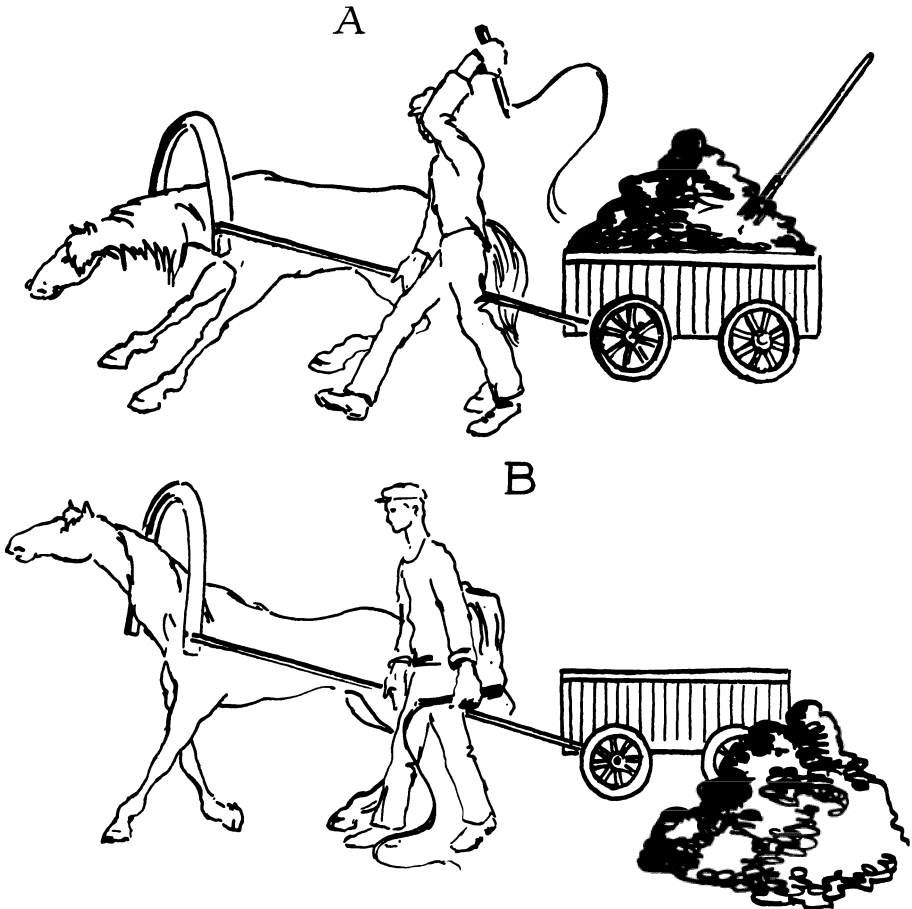
Overload of a heart already weakened by surgical intervention results in respectively decompensated circulation. It may be complicated by acute cardiac failure of functional genesis. To prevent this common complication and to provide conditions favourable for the surgically weakened myocardium, it is necessary to organize postoperative treatment in such a way as to minimize the load on the surgically traumatized heart. It is hardly feasible without a mathematical model since unloading may be easily transformed into haemodynamic shock.

By the present time much experience has been gained at the Bakulev Institute of Cardiovascular Surgery in the realization of a treatment based on maximal sparing of the heart. Such studies revealed the possibility of left and right ventricular unloading by an average of 50% with confident and effective improvement of a functional status of the myocardium after (and as a result of) unloading. Simultaneously, we revealed that for a patient's organism (for the heart in particular) the way in which reduced cardiac efficiency is achieved matters in every individual case. If there is no individual correction then reduced cardiac efficiency may produce circulation centralization, lowered myocardial contractility, etc.

Thus, though we learned much in the last decade the problem of serious complications remains urgent. Such a situation is shown in Figure 7. But the most difficult task is choosing the cardiogenic/vasodilator ratio in complex disorders of cardiac and vascular genesis.

### *3.2. Estimation of treatment efficiency*

Treatment with cardiogenics may elevate the cardiac index and at the same time may have a deleterious effect. Let's discuss such a situation, which may be seen during the treatment of low output syndrome with dopamine. Indices in patient K, with acute cardiac failure after valve replacement, seen in response to intravenous dopamine  $8 \mu\text{g}/(\text{kg min})$  are shown in Figure 8. In Table 4 it is shown that the cardiac index was elevated by 28%, arterial pressure by 50% and the left and right ventricular loads were nearly doubled. Having examined these



*Fig. 7. (A) You may drive on (cardiotonics). (B) You may reduce a load (vasodilators). Only a model of an individual status helps to find an adequate decision.*

properties we find that the functional status of the left ventricle is reduced by 18%. With the help of such a model we can find the causes of these changes (see Table 5): elevation of the cardiac index and pressure is caused not by improved cardiac activity but by its loading which, in turn, results in reduced venous elasticity and elevated systemic resistance.

This picture illustrates the possibilities of mathematical models in the definition of the value of the effect (but not of volume of a response or close relationship) of pharmacological treatment.

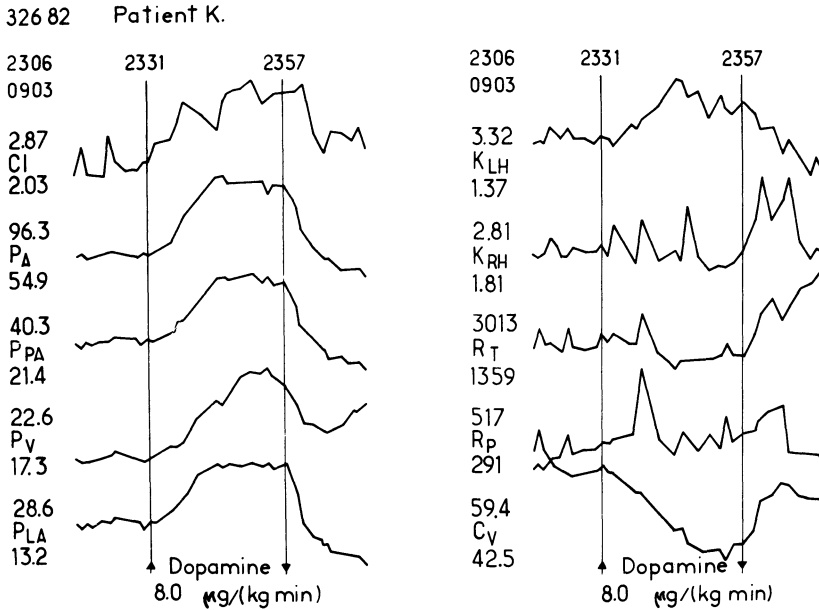


Fig. 8. Circulatory indices of the patient with acute circulatory insufficiency following valve replacement in response to intravenous dopamine  $8.0 \mu\text{g}/\text{kg min}$ .

Table 4. Effect of dopamine on cardiovascular functions and properties.  
(Patient K;  $8.0 \mu\text{g}/(\text{kg min})$ ; 9.03.82; 11.06-11.39 pm)

functions	Dopamine		Deviation %	Properties	Dopamine		Deviation %
	$0 \mu\text{g}/(\text{kg min})$	$8 \mu\text{g}/(\text{kg min})$			$0 \mu\text{g}/(\text{kg min})$	$8 \mu\text{g}/(\text{kg min})$	
CI	2.10	2.69	28	R <sub>T</sub>	1800	2250	25
P <sub>A</sub>	64.8	97.6	50	R <sub>P</sub>	342	332	3*
P <sub>PA</sub>	27.4	39.0	43	C <sub>A</sub>	0.67	0.50	-34
P <sub>LA</sub>	18.4	27.8	52	C <sub>PA</sub>	1.41	0.78	-82
P <sub>V</sub>	17.4	21.3	22	C <sub>V</sub>	33.9	26.5	-31
N <sub>RV</sub>	0.15	0.30	100	C <sub>PV</sub>	4.90	4.29	-31
N <sub>LV</sub>	0.35	0.67	94	K <sub>R</sub>	2.01	2.11	5**
				K <sub>L</sub>	1.91	1.61	-18

N<sub>RV</sub> N<sub>LV</sub> – capacities of right and left ventricles.

\* Sign (-) shows value of magnitude of the final state to be less than the initial one.

\*\* The difference between values is not reliable ( $P > 0.05$ ).

Table 5. Changes in haemodynamic function relative to changes in estimates of properties in response to dopamine administration ( a model analysis).  
(Patient K; 8,0  $\mu\text{g}/(\text{kg min})$  ; 9.03.82; 11.06-11.39 pm)

Properties	Functions				%
	CI	P <sub>A</sub>	P <sub>PA</sub>	P <sub>V</sub>	P <sub>LA</sub>
R <sub>T</sub>		20			
K <sub>LH</sub>	- 5		9		14
C <sub>V</sub>	24	24	22	23	23
All properties	25	41	37	20	43

### 3.3. Balance of circulating blood volume

Different clinical schools stick to various criteria of support or change of blood volume in the patients who underwent open-heart surgery. Our experience shows that both tactics of 'a drop for a drop' and hyper- or hypovolemia may have positive or negative sequelae in relation to an individual patient's status [14]. Normalization of circulating blood volume cannot be viewed as an end in itself. In cases of acute cardiac failure transfusion of blood or its substitutes or stimulation of diuresis are necessary not because of the fact that blood volume per kg of body weight is below or above the normal values, but because of the fact that a criterion 'improved or depressed circulating blood volume' points to a cardiac functional status (but not a function) (or any other organ, if it results in the development of a pathological process).

### 3.4. Classification of circulatory distresses

Fundamental physiological knowledge united into the systems by mathematical models of the heart and circulation on one hand and the gained clinical experience on the other have contributed to the development of the classification of acute circulatory and cardiac disorders, which is algorithmically represented [6]. Classification generalizes basic theses of the other ones. The development has been realized in the following way. By means of the model every type of disorder is matched with that property (for example, contractility, elasticity, tonus) which results in pathological manifestation of a function (for example, changes in blood flow, pressure, volume, etc.). A quantitative estimate of the property that has the greatest influence on the development of a pathological process is used as a criterion of classification. A diagnostic method is based on a principle of the weakest link. A sequential analysis of the influence of all properties influencing a circulatory function has allowed differentiation between pathological and compensatory shifts. Differentiation between pathological and compensatory shifts has the following basis: a compensatory

response to changes in cardiac and vascular properties reacts in such a way as to restore deficit of the damaged function; protective responses prevent refusal, reduce a load (even to the detriment of a function) on the most strained (or weakened) subsystem (organ). This method is realized in computer program form (IBM PC, APPLE, etc.). Use of this classification immediately reveals pathogenic links and hence provides a therapeutic decision. Real time examination of the latter confirms their adequacy in a great number of decisions [7]. We have undertaken a comprehensive analysis of diagnostic results in some clinical precedents with the help of the most common degrees of acute circulatory distress seen in cardiology. A comprehensive analysis showed that different diagnostic methods do not always reveal the same disorders, and more than that, we may say, it happens rather often (see Table 6).

On the whole, these studies have shown that the developed classification solves a problem of completeness (definiteness), necessity (sufficiency) and noncontradiction (consistency) on the basis of a system of independent signs (properties). It allows the evaluation of associated distresses and provides details agreement with joint body subsystems and quantitative evaluation of severity.

*Table 6.* The coincidence of types of disorders revealed by different methods. The indices 1, 2, 3, 4, 5, 6, 7 designate left ventricular failure, right ventricular failure, cardiac failure, hypovolemia, peripheral spasm, concomitant peripheral spasm and favourable state, respectively.

Classification	J.W. Kirklin	M.V. Braimbrige	R.N. Lebedeva	CMC*
J.W. Kirklin	12 <sub>1</sub> — — 5 <sub>4</sub> 10 <sub>5</sub> 3 <sub>6</sub> 10 <sub>7</sub>	3 <sub>1</sub> — — — 3 <sub>4</sub> 4 <sub>5</sub> 2 <sub>6</sub> 4 <sub>7</sub>	— — — 3 <sub>4</sub> 5 <sub>5</sub> 3 <sub>6</sub> 7 <sub>7</sub>	5 <sub>1</sub> — — — 6 <sub>5</sub> 2 <sub>6</sub> 1 <sub>7</sub>
M.V. Braimbrige	3 <sub>1</sub> — — 3 <sub>4</sub> 4 <sub>5</sub> 2 <sub>6</sub> 4 <sub>7</sub>	9 <sub>1</sub> 2 <sub>2</sub> 11 <sub>3</sub> 3 <sub>4</sub> 6 <sub>5</sub> 7 <sub>6</sub> 5 <sub>7</sub>	2 <sub>1</sub> — 7 <sub>3</sub> 3 <sub>4</sub> 4 <sub>5</sub> 5 <sub>6</sub> 5 <sub>7</sub>	4 <sub>1</sub> 1 <sub>2</sub> 6 <sub>3</sub> — 4 <sub>5</sub> 2 <sub>6</sub> 1 <sub>7</sub>
R.N. Lebedeva	— — — 3 <sub>4</sub> 6 <sub>5</sub> 3 <sub>6</sub> 7 <sub>7</sub>	2 <sub>1</sub> — 7 <sub>3</sub> 3 <sub>4</sub> 4 <sub>5</sub> 5 <sub>6</sub> 5 <sub>7</sub>	2 <sub>1</sub> — 17 <sub>3</sub> 4 <sub>4</sub> 6 <sub>5</sub> 6 <sub>6</sub> 8 <sub>7</sub>	— — 4 <sub>3</sub> — 4 <sub>5</sub> 2 <sub>6</sub> 1 <sub>7</sub>
CMC*	5 <sub>1</sub> — — — 6 <sub>5</sub> 2 <sub>6</sub> 1 <sub>7</sub>	4 <sub>1</sub> 1 <sub>2</sub> 6 <sub>3</sub> — 4 <sub>5</sub> 2 <sub>6</sub> 1 <sub>7</sub>	— — 4 <sub>3</sub> — 4 <sub>5</sub> 2 <sub>6</sub> 1 <sub>7</sub>	9 <sub>1</sub> 1 <sub>2</sub> 8 <sub>3</sub> — 9 <sub>5</sub> 10 <sub>6</sub> 1 <sub>7</sub>

\* Clinico-Mathematical Classification (according to V.I. Burakovsky, D.Sh. Gazizova).

### 3.5. Necessity of individual treatment

The form (but not the quantitative relationship) of a greater part of correlations is stable. As examples we may refer to Starling's, Pousseulet's and Frank's laws, specific features of the valves, homeometric self-regulation and several others. However, a part of a relationship varies greatly in relation to a patient's status. For example, the correlation coefficients  $r$  (CI,  $C_V$ ), estimated according to the data of many-hour monitoring in 20 patients who underwent open-heart operations, vary from 0.33 to 0.61. Correlation of the cardiac index and pump

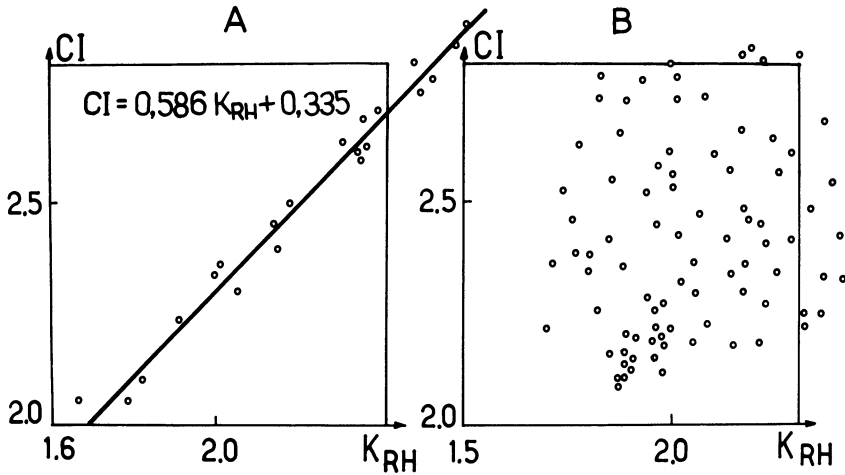


Fig. 9. Relationships of cardiac index and pump coefficient of the right ventricle (A), in the same patient registered in 20-minute intervals (B).

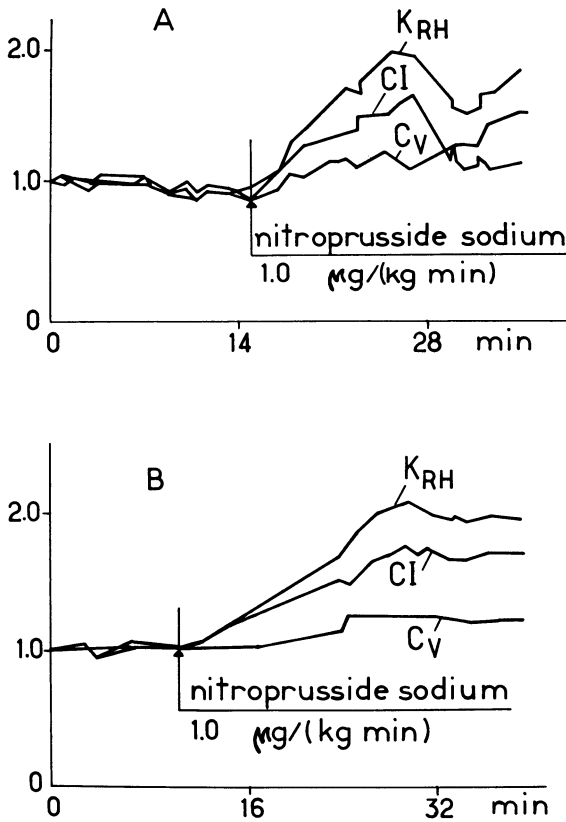


Fig. 10. Time changes in cardiac index, pump coefficient of the right heart and venous elasticity (A, in relative values) influenced by nitroprusside sodium, 1.0 µg/(kg min), and the same graph 20 minutes later (B, see explanations in the text).

coefficient of the right ventricle may be changed in a 20-minute period from rather high to very low (see Figure 9). Two following graphs reveal the causes of these changes (Figure 10). If  $C_V$  remains practically unchanged in response to nipride administration and  $K_{LH}$  is increasing (in response to intravenous nipride), then the cardiac index is elevated. If  $C_V$  is also elevated then the cardiac index is insignificantly changed and the relationship between CI and  $K_{RH}$  weakens. Probably in severe complications these combinations are numerous and have combinatorial character. This results in essentially individual (like in chess playing) responses and hence in the necessity of treating the patient, but not the disease.

## Conclusion

The urgency of an individual approach based on scientific methods (and skill in particular) is principal. We speak about the fact that a specialist in intensive care or a reanimatologist must undertake valuable scientific studies in an extreme situation and the obtained results must be used to choose or correct treatment. This organic combination of clinical practice and scientific study is the next age of medicine.

## References

1. Beneken JEW (1965) A mathematical approach to cardiovascular function. The uncontrolled human system. Institute of Medical Physics Report. Utrecht 194 p
2. Braimbridge MV (1981) Postoperative cardiac intensive care. 3-3 ed Oxford, 233 p
3. Brower RW, ten Katen HJ, Meester GT (1984) Problems and pitfalls in a clinical research data management system. *Comput Programs Biomed* 19(1): 13-30
4. Burakovsky VI, Lishchouk VA (1977) An analysis of myocardial hemodynamic overload during open-heart operations. *Grudnaya Khirurgiya*, 12-23
5. Burakovsky VI, Lishchouk VA (1985) Results with an individual diagnosis and treatment of the patients with acute circulatory disorders (based on mathematical models). Preprint 85-29. Kiev, 53 p
6. Burakovsky VI, Lishchouk VA, Gazizova DSh (1983) Classification and diagnosis of acute circulatory disorders by means of mathematical models. Preprint 83-47. Kiev, UkSSR Academy of sciences. Institute of cybernetics. 28 p
7. Burakovsky VI, Lishchouk VA, Kertsman VP, Mironenko VI (1989) Acute circulatory disorders and heart failure after open heart surgery. *Vestn Acad Med Nauk SSSR*. 10: 3-14
8. Evans RL (1962) A unifying approach to blood flow theory. *J Theor Biol* 3: No. 3: 392-411
9. Frank O (1899) Die Grundform des arteriellen Pulses. *Zeitschr Biol Bd*. 37: 483-526
10. Kirklin JW (1970) System analysis in surgical patients. Birmengam: Glasgow University Publication. 23 p
11. Listschuk WA (1970) Die Selbstregulierung des Herzens. *Der Mensch als Regler*. Berlin 43-67
12. Lishchouk VA (1973) Automatized system in investigative programs and professional training. USSR Ministry of Health, Central Institute for Advanced Training. 34 p. Deposited in the All Union Research Institute of medicine and medical technical information. N 169-173
13. Lishchouk VA (1978) Experience in application of mathematical models in the treatment of patients following surgical operations on the heart. *Vestn Akad Med S*. No. 8: 38-49

14. Lishchouk VA (1981) A formalized theory of circulation oriented to a cardiosurgical clinic: Diss Doct Sci. Moscow 454 p
15. Pincioli F, Turri E (1987) A first step quantification by correlation techniques of the variability in some time intervals measured on the Politecnico/Ca- Granda VCG Database On Young Normal Subjects. *Comput Biol Med* 17(2): 127-141
16. Sheppard LC, Kirklin JW (1974) Cardiac surgical intensive care computer system. *Fed Proc* 33: 2326-2328
17. Shortliffe EN, Buchanan BG, Feigenbaum EA (1979) Knowledge Engineering for Medical Decision Making: A Review of Computer Based Clinical Decision Aids. *Stanford Univ Stanford*, 1207-1224



# A pediatric cardiology diagnostic coding system and database

R. W. BROWER

*Medical Computing Consultants, Rotterdam, The Netherlands*

E. HARINCK

*Dept. of Cardiology, Wilhelmina Children's Hospital, Utrecht, The Netherlands*

A. C. GITTENBERGER-DE GROOT

*Dept. of Anatomy and Embryology, University of Leiden, Leiden, The Netherlands*

## **Abstract**

The design and structure of a pediatric cardiology and cardiac surgery coding system is described. This system supports a practical diagnostic & intervention coding system as well as the recording of essential patient administrative and medical information. The aim of the program is to improve patient care and research, stimulate study projects between cooperating centers, and enhance the efficient management of departmental resources. The coding system contains over 1700 specific items covering virtually every anatomical, surgical, and cardiological related diagnosis and intervention in pediatric cardiology. Many of the diagnoses are based on the (Brompton) segmental analysis approach with direct access to each separate item by code number or key words, or by a stepwise approach through extensive diagnostic 'trees'. The database system includes information on resource utilisation, an integrated growth chart program, follow-up diary and mailing system as well as several standard retrieval and reporting programs. The system is now operational in all the Pediatric Cardiac Centers in the Netherlands as well as a number of major hospitals in the European Community.

## **Introduction**

Pediatric cardiology differs in many respects from the practice of adult cardiology, the principal difference being the nature of the disease confronting the pediatric cardiologist. This can be described as continually dealing in a wide spectrum of rarities, usually congenital rather than acquired. This contrasts with the 'production' work of adult cardiology often consisting of coronary artery disease, myocardial infarction, and valve disease. The pediatric cardiologist must possess a profound knowledge of cardiovascular anatomy in all its pathological expressions, as well as the developmental processes.

Although the diagnostic tools available to both specialities are similar, their

pattern of use differs significantly. Often diagnoses are formulated on the basis of a physical examination and Echo-Doppler – with such diagnostic testing as cardiac catheterisation much less frequently performed. Functional testing and/or stress testing are often impractical and rarely relevant. By the time the patient is referred to a 'specialist pediatric center' the focus is often on developing an accurate and complete anatomical diagnosis essential for the surgeon to fulfill his task.

It is clear that a database for pediatric cardiology must differ in some profound ways from that designed for an adult practice. This difference is due to the necessity to record the pertinent facts in a situation where hundreds of distinct syndromes, associated conditions, and isolated defects prevail. This is realized in the diagnostic coding system described below. The database is used not only to store and retrieve patient information, but also to provide an efficient 'gateway' to the coding system.

The aim of this report is to describe the design and structure of the diagnostic coding system and to show how this has been integrated with the pediatric database. This is preceded by a brief description of the database, and followed by a discussion of reports generated by the system.

## **Materials and methods**

The database and diagnostic coding system run on IBM compatible PS-2 models 30 to 80, or AT equipment (using a 80286 or 80386 CPU). The minimum free random access memory is 512 kb. The program does run on the older XT class of personal computers, but speed of operation (particularly disk access) is not sufficient for most routine work. A 30 Mb hard disk is recommended, although the system software occupies about 1.5 Mb.

The system may be installed for either color or monochrome monitors. The VGA system is advised as the software can take advantage of the full resolution for displaying growth curves. The use of a color monitor is also to be preferred.

The use of a tape streamer or dual disk system is strongly recommended for performing regular backups, although the modern high density (1.4 Mb) disks provide a viable alternative for small databases (about a few thousand cases).

Both laser printers and dot matrix printers are supported, although the former is recommended because of its relatively quiet operation and capacity to print high resolution graphics.

The disk operating system (DOS) should be PC Dos vs 3.1. or higher. The programs are all written in the run-time version of Foxbase (a dBase III compiler). The database files, however, are all compatible with dBase III. A network version is available for Novell network users, although this requires the availability of a trained network supervisor.

The current design capacity is for a maximum of 10000 cases. This limit presents no fundamental obstacle as the database can be easily extended to 100000 cases providing sufficient disk capacity exists.

## Database

The principal design goals of this database were to:

1. record the patients diagnosis accurately and in an efficient, consistent way.
2. provide a useful overview of the diagnostic/therapeutic history of any given patient.
3. allow an individual doctor to document his activities in terms of patient care and use of resources.
4. assist departmental administrators by providing the tools to better manage human and material resources.

Support for research activities has become an increasingly active area of interest and one which the database does support via several standard report programs and user specified search and list criteria.

As in any clinical database, a number of 'administrative' items must be endured. As this is a common feature of virtually all clinical databases, this aspect will not be dwelt upon. Our policy has been to record only the essential administrative information. This also extends to diagnostic tests such as the ECG, Chest X-ray, Phonocardiogram, etc. The date and number of procedures are recorded, but the details are left in the original laboratory reports. With the possible exception of Echo-Doppler and cardiac catheterisation, only the conclusions, in as far as they contribute to a diagnosis, are completely recorded.

The database is organized by the type of patient encounter as shown in Table 1 below. The section on 'Administrative data' registers the patient in the database and must be completed before any of the other items can be selected. This includes such records as the patients identification, birth weight and height, address, parents or guardians, general practitioner, referring specialist, and type of insurance.

*Table 1.* Type of patient encounters recorded in the database

- 
1. Administrative data
  2. Out-Patient consult
  3. Intercolleague consult
  4. Admission cardiology
  5. Admission thoracic surgery
  6. ECG/PHONO/ECHO etc. No consult
  7. Mortality
  8. Diagnosis
- 

An 'Out-Patient consult' is classified as either a first visit or follow-up visit. This is useful for documenting departmental activities in relation to new patients vs. the existing case load. The data recorded here are limited to date of consult (thus age), weight and height, responsible physician, routine laboratory tests (ECG, Phono, Chest X-ray, Echo-Doppler, Rhythm/Conduction-Exercise ECG, Holter, Pacemaker check), and conclusion.

The 'Intercolleague consult' is very similar to the 'Out-patient consult' in structure. The intention is to note the performance of professional activities which have a bearing on the staff work load.

'Admission to cardiology' contains on one screen the essential information: admission indication, date admission (thus age), weight and height, responsible cardiologist, laboratory tests (ECG, Phono, Chest X-ray, Rhythm/conduction-Exercise ECG, Holter, Pacemaker check, Programmed electrical stimulation, Echo-Doppler) as well as the performance of balloon dilatation, cardiac catheterisation and re-catheterisation. The date of discharge is recorded and the number of bed-days is computed automatically. The existence (date) of the routine laboratory tests are noted. Essential data from Echo-Doppler and Cardiac catheterisation can be entered at the user's discretion (the option is only offered if the date is already entered).

'Admission to thoracic surgery' is very similar to 'Admission to cardiology', but with the addition of date of surgery, whether 'Open' or 'Closed' heart surgery performed, and date of re-operation. Note that all the details surrounding the sort of operation and complications are recorded via the 'diagnostic coding system'.

The option 'ECG/PHONO/ECHO etc. No consult' is used to note the use of departmental material resources, but without involving the activities of professional staff. This is of interest for preparing the departmental budget and for resource planning.

'Mortality' allows for the apart recording of death. This includes such information as date of death, location (coded as home, during admission, etc.), cause of death, date of post-mortem examination, post-mortem registration number and heart preparation number.

Finally 'Diagnosis of a patient' may be independently accessed in connection with any of the above described options. The structure of the coding system is described in the following sections.

### **Diagnostic coding system**

The Diagnostic Codes are based on the sequential segmental analysis [1] originally developed for use in the Brompton Hospital in London. The segmental analysis used here is particularly appropriate for those patients who have a heart characterized either by abnormal atrioventricular and/or ventriculo-arterial connections or abnormal relations between atria, ventricles, or great arteries. These features are described step by step following a flow chart from the venous to the arterial pole of the heart.

- Atrial situs, A-V Connection, V-Art. Connection
- Great veins and coronary sinus
- Atrium and atrial septum
- Atrioventricular valves and AV septal defect
- Ventricle and ventricular septum

- Arterial valve level and great arteries.

This list has been supplemented by the extensive diagnostic code for congenital heart disease developed by van Mierop [2]. As many patients have normal connections and relations, the above items can be bypassed and other aspects of the diagnostic coding system accessed. As both diagnostic coding systems focus on the anatomical description of congenital cardiovascular malformations a number of topics have been added to the main menu. These include:

- Acquired heart diseases and Other
- Position heart and its component parts in thorax
- Position thoraco-abdominal organs
- Rhythm and conduction disturbances
- Associated non-cardiac anomalies

Finally, for practical reasons it proved necessary to include a number of therapeutic activities in the 'diagnostic' coding system. These include:

- Medication
- Operations and interventions

### **Coding system design requirements and policy**

The above thirteen information categories form a basis for identifying over 1700 individual diagnostic/therapeutic codes. It is essential to be able to quickly and consistently identify the required codes.

Since most users are not familiar with database languages, a command structure is contraindicated in favor of a menu structure. This consists of a logical grouping of a reasonable number of sub-menus and items per menu. Furthermore, it allows for the selection of multiple menu items wherever appropriate. It is possible to step backwards through a menu to undo mistakes, and support the entry of 'free text' for some items in order to further elaborate on a standardized diagnosis. At any stage it is possible to examine the current status of the selected codes.

The use of a menu structure has two additional beneficial effects: 1) the user is continuously reminded of the alternatives/options available which encourages a more complete diagnostic coding; 2) this structure forms a useful teaching tool for students in anatomy, pathology, (pediatric) cardiology, and cardiac surgery.

It is recognized, however, that a menu-selection system is not always the preferred approach. Three other options are also available: 1) selection of diagnosis/therapy by direct entry of the code number; 2) search for diagnoses on the basis of a keyword; 3) search for diagnoses on the basis of a word or phrase contained in the diagnosis. We mention in passing that all three approaches are supported, but that the operational details are not presented in this report.

Each diagnosis/therapy is assigned a six digit code number. Whenever possible this is consistent with the original Brompton coding. Generally the

leading digits of the code number indicate the class of diagnosis/therapy. For example, all surgical interventions begin with '12', all operations on the mitral valve begin with 1203, while the item 'mitral valve replacement with Bjork-Shiley valve' is code '120317'. This facilitates the retrieval of information related to classes of diagnoses/interventions. The final diagnosis of a patient can be easily translated and reported in different languages, facilitating international exchange of information as well as unambiguous communication between referring institutions.

In total over 1700 individual cardiovascular related diagnoses /therapies are defined, although only the code number is actually stored in the database. The text belonging to the code is quickly retrieved via a related database dictionary indexed on the code number. This allows for efficient storage of the essential information and provides for faster retrieval than would be the case if the full text for every diagnosis were recorded. Also, it has the potential advantage of supporting diagnostic dictionaries in more than one language without having to change a single item in the patient's database.

### Coding system implementation

Upon selecting 'Diagnosis' from the main database menu, the current diagnosis of the default patient is displayed (see Figure 1). This includes the data attached to the diagnosis, the source or basis of the diagnosis, a column for user selected codes (note \* = 'incorrect diagnosis'), and finally the text of the diagnosis/therapy itself.

NR	DATE	INVEST.	COM.	TEXT	B__,A.B. (ALBERT)
1	06/03/88	D		Fallot	
2	07/03/88	AB	*	Turner syndrome	
3	11/03/88	E		Correction Fallot with resection muscle tissue RV outflow area and P. valvulotomy	
4	13/03/88	E		Postoperative complication= rhythm- and conduction disturbances	

[I]nput from "Diagnostic Coding System", [A]ppend,  
[C]hange sequence, [#] = modify line nr., or ←: █

*Fig. 1.* Existing diagnosis/therapy as displayed for the current patient. Selecting [I]nput accesses the diagnostic coding system, while [A]ppend supports adding diagnoses via selections from key words or directly via specifying the diagnosis code number.

At the end of the list the user is offered the option of accessing the diagnostic coding system for the purpose of adding new information (this can also be done using the key word retrieval system if desired). Upon selecting the option to access the coding system, this aspect of the database is loaded, and the coding procedure begins.

Since only the diagnostic codes are stored in the patient's file, the entire coding system can be viewed as an independent program whose task is to deliver, via one-way traffic, user selected 'diagnostic codes'. As such the coding system could be developed in isolation from the patient database. The implication of a design based on choices selected from a menu which point to other sub-related menus, is that the definition of the menu dictionary forms a logically structured database, or 'intra-related' database. The medical intelligence is fully contained in the logical structure while speed of operation, ancillary features, and 'user friendliness' is the province of the program accessing the information. The use of the term 'database' in this context may appear unusual in the sense that it is strictly a read-only database and would not be frequently up-dated. It is certainly not the intention that database users would modify its contents. Indeed, because of the intra-related aspects, a separate program is used to generate the database.

### Main access menu

Often the patient's diagnosis consists of one of the commonly occurring cardiovascular syndromes. It may then be preferred to specify the syndrome rather than all the distinct anatomical components. The main access menu allows for this possibility. It also allows the user to directly enter the more

Diagnostic Coding System Pediatric Cardiology

24 September 1989

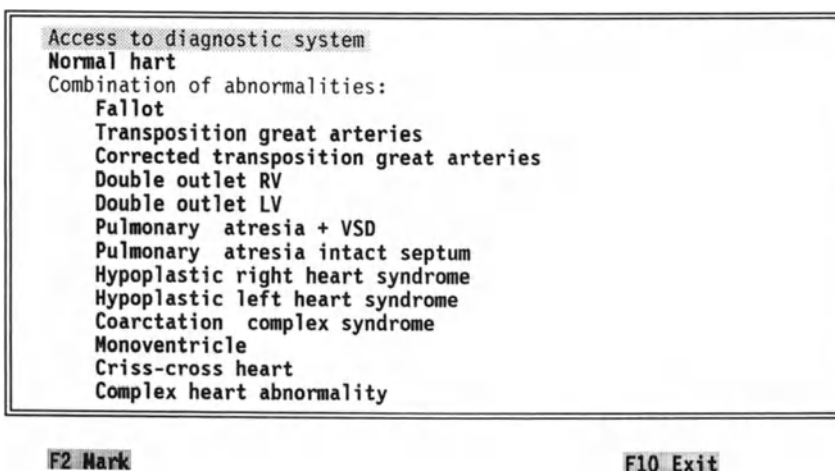
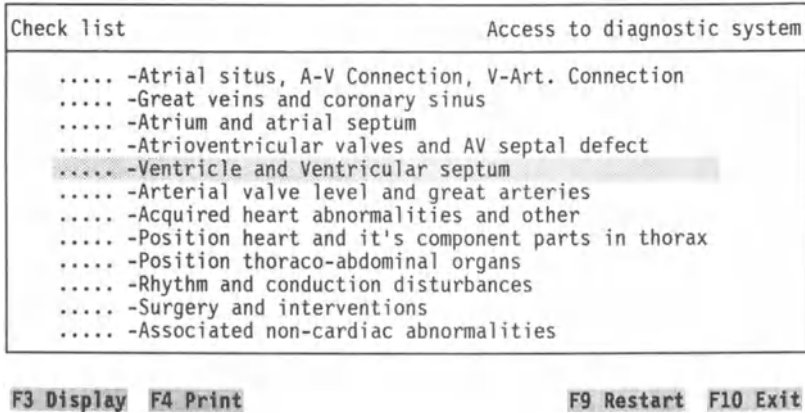


Fig. 2. Main access menu displaying the commonly occurring cardiovascular syndromes. Any one or more of these items can be selected. In some cases this completes the diagnostic coding. In others the 'Check list' shown in figure 3 must be accessed for a more complete description.



*Fig. 3.* The 'check list' is displayed after selecting a specific cardiovascular syndrome (see figure 2) or directly by selecting 'access to coding system'. This menu controls access to the detailed diagnostic codes. The cursor is initially positioned on the first line, although the cursor may be moved and the items selected in any order. Upon completing the specification of the diagnostic codes pertaining to each line, the leading dots (.....) are replaced with the flag 'done', and the cursor automatically re-positioned at the next item.

systematic coding system proper. The current options are shown in Figure 2. Note that this includes the item 'Normal heart' to make evident the fact that the pathology lies elsewhere. As these items include the most frequently occurring diagnoses, the entire diagnostic coding procedure may be concluded at this stage, while the option is available to go into much greater detail.

### Check list

The menu shown in Figure 3 is displayed upon entering the diagnostic/therapeutic coding system proper (with or without previously specifying a syndrome from the main access menu). This menu is referred to as the 'Check list' and provides for efficient access to all the available diagnostic/therapeutic codes. As each item is processed the leading dots are replaced with word 'done' and the cursor automatically moves on to the next line item. The cursor control keys are all active, and any item may be ignored or their order of execution altered.

It is beyond the scope of this report to describe the entire coding system accessible via this 'check list'. However, its basic structure together with a simple example is presented below. A complete specification is available from the authors[3].



### Structure of diagnostic system

The 'diagnostic codes' database consists of paragraphs which appear on the screen as a unit. The data file containing the paragraph is structured as follows:

```
N1, N2, Paragraph header text, N3
N4, Line text 1, N5, N6
N4, Line text 2, N5, N6
...
N4, Line text N2, N5, N6
```

N1 = Paragraph 'pointer' number  
 N2 = Number of line items in menu  
 N3 = Selection criteria/quality control code  
 N4 = Display status per line item  
 N5 = Pointer to intra-related paragraph number  
 N6 = Six digit diagnosis code number per line item.

For example, suppose that '- Ventricle and Ventricular septum' is selected. The first paragraph in the respective database file consists of the following:

```
1,5, 'Ventricle and Ventricular septum', 5
-1, 'RV, general          ', 10,-1
-1, 'RV, outflow tract   ', 20,-1
-1, 'LV, general         ', 30,-1
-1, 'LV, outflow tract   ', 40,-1
-1, 'Ventricular septum  ', 60,-1
```

Five line items are displayed (see Figure 4) and up to all of these may be selected at any one time. The 'display status' (code (-1)) indicates that all these items serve as pointers to sub-menus and that the text itself is not used to form the text of the diagnosis. Each line item points to a different paragraph elsewhere in the same database file. The diagnostic code number is not yet relevant, and its position is held by the code '-1'.

The line items listed in the menus/sub-menus have two functions: the line may represent a specific diagnostic code, or it may serve as a 'pointer' leading

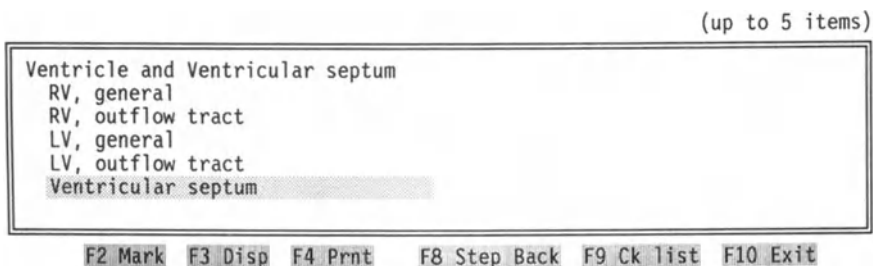


Fig. 4. Menu as displayed on the screen allowing selection of 'Ventricle and Ventricular septum'. All five items may be selected via the 'Mark' key.

to more specific codes. Some items serve both functions. Those which serve exclusively as 'pointers' are not printed in the report as the actual diagnosis will be coded at a subsequent stage. In order to distinguish between items which form part of the diagnosis and those which serve only as 'pointers', the diagnostic items are displayed with high intensity on the screen (shown as bold in the figures). Line items which serve only as 'pointers' are shown in normal intensity. All the line items shown in Figure 4 are general diagnostic categories which point to more specific diagnoses.

Suppose the first line item is chosen. This points to paragraph 60 which contains the following information.

```
60,2 'Ventricular septum', 2
+ 1, 'Ventricular septal defect (VSD)      ', 65, -2
- 1, 'other abnormalities ventricular septum', 90, -1
```

Here we may make up to two selections. The first line item (...VSD) points to paragraph 65. The diagnosis code number (-2) indicates that the precise code number will be assigned later, but that the currently displayed text will be included with the full text describing the diagnosis.

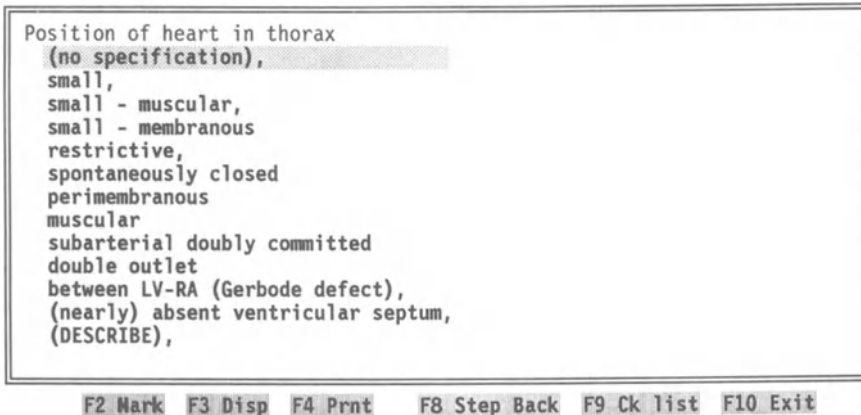
Paragraph 65 then allows for the specification of the details.

```
65,13, 'Ventricular septal defect (VSD)',9
+ 1, '(no specification)                ', 1,071000
+ 1, 'small                             ', 1,071501
+ 1, 'small - muscular                   ', 1,071502
+ 1, 'small - membranous                 ', 1,071503
+ 1, 'restrictive                        ', 1,071401
+ 1, 'spontaneously closed                ', 70,-2
+ 1, 'perimembranous                     ', 75,-2
+ 1, 'muscular                            ', 80,-2
+ 1, 'subarterial double committed        ', 85,-2
+ 1, 'double outlet                       ', 86,-2
+ 1, 'between LV-RA (Gerbode defect)     ', 1,071402
+ 1, '(nearly) absent ventricular septum ', 1,071403
0, '(DESCRIBE)                           ', 1,071403
```

Thirteen line items are displayed under the paragraph header (see Figure 5), but in this example no more than nine may be selected at any one time. Several line items have a unique diagnostic code number (071000 ... 071403) and all of these point to paragraph '1' which redisplay the main menu (providing there are no outstanding branches previously marked for processing). However, several items also point to further paragraphs where more detailed selections can be made. Where appropriate, free text can be entered and this is designated in the paragraph by the text '(DESCRIBE)': the entered free text replaces the instruction to describe the VSD. The code 'display status' code (0) indicates that the item allows for free text and should be so processed on the screen.

Naturally these details are not of interest to the clinician using the system.

(up to 9 items)



*Fig. 5.* Menu as displayed on the screen allowing selection of 'Ventricular septal defect'. Up to nine items may be selected.

The screen displays the essential information which can be selected with the minimum number of keystrokes.

Each menu/sub-menu is displayed on one screen together with several function keys (see Figures 4 and 5). Line items are selected by moving the cursor using the [up], [down], [PgUp], [PgDn] cursor control keys.

The special function keys execute the following operations.

- F2 Mark:** If more than one item is to be selected from a screen, the F2 key is used to flag all relevant selections. The system will first process the first marked branch to completion then return to follow-up on all outstanding open branches in the order they appear on the screen.
- F3 Display:** Displays on the screen all selected diagnoses.
- F4 Print:** Prepares a report listing all selected diagnoses.
- F3 Step Back:** Returns to the previous screen in the diagnostic tree, used to recover from mistakes.
- F9 Ck list:** Goes directly back to the Check List (Figure 3).
- F10 Exit:** Leaves the program but saves all selected diagnoses.

The [Enter] key commits all marked branches to memory for subsequent processing, which also includes the current line item whether or not marked. That is, if only one item is to be selected the cursor can be moved to the required item and the [Enter] key pressed.

When the coding is completed, control returns to the data base program. This requests some further information such as the date of diagnosis, and a user definable code. The new diagnostic codes are then automatically appended to the patient's file and will be displayed in chronological order the next time the 'diagnosis' is accessed.

## Reports

Several classes of reports are available via menu selections. This generally falls into four classes of reporting:

1. Patient oriented, pertains to one specified patient
2. Group oriented, pertains to groups of patients
3. Resource oriented, i.e. use of material resources and manpower
4. User specified list and search criteria

A full description of these reports is beyond the scope of this report, although extensive documentation is available from the authors [3].

Patient oriented reports provide information ranging from a complete printout of the patients' administrative and medical history to a very abbreviated summary of the essential conclusions. Also the option exists to plot growth curves (height and weight vs time) within the framework of the database.

Group oriented reports are available for preparing the agenda of planned patient consults and admissions. The option also exists to retrieve patients names and medical histories for all cases with one or more user specified diagnoses.

An extensive set of reports have been developed which are of value to department administrators. This includes the preparation of departmental monthly/quarterly and annual reports documenting use of all recorded material resources, broken down by type of insurance coverage. Also, the level of activity (patient consults, admissions, etc.) for each physician in the department as well as an inventory of referrals per institution is available.

Finally, reports are available which allow the user to specify his own search criteria and items to be listed. This is intended to support research activities and provide the user with the capability to extract information from the database which is not yet covered in the other standard reports. The option exists to export this data to floppy disk as a standard ASCII file which can be further processed by most standard statistical packages (SPSS, SAS, BMDP, etc.), or displayed using one of the standard graphics packages available for the personal computers.

## Summary

A working group, established in 1986 by the Dutch Pediatric Cardiology Centers, has completed development of a database management system pertinent to pediatric cardiology and pediatric thoracic surgery. This system supports a practical cardiovascular diagnostic system as well as recording essential patient administrative and medical information. The primary aim of setting up the program was to improve patient care. This required the development of a diagnostic coding system to allow the efficient entry of complete and accurate diagnoses in a consistent way. Additional goals were to

enhance the efficient management of departmental resources, stimulate inter-institutional cooperative studies, and provide a research platform for individual physicians. The coding system contains over 1700 specific items covering practically every anatomical, surgical and cardiological related diagnosis and intervention. Many other relevant pediatric anomalies have been added. The anatomical diagnoses are based on the (Brompton) segmental analysis accessed by a choice of 1) direct access to each separate item (via code number), 2) access by key word, 3) a stepwise approach through extensive diagnostic 'trees'. The database system includes, besides the patient administrative /medical /diagnostic /therapeutic information, an integrated growth chart program, follow-up agenda and mailing system as well as several standard retrieval and reporting programs. The system is now successfully operational in all pediatric cardiac centers in the Netherlands as well as a number of other major hospitals in the European Community.

### **Acknowledgements**

The authors gratefully acknowledge the generous support from the Dutch Heart Foundation. The project was subsidized through NHS project No. 35.005 and project No. 13.025.

### **References**

1. Miller GAH, Anderson RH, Rigby ML (1985) Commentary on the diagnostic code – The Brompton Hospital Diagnostic Code. In *The Diagnosis of Congenital Heart Disease*. GAH Miller, RH Anderson, ML Rigby (eds), Castle House Publications Ltd pp 73-120
2. Van Mierop LHS (1984) Diagnostic Code for Congenital Heart Disease. *Ped Card* 5: 331-362
3. Brower RW (1989) Pediatric cardiology database management system, Publ Medical Computing Consultants, POB 35010, 3005 DA Rotterdam, The Netherlands

# Databases and decision system for diagnosis of congenital heart disease

IVAN MALČIĆ<sup>1</sup> and DARKO IVANČEVIĆ<sup>2</sup>

*Department of Paediatric Cardiology<sup>1</sup> and Department of Nuclear Medicine<sup>2</sup>, University Hospital Rebro, Zagreb*

## Introduction

In paediatric cardiology a database is indispensable in order to conduct more extensive and detailed investigations. A database should contain large amounts of data since even the most ordinary information from medical history and status praesens of a patient provides better understanding of the disease. A particular problem is coding of information because of the difference between qualitative and quantitative information. In our study database served as 'raw data' in developing for diagnostic decisions in case of congenital heart disease (CHD). In this work we shall describe the collection of information/data (Part one) and the system of decision analysis in the process of diagnosing congenital heart disease (Part two). Large amounts of data, both qualitative and quantitative could be collected using following methods:

1. classification of data,
2. coding of data by means of codex of chiffres,
3. storing (retrieving in large systems),
4. data processing using suitable statistical programmes (Part one).

Diagnostic decisions during the diagnostic process in congenital heart disease (CHD) must be supported by suitable statistical and mathematical principles. In the majority of children with CHD, the diagnosis should be made before the complete diagnosis procedure has been accomplished. Our aim is to support the rationalisation by diagnostic procedure with an estimate of every diagnostic step. This approach is possible with the following principles:

1. ROC analysis of all diagnostic step procedures
2. Inclusion of Bayesian analysis (the principles of conditioned probabilities)
3. The value analysis of diagnostic step methods incorporated in an elaborate theoretical model, and this model in flow diagram for every type of CHD
4. The elaborated theoretical and mathematical model beneficial in the routine with the aim to achieve: maximal information (possible probability of correct diagnosis 1.00) with minimal risk for patient and doctor (the minimal number of diagnostic steps) and cost benefit.

The system of diagnostic decisions during the diagnostic process in CHD

would be developed by introducing the data in the decision process with a concurrent application of the methods described in Part two of this study.

## PART ONE

### 1. Databases in paediatric cardiology

Data were collected for each patient by means of a questionnaire. In order to enable further data processing the following procedures were applied: classification of data, coding, recording of data on magnetic tape, data processing.

#### *1.2. Classification of data*

The data were classified into three basic groups in accordance with the groups of patients included in our study:

1. thoroughly examined patients with non-invasive diagnostic methods and catheterisation, with left-to-right shunt in final diagnosis.
2. thoroughly examined patients with non-invasive diagnostic methods and catheterisation without L-D shunt in final diagnosis.
3. patients examined with non-invasive diagnostic methods but without catheterisation and dismissed only due to non-invasively diagnosed innocent systolic murmurs. All data were coded using codex of chiffres.

#### *1.3. Coding of data. Codex chiffres*

The majority of data were of qualitative (nominal) character, and only a minority were of quantitative (continuous) character. The reason is that results, observations and examinations are in practice usually given qualitatively, that is nominal. They are by their meaning subjective and reflect nosology, that is a physician's knowledge of the disease. The smaller part of the data are numerical, and for these quantitative characteristics another statistical model was used.

In order to process different characteristics, the data (values) are translated into a common language using codex of chiffres. Quantitative data (values) are coded by order of numbers with the help of the decimal principle where necessary. Codex contains a total of 119 variables in 195 positions, each position having a precise symbol. In this way 23 205 symbols were coded (input) taking care of the quality, relationship and logic of the chiffres.

#### *1.4. Recording of data on magnetic tape*

Coded data were stored on a magnetic tape using a numerical system. The data stored in a central unit served to access the data by means of a message and

programmes from the terminal. The data were recorded in the form of series of data and tables which served as 'raw data' to be analysed and to obtain final results reported in Part two of this study.

### *1.5. Data processing*

Data were processed in the University Computing Centre (SRCE) in Zagreb using UNIVAC 1100 and OS 1100 EXEC-8 operating system. Analysis related to nominal variables was performed by means of CONTAB programme [1]. By means of this programme all interactions of marginal variables were performed and frequencies were calculated as well as probabilities that certain entity (a patient) belongs to this unit (cell) and conditional probabilities with regard to categories of marginal variables. Assuming that crossed marginal variables are statistically independent, a frequency by cell of contingent table was calculated and tested by  $\chi^2$  test. From the selected tables obtained by crossing of marginal variables, matrices of decisions were obtained, that is ROC-tables. Continuous (qualitative) variables were analysed by means of STATJOB programme package [2]. Samples were described (arithmetic means, standard deviation, variance, range) and variable dependence (correlation) was tested using DSAT 2 model from the mentioned programme package. Parameters were estimated using single and multiple linear regression with STATJOB package, model STEPREG 1. STEPREG 1 values in addition to basic statistics over the group of variables in process estimates parameters in the model:

$$Y = XB + \text{eta.}$$

Vector Y (N1 type) stands for the value of all N entities of the dependent variable, and the columns of X matrix (N × K type) stands for the values of all entities for K independent variables X... X<sub>k</sub> so that the vector of parameters estimated K × 1 and eta is the vector of the independent, random variable N × 1 type. STEPREG 1 programme at work yields the following important elements: regression coefficients, determination coefficients, corrected determination coefficients, multiple correlation, standard estimation error, standard regression coefficients, standard regression coefficient error, t-values for regression coefficient with N-k degree of freedom, partial correlation and table of analysis of variance for linear regression.

### *1.6. Databases in radioangioscintigraphy (detection of left to right shunt)- example for databases*

Radionuclide angiocardiology has been performed by injection of Tc-99m-pertechnetate into the subclavian or external jugular vein. The passage of radioindicator through heart and lungs has been followed by computerized large field of view gamma camera. After analog-to-digital conversion data from gamma camera were collected into Gamma-11 DEC computer system in 64 × 64 matrices with 5 or 10 frames per second during 20 or 40 seconds.



Dynamic studies were analysed by region of interest technique and transformed into dynamic curves of heart, right atrium, left ventricle, lungs, abdominal aorta, head etc. [3]. Qualitative detection of left to right shunt has been performed by inspection of the logarithmic plot of lung and left ventricle curves [4]. Quantitative analysis was done from both lung activity curves using the computer program based on the method of Maltz and Traves [5]. The original lung curve is smoothed and background subtracted and then fitted to gamma variate function creating a curve simulating pulmonary flow. This curve is subtracted from the original pulmonary curve to create a third curve representing pulmonary recirculation. This curve is fitted with the gamma variate too and simulates premature left to right recirculation. The program calculates QP/QS reflects the left to right shunt magnitude. Previously many comparisons were made between this method and the oxygen saturation method and proved its validity in left to right shunt quantitation.

PART TWO

2. Decision analysis supported with databases

2.1. ROC-analysis (Decision analysis)

2.1.1. Two by two: Decision matrices. Let us start with a simple situation: one of the diagnostic methods in paediatric cardiology (for example echocardiography) yields two results, positive (disease present) or negative (disease absent). There are four possible outcomes: true positive (TP) when the test is positive in patients with disease, true negative (TN) when the test is negative in patients without disease, false positive (FP) when the test is positive in patients without disease and false negative (FN) when the test is negative in a patient with disease which may be diagrammed in a decision matrix (Table 1). The sensitivity is the fraction of TP in all patients with disease = TP / (TP + FN). The specificity is the fraction of TN in all patients without

Table 1. A general decision matrix

Test results	Disease present (D +)	Disease absent (D-)	Total
Abnormal (T +)	a (TP)	c (FP)	a + c
Normal (T-)	b (FN)	d (TN)	a + d
Total	a + b	c + d	a + b + c + d

Derived ratios

- (1) True positive ratio (sensitivity) =  $a/a + b$
- (2) True negative ratio (specificity) =  $d/c + d$
- (3) False positive ratio =  $c/c + d$
- (4) False negative ratio =  $b/a + b$

disease =  $TN/(TN + FP)$ . The always positive diagnosis has 100% sensitivity and 0% specificity, the always-negative test has 100% specificity and 0% sensitivity [6, 7]. The reference for all other diagnostic methods is the definitive diagnosis.

**2.1.2. ROC-diagram.** Let us transform the general decision matrix in ROC-diagram: The outcome diagnosis can be expressed with five (or more/or less) nominal values. For example, the diagnosis with a suitable diagnostic method may be:

1. accurately positive,
2. probably positive,
3. suspect,
4. probably negative,
5. accurately negative.

Also, with five nominal values for the outcome of the diagnosis there are four criteria and four values for the derived ratios by ROC-diagram (the ROC term will be explained later with ROC curves) (Table 2) (in 'Example' Table 4) [8].

Table 2. ROC -diagram with five nominal values: The number of criteria = the number of values - 1.

basic diagnosis	Diagnosis with suitable diagnostic method					Total	
	a. pos.	A prob. pos.	B sus.	C prob neg.	D a. neg.		
CHD	+	$n_{11}$	$n_{12}$	$n_{13}$	$n_{14}$	$n_{15}$	N
	-	$n_{21}$	$n_{22}$	$n_{23}$	$n_{24}$	$n_{25}$	N
Total		N	N	N	N	N	N

**2.1.3. Bayes' theorem.** The primary role of any diagnostic test is to alter one's degree of suspicion of disease. Given a pretest probability (derived ratios from ROC diagram), one modifies this to a post-test likelihood based on the test results. Although the accuracy of a diagnostic method answers the question, 'how often does the test give the correct results?', it does not answer the more clinically relevant question, 'what is the probability that a patient has or does not have the disease, given a positive or negative test result?'. This question is concerned with conditional probability, the chance that a certain condition (i.e. the test result) is fulfilled. With the use of Bayesian analysis (first described by Rev. Thomas Bayes in 1763), one can calculate the influence of new factors (i.e. new test results) on previous probability and thus quantify the conditional probability of disease. Introducing the principle of conditional probabilities (diagnostic versus nosologic), the nosologic values in ROC-diagrams were marked with p (probability-p). This may be accomplished by dividing the number of patients in every column with the total number of patients (Table 3) (in 'Example' Table 5). The derived values of these tables (ROC values) were marked with p(SE) (probability of sensitivity), p(SP) (probability of specificity),

Table 3. ROC -diagram with conditional probabilities:: The outcome diagnosis with suitable method.

Diagnosis	Accurately positive	Probably positive	Suspect	Probably negative	Accurately positive	
CHD	+ P <sub>11</sub>	P <sub>12</sub>	P <sub>13</sub>	P <sub>14</sub>	P <sub>15</sub>	1
	- P <sub>21</sub>	P <sub>22</sub>	P <sub>23</sub>	P <sub>24</sub>	P <sub>25</sub>	1
	Z <sub>1</sub>	Z <sub>2</sub>	Z <sub>3</sub>	Z <sub>4</sub>	Z <sub>5</sub>	

$$P_{11} + P_{12} + P_{13} + P_{15} = 1$$

$$P_{21} + P_{22} + P_{23} + P_{25} = 1$$

p(FN) (probability of false negatives), p(FP) (probability of false positives). Bayesian analysis acts as a post-test. The optimal point of correct diagnosis in ROC diagram with conditional probabilities has the greatest probability of the correct diagnosis p(CD) and the lowest probability of the false diagnosis p(FD). This relationship was proposed in the following formulas:

- (1)  $p(CD) = p(SE) p(DP) + p(SP) p(DA)$
- (2)  $p(FD) = 1 - p(CD)$ .

Bayes’ theorem calculates the likelihood of disease by considering the pretest probability as well as the sensitivity and specificity of the examination. That is, the predictive value of a test (posterior probability) in a particular patient can be determined if the disease prevalence (a prior probability) is known. By comparing pre- and post-test probabilities over a range of population prevalence, Bayesian analysis permits comparison of two different tests as to which yields a greater change from pre-to-post test and estimation of the range of prevalence of disease for which given test is worth ordering. Tests, in general, are useful when the probability of disease is intermediate. When the pretest probability is low, a negative result has little effect, but a positive result challenges the clinical assumption. In this situation, a highly specific test is needed so that a positive result can be trusted. When the pretest probability is high, a negative result can substantially change expectations but a positive test has little effect. In this case specificity is unimportant but high test sensitivity is needed to validate a negative result. Using Bayes’ theorem one can determine the reliability of a test result at a given disease prevalence [9-12].

**2.1.4. Practical application.** In paediatric cardiology, the Bayesian analysis has been used to determine the place of suitable diagnostic method

1. routine non-invasive method including clinical investigation, x-ray, ECG,
2. echocardiography (ECHO) as well as M-mod, 2-D, 2-D with Doppler,
3. radioangioscintigraphy (RAS) for detection and quantitative evaluation of intracardiac shunts,
4. invasive diagnostic (catheterisation with oximetry and angiocardiology) in the diagnostic flow-diagram (see ‘Example’, Figure 4).

*2.1.5. Receiver operating characteristic (ROC) curves.* One way to quantitate the effect of differing thresholds for test positivity is the use of receiver operating characteristic (ROC) curves [10]. An ROC curve is a plot of false positive ratios versus true positive ratios over a range of interpretative criteria. The term is derived from the early days of radar when interpreters sought to distinguish signals caused by airplanes from 'noise' caused by other sources. The ROC curve is unique in that it can measure diagnostic accuracy and simultaneously describe utility. The method has been used in a variety of ways in medicine, but it is possible to use the ROC curves for all the methods in decision flow-diagram in diagnosis of CHD. The ROC curve can graphically represent the relationship between sensitivity and specificity as criteria vary (see 'Example', Figure 2). Thus, interpretative criteria yield the optimal test accuracy. ROC curves have been used to calculate the diagnostic gain and added costs resulting from the introduction of further tests. If we incorporate in ROC analysis the table of conditional probabilities (Bayes' analysis),  $p(SE)$ ,  $p(SP)$ ,  $P(FP)$ ,  $P(FN)$  with  $p(CD)$  (see Equations (1) and (2)), this approach can be applied in a series of test results, that is for different diagnostic methods in diagnostic flow diagram. The optimal number and type of diagnostic methods can be selected for a given condition with due consideration for cost and risk [13, 14, 15]. The diagnostic flow-diagram should be stopped at the point where the probability of correct diagnosis is 1.00 or near 1.00. (see flow diagram in 'Example', Figure 5.). In CHD it is important to know:

1. the type of CHD,
2. the size of the defect
3. the place of the defect and
4. excluded excessive pulmonary hypertension.

Also, although the probability of correct diagnosis is 1.00, diagnostic procedure can be enlarged with a suitable diagnostic method (see 'Example', Figure 5, interaction between echocardiography and radioangioscintigraphy).

*2.2. Theoretical and mathematical presentation of the model for the flow diagram in the diagnostic procedure- 'input-output system'. Practical application of the principles described above*

Diagnostic procedure can be represented as a group of tests (a system of tests) the aim of which is a correct diagnosis. Every test may be defined as a single 'input-output' system. For accurate test definition the following should be determined:

1. system variables;
2. output of system transfer;
3. output function of the system.

When considering variables one has to bear in mind the input variables, the conditional variables and the output variables. The transfer function of the system serves to define the transition of input variables into the output variables. This may be presented graphically where the input variables are marked with  $X_1$  and  $X_2$ , conditional variables with  $Z_1$   $Z_2$ ... $Z_5$ , and output variables with

R O C

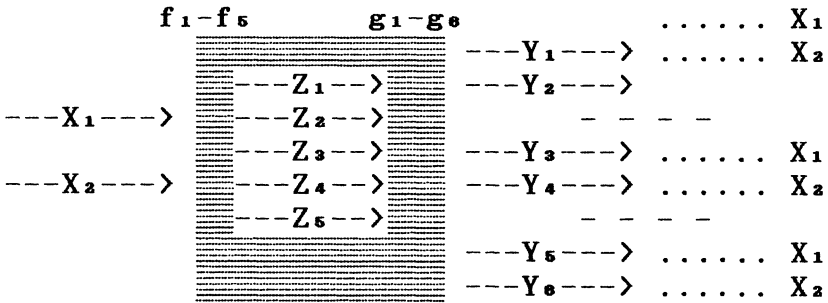


Fig. 1. Theoretical and mathematical presentation of the model for the flow diagram in the diagnostic procedure – ‘input-output system’.  $X_1$  and  $X_2$  = input variables, ROC = Receiver operating characteristic,  $Z_1$ - $Z_5$  = conditional variables,  $Y_1$ - $Y_6$  = output variables,  $f_1$ - $f_6$  and  $g_1$ - $g_6$  = transition functions

$Y_1...Y_2...Y_6$ . The graphic presentation of ‘input-output’ system (Figure 1) is specified for every decision making point.

$X_1$  = number of patients that undergo suitable diagnostic methods

$X_2$  = probability that a patient who undergoes an appropriate diagnostic method has the expected congenital heart disease.

The transition functions may be marked with  $f_1, f_2...f_5$  and they imply the change of input variables into the condition variables

$$Z_1 = f_1 (X_1, X_2)... \dots Z_5 = f_5 (X_1; X_2).$$

In this paper, the ROC analysis is taken to be the transmission function. The output function is marked with  $g_1$ - $g_6$ , and it actually implies the change of the condition variables into the output variables:

$$Y_1 = g_1(Z_1, Z_2, Z_3, Z_5, X_1, X_2)... \\ \dots Y_6 = g_6(Z_1, Z_2, Z_3, Z_4, Z_5, X_1, X_2).$$

The results of the diagnosis in each of the decision matrix sites may be represented by the diagram of conditional probabilities if the diagnosis is marked as  $j$ , and came from group  $i$ . Thus,  $p_{23}$  means the probability that the diagnosis is suspected in a patient belonging to the group without CHD. Presuming that the  $p_{ij}$  conditional probabilities are known, the transition functions may be determined. Therefore the following is obtained

$$Z_1 = p_{11}X_1X_2 + p_{21}X_1(1-X_2)... \\ \dots Z_5 = p_{15}X_1X_2 + p_{25}X_1(1-X_2).$$

Now the output variable should be determined. Let us presume that three patient groups are observed:

$$Y_1 = Z_1 = \text{number of patients with CHD}$$

$$Y_2 = \frac{p_{11}X_2}{p_{11}X_2 + p_{21}(1-X_2)} = \begin{array}{l} \text{probability that a patient has a CHD if} \\ \text{he belongs to patient group } Y_1 \end{array}$$

$$Y_3 = Z_2 + Z_3 + Z_4 = \text{number of patients in whom new diagnostic phase is required due to a suspect CHD}$$

$$Y_4 = \frac{(p_{12} + p_{13} + p_{14})X_2}{p_{12} + p_{13} + p_{14})X_2 + (p_{22} + p_{23} + p_{24})(1-X_2)} = \begin{array}{l} \text{probability that} \\ \text{the patient has} \\ \text{CHD if he be-} \\ \text{longs to patient} \\ \text{group } Y_3 \end{array}$$

$$Y_5 = Z_5 = \text{number of patients in whom CDH has been eliminated}$$

$$Y_6 = \frac{p_{15}X_2}{p_{15}X_2 + p_{25}(1-X_2)} = \begin{array}{l} \text{probability that a patient has a CHD if} \\ \text{he belongs to patient group } Y_6. \end{array}$$

Every output variable of one diagnostic procedure is an input variable for another diagnostic step. Thus  $Y_1$  is transferred into new  $X_1$  and the whole process is continued on the site of the new diagnostic decision. For the new corrected diagnostic step ROC analysis should be made. The number of those who were eliminated, by previous and by corrected diagnostic procedures should be compared [3].

### 2.3. Example

The study was based on the final outcome of diagnostic results in 100 examined patients in whom intracardiac left-to-right (L-R) shunt was suspected. All the patients were children (50 males, 50 females) aged 1-15 years. Of 80 thoroughly examined patients, 28 had atrial septal defect(ASD), 28 ventricular septal defect(VSD) and 14 persistent ductus arteriosus(PDA). Twenty patients were included by the random choice method from the group to patients with suspected CHD, and were dismissed with an innocent murmur, diagnosed only owing to routine non-invasive diagnostic (RND) [16]. In their final diagnostic, 30 patients were within a group in which thorough cardiologic examination had been undertaken.

The diagnostic flow with decision sites has been divided into several diagnostic phases: 1) RND, 2) echocardiography (ECHO) (without Doppler) [17, 18], 3) radioangioscintigraphy (RAS) [19, 20, 21], 4) catheterisation (CATH) [22, 23] and angiocardiology (ANGIO [24]. The first diagnostic decision was made after RND (including: physical examination, inspection, palpation, auscultation, ECG, X-ray, PHONO) the outcome of which has been arranged in five variations: 1) accurately positive, 2) probably positive, 3) suspect, 4) probably negative, 5) accurately negative (for RND and RAS). The outcome of diagnosis for ECHO and CATH has been arranged in three variations: 1) accurately positive 2) suspect and 3) accurately negative. The fourth diagnosis

was based on the outcome of oximetric analysis. In order to make the ROC analysis for oximetry it was graded in three steps: 1) negative finding (saturation difference lower than 0.5 vol%) 2) insignificant finding (saturation difference higher than 1.51 vol%) 3) significant findings (saturation difference higher than 1.51 vol%). This is an example of how it is possible to transform quantitative values into qualitative ones for the purpose of ROC analysis.

Table 4. ROC -diagram (nosologic probabilities) for RND in diagnosis ASD

basic diagnosis		Diagnosis in RND					Total
		A a. pos.	B prob. pos.	C susp.	D prob neg.	a. neg.	
ASD	+	12	11	2	3	0	28
	-	0	5	14	31	22	72
Total		12	16	16	34	22	100

The ROC diagram (Table 4) shows that the ascending flow comprised 28 patients with and 72 without ASD. ASD was positively diagnosed in 12, probably in 16 subjects; it was suspect in 16 subjects, probably negative in 34 and negative in 22 patients. The ROC diagram of conditional probabilities is obtained by dividing the total number of patients in each line with every number in the column. From the table of conditional probabilities (Table 5) which has five criteria, it is

Table 5. ROC -diagram of conditional probabilities for RND in diagnosing of ASD

basic diagnosis		Diagnosis probabilities of the RND					Total
		A a. pos.	B prob. pos.	C susp.	D prob neg.	a. neg.	
ASD	+	0.428	0.392	0.071	0.107	0.000	1.000
	-	0.000	0.069	0.194	0.430	1.000	1.000

possible to calculate four derived ratios obtained by ROC analysis, by changing criteria from A-D; p(SE), p(SP), p(FP), p(FN). We calculated p(CD) and p(FD) for every item of the ROC curve according to the equations described earlier (Table 6). In the case of ASD the following may be derived from ROC diagram p(DP) = 0.28 and p(DA) = 0.72. The point with the highest p(CD) and the lowest p(FD) value is the optimal point in the ROC curve. In the example presented this is the point B with p(CD) = 0.90 and p(FD) = 0.10 (the optimal points are asterisked in all analyses). Diagnostic procedures may be arranged according to the degree of significance related to the optimal point. The diagnostic method in diagnostic procedure must have the highest p(CD) value in the optimal ROC curve. Figure 2 shows the results graphically.

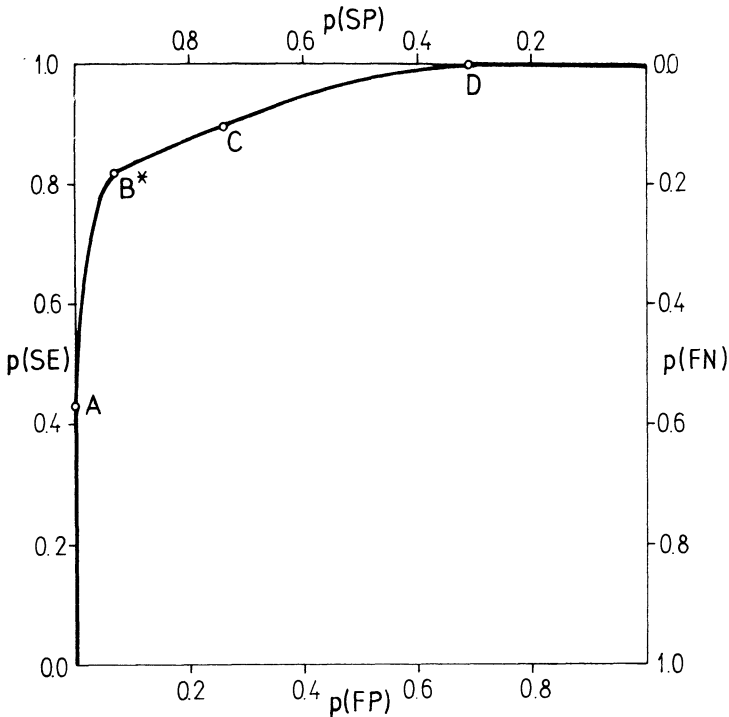


Fig. 2. ROC curve for RND by ASD. ROC = Receiver operating characteristic; RND = Routine non-invasive diagnostics; ASD = atrial septal defect; p = probability; SE = sensitivity; SP = specificity, FN = false negatives; FP = false positives; \* = optimal point; A-B-C-D = criteria.

Table 6. The ROC analysis for RND in diagnosing the ASD

Point	p(SE)	p(SP)	p(FP)	p(FN)	p(CD)	p(FD)
A	0.428	1.000	0.000	0.371	0.840	0.160
B*	0.821	0.930	0.694	0.178	0.900	0.100
C	0.895	0.736	0.263	0.104	0.780	0.220
D	1.000	0.396	0.694	0.000	0.500	0.500

**2.3.1. ROC analysis for the total diagnostic procedure in the diagnosing of atrial septal defect.** RND is only the first decision site in the diagnosing of ASD. Other decision methods are: ECHO, RAS and CATH (oximetry). Following the principles described in the introductory part and prior depicting of ROC analysis in RND, the results in Table 7 and Figure 3 have been obtained. The optimal ROC analysis points are asterisked. In the present, the optimal points of RND are:  $p(\text{SE})=0.8215$ ,  $p(\text{SP})=0.9305$ ,  $p(\text{FP})=0.0694$ ,  $p(\text{FN})=0.1785$ ,  $p(\text{CD})=0.90$ ,  $p(\text{FD})=0.10$ . ECHO was made for ASD in three axes (apical, subcostal, short at the base) and the results were graded as: 1) positive,



Table 7. ROC analysis for the whole diagnostic procedure in diagnosing of ASD

R O C	Criteria	diagnostic method for ASD			
	Point	RND	ECHO	RAS	CATH
p(SE)	A	0.428	*0.560	0.571	0.785
	B	*0.821	0.800	*0.892	*1.000
	C	0.892		0.892	
	D	1.000		1.000	
p(SP)	A	1.000	*0.945	1.000	0.980
	B	*0.930	0.781	*0.980	*0.961
	C	0.736		0.942	
	D	0.694		0.634	
p(FP)	A	0.000	*0.054	0.000	0.019
	B	*0.069	0.218	*0.019	*0.038
	C	0.263		0.057	
	D	0.694		0.634	
p(FN)	A	0.571	*0.440	0.428	0.214
	B	*0.178	0.200	*0.107	*0.000
	C	0.107		0.107	
	D	0.000		0.035	
p(FD)	A	0.160	*0.175	0.150	0.087
	B	*0.100	0.212	*0.500	*0.025
	C	0.220		0.475	
	D	0.500		0.425	
p(CD)	A	0.840	*0.825	0.850	0.912
	B	*0.900	0.787	*0.950	*0.975
	C	0.780		0.925	
	D	0.500		0.575	

2) suspect, 3) negative. The results of the ROC analysis were as follows:  $p(SE) = 0.5600$ ,  $p(SP) = 0.9454$ ,  $p(FP) = 0.0546$ ,  $p(FN) = 0.4400$ ,  $p(CD) = 0.8250$  and  $p(FD) = 0.01750$ . The outcome of the RAS analysis was marked with five values. Optimal ROC values were obtained only when probable positives were considered definitely positive together with their accompanying values:  $p(SE) = 0.8928$ ,  $p(SP) = 0.9808$ ,  $p(FP) = 0.0192$ ,  $p(FN) = 0.1072$  and  $p(CD) = 0.9500$  and  $p(FD) = 0.5000$ .

Oximetric analysis is incorporated in the ROC analysis following the above described criteria. The usual criteria are considered too strict, which was proven by determining the optimal oximetric values, if the shunt calculated for each existing difference was higher than 0.5 vol%. The accompanying oximetric values out of the ROC analysis with thus moderated criteria are, in their optimal point, as follows:  $p(SE) = 1.000$ ,  $p(SP) = 0.9600$ ,  $p(FP) = 0.0384$ ,  $p(FN) = 0.0000$ ,  $p(CD) = 0.9750$  and  $p(FD) = 0.0250$ .

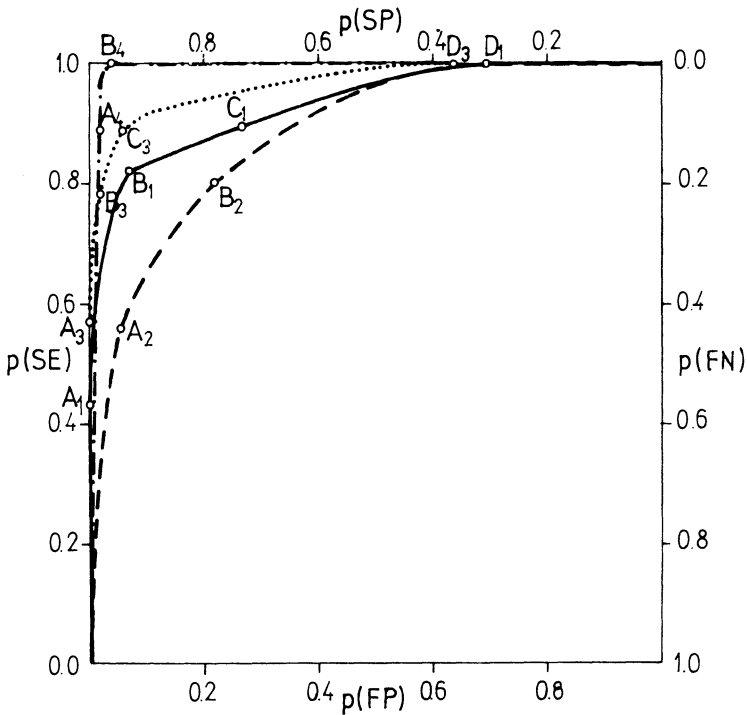


Fig. 3. ROC curves of the total ROC analysis in the diagnostic procedure by ASD. ROC = receiver operating characteristic; ASD = atrial septal defect; p = probability; SP = specificity; SE = sensitivity; FN = false negative; FP = false positive; A1-B1-C1-D1 = RND = routine non-invasive diagnostic; A2-B2 = ECHO = echocardiography; A3-B3-C3-D3 = RAS = radioangiostintigraphy; A4-B4 = CATH-oxym. = catheterisation-oximetry; A-D = criteria.

**2.3.2. Flow diagram analysis of the previous and newly recommended diagnostic procedure.** Diagram of the previous (Figure 4) and newly recommended diagnostic procedures (Figure 5) has been derived by introducing the results of the model described. The model is involved in every diagnostic step. Thus, a new ROC diagram (corrected) was constructed from each of them. It differs from the previous one in the number of patients excluded from diagnostic phase (regardless of whether the error was eliminated with high probability for a negative diagnosis or confirmed with high probability for a correct one). The newly proposed flow diagram differs from the previous one in the different patients flow. After RND, only the definitely positive were sent to ECHO test. However, the same patients may be instructed to have the extent of the defect measured by ROC analysis. Sixteen out of 28 patients with ASD may be singled out for surgery by a combination of ECHO and RAS analyses. This implicates 57.15% of all patients included in the study. The probability of correct diagnosis is not higher in the probably positive or suspect patients recommended for catheterisation. The results show that a combination of non-invasive methods (RND, RAS nad ECHO) plays a crucial role in deciding the



*Table 8. Comparative analysis of the former and the newly proposed flow diagram*

- 
1. Probability of correct diagnosis  $p$  (CD) in the whole procedure  
 $p$  (CD) = 0.97     $p$  (FD) = 0.03    in both diagrams
  2. The difference in the number of diagnostic steps in all patients with ASD:  
 RND 0, ECHO 5, RAS 6, CATH 16
  3. Staging of diagnostic steps: RND = 3, ECHO = 4, RAS = 2, CATH = 1
- 

analysis 10 more patients may be singled out with the probability of having ASD 1.00. This is 57.15% of the total number of patients with ASD.

#### *2.4. Conclusion*

Thanks to the analysis of decision making by means of the methods described (ROC analysis and Bayes' theorem included in 'input-output' model), it is possible to obtain maximal information on congenital heart defect before completing the total diagnostic procedure. In this way the diagnostic process is maximally rationalised. The diagnostic model proposed here reduces the number of examination tests without risk to the accuracy of the diagnosis. This protects both patient and doctor from overdiagnosis. Economic considerations are not negligible, either. The model could be applied in other branches of medicine, as well.

#### **References**

1. Zlobec L, Varga M, Momirovic K (1974) Program za analizu nominalnih i ordinalnih varijabli (in Croatian) Programme for analysis of nominal and ordinal variables. *Kineziologija* 4(1)
2. STATJOB (1973) Reference Manual of UNIVAC 1100 Series Computers, Academic Computing Centre the University of Wisconsin, Madison
3. Ivančević D, Price R, Rhea T, Graham I, Atwood GT, Brill AB (1975) Computer analysis of radiocardiograms of patients with intracardiac shunts. Reprint of radiocardiograms of patients with intracardiac shunts. Reprint from 'Dynamic studies with radioisotopes in medicine 1974', vol.II: 229-250: International atomic energy agency Vienna
4. Ivančević D, De Vernejoul P, Kellershohn C (1970) Radiocardiographic determination of systemic and pulmonary blood flow an atrial septal defect. Proceedings of the IST International symposium on nuclear medicine. Carsbad 67-74
5. Maltz DL, Treves S (1973) Quantitative radionuclide angiocardigraphy: determination of QP/QS in children. *Circulation* 47: 1049.
6. McNeil Barbara, Adelstein SK (1976) Determining the value of diagnostic and screening test. *J Nucl Med* 17: 439-48
7. Evans AL (1981) The evaluation of medical images. A. Hilger (ed) Ltd, Bristol in collaboration with the Hospital Physics Association, Glasgow.
8. Malčić I, Ivančević D, Tepeš B (1989) Decision analysis for diagnosis of congenital heart disease with left-to-right shunt. *Medicina e infomatika* 6: 167-177
9. Kassirer JP, Moskowitz AJ, Lau J, Pauker SG (1987) Decision analysis: A progress report. *Ann Intern Med* 106: 275-91

10. Pauker SG, Jerome PK (1980) The threshold approach to clinical decision making. *N Engl J Med* 302: 1109-17
11. Frybak DG (1978) Bayes' theorem and conditional nonindependence of data in medical diagnosis. *Comput Biomed Res* 11: 423-34.
12. Gorry GA (1978) Pauker SG, Schwartz WB (1978) The diagnostic importance of the normal findings. *N Engl J Med* 298: 486-89
13. Metz CE (1978) Basic principles of ROC analysis. *Semin. Nucl Med* 8: 283-298
14. Swets JA (1978) ROC analysis applied to the evaluation medical imaging techniques. *Invest Radiol* 14: 109-121
15. Turner DA (1977) An intuitive approach of receiver operating characteristic curve analysis. *J Nucl Med* 19: 213-20
16. Walker JK, Jones PRM, Dighton GP (1983) Non-invasive assessment of isolated atrial defect. *Br Heart J* 49: 163-66
17. Snider AR (1984) Use and Abuse of the Echocardiography Pediatric Clinics of North America 31/16, 1346-66
18. Radtke WE, Tajik AJ, Gau GT, Schattenberg TT, Giuliani ER, Taucredi RG (1976) Atrial septal defect: Echocardiographic observations. Studies in 120 patients *An Intern Med* 84: 246-53
19. Hurwith RA, Treves S, Keane JF, Gierod DA, Caldvelld RL (1982) Current value of radionuclide quantification and management in patients with secundum atrial septal defect. *Am Heart J* 103: 421-25
20. McIlveen BM, Murray PC, Giles RW *et al.* (1982) Clinical application of radionuclide quantitation of L-R cardiac shunts in children. *Am J Cardiol* 47: 1273-78
21. McIlveen BM, Hoschl PC, Murray PC *et al.* (1978) Radionuclide quantitation of left-to-right shunts in children. *Aust NZJ Med* 8: 500-508
22. Rudolph AM, Cayler GC (1958) Cardiac catheterisation in infants and children. *Pediatric Clinics of North America* 5: 907-43
23. Barratt-Boyes BG, Wood EH (1957) The oxygen saturation of blood in the venae cavae, right heart chambers, and pulmonary vessels of healthy subjects. *J Lab Clin Med* 50: 93-106
24. Bargeron L (1985) Is there a role for angiography in the diagnosis of congenital heart disease? *Pediatric Cardiology Congresses*. I Doyle (ed). II World Congress of Pediatric Cardiology, Abstr, New York

# Practical PC-based data management in paediatric cardiology

K.J. SPRENGER and M.M.A. DE MOOR

*Paediatric Cardiac Unit, Dept. of Paediatrics & Child Health, Red Cross War Memorial Children's Hospital, Rondebosch 7700, Cape Town, South Africa*

## Abstract

A cross-referencing data base is essential in a paediatric cardiology unit for research and audit of the results of the unit. In a busy unit it must be user-friendly, accessible, and practical. We describe a classification and coding system for paediatric cardiology. The system enables coding of segmental sequential analysis of cardiac pathology. It provides codes for diagnosis, operations and complications. It is practical and easy to remember and use, and runs on a local area network of personal computers. Our system has been in operation for 4½ years, and problems which have been encountered are discussed.

## 1. Introduction: the peculiarities of paediatric cardiology

Congenital heart disease is relatively rare, occurring in only 8-10/1000 live births, and acquired heart disease in children is even less common. The incidence of acquired heart disease in adults is considerably higher than in children, and there are many more adults than there are children, particularly in developed countries. These factors result in a much higher prevalence of adult cardiac pathology than of paediatric cardiac disease.

The greater work load in adult cardiology results in more financial investment in research and development in this field. Paediatric cardiology remains the Cinderella with slower progress in the development of technology and expertise. The use of computers in particular in paediatric cardiology has suffered this fate, lagging far behind the adult usage.

The variability and diversity of congenital heart disease has always been a stumbling block to the accurate description and coding of the pathology. This variability pervades all aspects of investigation and management of congenital heart disease in children. There is no 'standard catheterisation' in children and no 'standard echocardiographic' report.

Another obstacle to the description of the pathology in congenital heart

disease has been lack of a standard nomenclature. Elaborate, and very comprehensive, codes have been developed by some institutions [1, 2, 3] to classify congenital heart disease. Widely differing systems of description have, however, further impeded the development of computer database systems.

## **2. The data base problem**

Until October 1985 all cardiological data at our hospital was stored in A4 size envelopes, filed alphabetically. This system allowed only retrieval by patient name, and no method existed for extraction of data in any other form. It was not possible to easily establish mortality statistics, disease prevalences or any other information. Yearly audits of the department activities took weeks to prepare.

A cross reference data base was essential if our department was to be progressive. In order to solve this problem we established a data base on an in-house, PC-based, local area network.

## **3. Development of our computerised data base system**

In developing a software application there are two options. The development can be done by a computer scientist under the direction of the end user, or by the end user himself. Ideally both need to be involved in the process, but a more 'user friendly', practical application is produced if the primary input is from a computer-literate end user.

A computer system should not however be created without consulting a computer scientist who has expertise in the field of medical computing.

## **4. Requirements for practicality**

### *4.1. Simplicity and user friendliness*

The application in our environment had the same requirements as do all other software applications [4, 5, 6, 7], but simplicity and user-friendliness were emphasised in order to make the system practical. An important factor in our environment was staff shortage. A computer solution had to generate as little extra work as possible.

### *4.2. Accessibility of the hardware and software*

The system had to be multiuser, with the possibility of simultaneous access to the same data base by different users. The computer terminals had to be easily accessible to all staff members. The software had to be menu driven and easy

to use, in order that it could be safely used by staff who were not computer trained.

### *4.3. Applicability*

The software had to be practical in that all work generated by the system had to be seen to be useful and of direct benefit to the staff, or the patient, or both.

## **5. Coding systems**

### *5.1. Comprehensive vs grouping of morphological diagnoses – the ‘lumpers’ vs the ‘splitters’*

There have been a number of coding systems developed by various centers [1, 2, 3]. Most are hierarchical and very detailed. All eventualities are catered for and the codes are very precise. Small errors in coding however can result in major inaccuracies. Our philosophy was rather that we would develop an abbreviated code which was less precise but that was less prone to inaccuracy.

Although Weinberg [3] considers that a classification system must be all-inclusive, precise and accurate, we contend that this type of code is difficult and time consuming to use. Its very precision makes it liable to error and inaccuracy. In a very busy cardiological practice a classification which uses broader categories, but remains accurate is preferable to one that is very precise but loses accuracy in its usage. We opted therefore to ‘lump’ codes together rather than ‘split’ them as far as possible. Where more precision is required this can be obtained from the patient’s hospital records.

A 5 character alphanumeric code was chosen in preference to an entirely numeric code because it is much easier to use and can be easily remembered.

Obviously a pathological diagnosis will have a higher confidence level than a clinical or even an angiographic or echocardiographic diagnosis. The true cardiac morphologist specialised in congenital cardiac defects may consider our coding and classification system a ‘laundry list’. We would disagree with this opinion in that it can still be very useful and perhaps even more so to the clinical cardiologist in a busy practice! The fact is that there are many more paediatric cardiology clinicians than congenital cardiac morphologists.

The clinician needs to review his clinical data. He must assess the morbidity and mortality of the patients in his care. The results of cardiac surgery must be regularly reviewed so that an audit of the surgeon’s results is available. In this manner the cardiologist is able to give the parents a true reflection of the child’s prognosis and also the results of a particular cardiac surgical center. Perhaps the results of a particular operation are worse in the hands of surgeon X than of surgeon Y. These facts should be readily available, and this is the case with the data base and coding system presently being used at our hospital.

The real time data base which we have developed allows us to have all the



clinical, echocardiographic and angiographic data available on any one patient. We can immediately see which operations the child has had, and what complications developed. Cardiac catheterisation reports as well as echocardiography (both M-mode and two-dimensional) reports are readily available and printed from the data base.

Our classification does not in any way detract from the complex coding and classification systems previously developed by the morphologists. Their classifications are obviously far more accurate and concise. For epidemiological purposes, and for the study of congenital cardiac defects, these 'morphological' classifications will contribute vital information.

Our system provides valuable local information and data but is unlikely to contribute in any major form to the world body of scientific knowledge as far as congenital cardiac defects are concerned.

## 5.2. *Standardisation*

Eventual national and international standardisation of coding systems in paediatric cardiology is important for progress in this field. The problem arises that already different coding systems have been developed on either side of the Atlantic Ocean, as well as in the Northern and Southern hemispheres, and probably on different continents. There are also the 'splitters' versus the 'lumpers'.

From a practical point of view there will probably never be international consensus. Paediatric cardiac centers should at least obtain a national consensus. Those countries with an abundance of paediatric cardiologists and cardiac morphologists will probably favour a more complex, accurate system. In countries where cardiologists are in short supply a more clinical coding system is likely to be more appropriate.

## 5.3. *The Red Cross War Memorial Children's Hospital Paediatric Coding system*

The system is subdivided into three groups:

1. Clinical, Echo and Cath
2. Operations, Procedures
3. Complications

Figure 1 shows the first page of the clinical, echo and cath codes. There are 323 codes in total: 186 clinical, echo and cath codes, 76 operations and procedures codes and 61 complication codes.

### 5.3.1. *Clinical, Echo and Cardiac catheterisation diagnoses are further subdivided into:*

5.3.1.1. Segmental sequential analysis: Codes have been devised to describe the entire segmental analysis as described by Anderson *et al* [8]. For example:

Fig. 1. The first page of clinical, echo and cath codes, listed in alphabetical order by description of the code.

---

CLINICAL, ECHO AND CATH CODES

A-V CONNX MODE: COMMON VALVE .....	CAVV..
A-V CONNX MODE: OVER-RIDING L VALVE .....	QAVVL
A-V CONNX MODE: OVER-RIDING R VALVE .....	QAVVR
A-V CONNX MODE: STRADDLING L VALVE .....	SAVVL
A-V CONNX MODE: STRADDLING R VALVE .....	SAVVR
A-V CONNX: BIVENTRICULAR AMBIGUOUS .....	BIAVC
A-V CONNX: BIVENTRICULAR CONCORDANT .....	BCAVC
A-V CONNX: BIVENTRICULAR DISCORDANT .....	BDAVC
A-V CONNX: UNIVENTRICULAR, ABSENT L .....	AAVCL
A-V CONNX: UNIVENTRICULAR, ABSENT R .....	AAVCR
A-V CONNX: UNIVENTRICULAR, DOUBLE INLET TO INDET V .....	UDIVI.
A-V CONNX: UNIVENTRICULAR, DOUBLE INLET TO LV .....	UDIVL
A-V CONNX: UNIVENTRICULAR, DOUBLE INLET TO RV .....	UDIVR
ABSENT PULMONARY VALVE .....	PVABS
ACUTE RHEUMATIC FEVER .....	ARE...
ANEURYSM: CORONARY ARTERY .....	ANCQR
ANEURYSM: PATENT DUCTUS ARTERIOSUS .....	ANPDA
ANEURYSM: AORTA .....	ANAO.
ANEURYSM: ATRIAL SEPTUM .....	ANASD
ANEURYSM: LV .....	ANLV..
ANEURYSM: PA PATCH .....	ANPAT
ANEURYSM: PULMONARY ARTERY .....	ANPA.
ANEURYSM: REPAIR .....	ANR...
ANEURYSM: RUPTURED SINUS OF VALSALVA ANEURYSM .....	ANSVR
ANEURYSM: RV .....	ANRV..
ANEURYSM: SINUS OF VALSALVA .....	ANSV..
ANEURYSM: UNSPECIFIED .....	ANOTH
ANOMALOUS L PULMONARY ARTERY: REPAIR .....	ALPAR
ANOMALOUS LEFT CORONARY ARTERY .....	ALCA.
ANOMALOUS LEFT PULMONARY ARTERY .....	ALPA.
ANOMALOUS PULMONARY VENOUS RETURN: CARDIAC .....	TAPVC
ANOMALOUS PULMONARY VENOUS RETURN: INFRACARDIAC .....	TAPVI
ANOMALOUS PULMONARY VENOUS RETURN: MIXED .....	TAPVM
ANOMALOUS PULMONARY VENOUS RETURN: SUPRACARDIAC .....	TAPVS
ANOMALOUS R PULMONARY ARTERY: REPAIR .....	ARPAR
ANOMALOUS R SUBCLAVIAN ARTERY: REPAIR .....	ARSAR
ANOMALOUS RIGHT SUBCLAVIAN ARTERY .....	ARSA..
ANOMALOUS SYSTEMIC VEINS: AZYGOUS-HEMIAZYGOUS .....	ASVAZ
ANOMALOUS SYSTEMIC VEINS: CORONARY SINUS .....	ASVCS
ANOMALOUS SYSTEMIC VEINS: IVC .....	ASVIV
ANOMALOUS SYSTEMIC VEINS: SVC .....	ASVSV
ANOMALOUS RIGHT PULMONARY ARTERY .....	ARPA.
AORTIC ARCH ANOMALY: COARCTATION .....	AAAC.
AORTIC ARCH ANOMALY: DOUBLE .....	AAAD.
AORTIC ARCH ANOMALY: HYPOPLASTIC .....	AAAH
AORTIC ARCH ANOMALY: INTERRUPTION .....	AAAI..
AORTIC ARCH ANOMALY: RIGHT-SIDED .....	AAAR.
AORTIC INCOMPETENCE: CONGENITAL .....	AVIC..
AORTIC INCOMPETENCE: INFECTIVE .....	AVII..
AORTIC INCOMPETENCE: PROSTHETIC .....	AVIP..
AORTIC INCOMPETENCE: RHEUMATIC .....	AVIR..

---

Situs inversus is coded SITI,  
 AV connection: univentricular, absent right is coded AVVCR,  
 VA connection: discordant is coded DVAC.

Similarly relationships are also coded for example

Aortic relations: anterior and to the right is coded AOAR

Right ventricle lies left, anterior and inferior is coded RVLAI.

Diagnoses regarding sequential segmental analysis or morphological relations are made either on clinical grounds or at echocardiography or at cardiac catheterisation.

5.3.1.2. Anatomical diagnoses are coded as close to clinical abbreviations as possible. For example a ventricular septal defect is coded clinically according to the size of the left to right shunt or degree of pulmonary hypertension (Shrire *et al* [9]) as VSD1, VSD2 or VSD3. At echocardiography or cardiac catheterisation a further anatomical diagnosis is made depending on the site of the defect for example a perimembranous VSD is VSDP or a muscular inlet VSD is VSDMI. A secundum atrial septal defect is coded ASD2. Certain anatomical diagnoses are not made clinically and the diagnosis can only be made by the special investigation, for example right ventricular tumour is coded RVT. In other situations the diagnostic code is entirely clinical, for example INN is innocent murmur.

5.3.1.3. Haemodynamic diagnoses are similarly coded according to clinical impression or echocardiographic or cardiac catheterisation diagnosis. A child who clinically has mitral incompetence is coded as MVIR if the mitral incompetence is thought to be due to rheumatic heart disease or MVIC if the incompetence is congenital in origin. It is coded as MVII if infective endocarditis is the etiology of the mitral incompetence on clinical as well as echocardiographic factors.

5.3.1.4. Generalised or systemic conditions are also coded in the data base if these conditions affect the cardiovascular system, for example Mucocutaneous Lymph Node Syndrome or Kawasaki disease is coded MCLNS and Takayashi Arteritis is coded ARTTA. Hyperlipidaemia is coded LIPID.

5.3.2. *Operations and Procedures.* Operations and procedures have also been coded using a maximum of 5 letters with abbreviations closely related to the clinical setting. Some examples are as follows;

Fontan repair: FONT

Modified Blalock-Taussig Shunt: MBTS

Pulmonary Artery Banding: PAB

End to End Coarctation Repair: COEE

Balloon Valvuloplasty: Pulmonary Valve: ANGPV.

After the initial coding system was developed it was found to be necessary to add some new codes for new operations and procedures. Extensive additions and modifications are not however advocated.

*5.3.3. Complications.* Complications following cardiac surgery or procedures such as balloon valvuloplasty can be monitored by coding for specific incidents. This allows the clinician an up to date assessment of the morbidity (and mortality) following surgery and to analyse trends regarding surgery, infections etc. An accurate documentation of arrhythmias, CNS lesions, respiratory complications such as stridor (STRID) or phrenic nerve palsy (PHRN) and so on can be made. Trends can be evaluated and also long term retrospective analysis performed. Once again complications are coded according to fairly major groups for example renal dialysis and/or renal failure is 'lumped' together as 'RENAL'; bowel or liver complications are grouped under the code 'GIT'. Further analysis must then be done manually. Similarly for infection codes, the organism is not coded – this information has to be retrieved from the patient notes or possibly from the microbiology laboratory records.

## **6. The computer solution**

### *6.1. The hardware*

The system runs on a Local Area Network (LAN) in order to satisfy the requirement that it be multiuser. A LAN was chosen in preference to a multitasking system. Reasons for this included the fact that at the time our system was being developed multitasking PCs were expensive and required operating systems other than DOS. A change of operating system and software did not outweigh the perceived advantages of a multitasking system. We continue to believe that distributed data processing is preferable to centralised processing. This allows each end user to use his terminal as a stand-alone PC, but can at any time enter the network to access the 'corporate' data base.

Novell Advanced Netware 286 was selected as the network software, and ARCNET as the hardware. The choice of Novell was based on its reputation as a reliable, fast network, and also the existence of local expertise. It is important to choose a network and software that are compatible. The software we use, dBase III plus (Ashton Tate), specifies that it can be run on Novell networks. We have had the experience of writing software for a relatively unknown network in another hospital which would not run on their LAN.

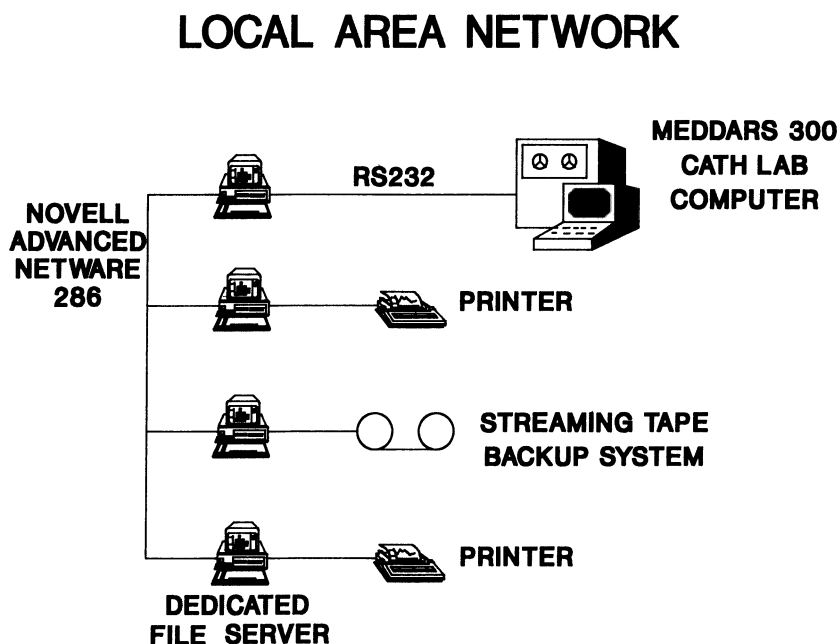
The three terminals used in our system are IBM compatible PCs. As long as they are standard IBM compatible machines the choice of PC should be based on the local supply company. A reputable company with a good maintenance record and fast turn-around times should be selected.

We elected to use our file server in a dedicated mode, thereby allowing maximum speed and protection of data on the network. The file server is a

Sperry IT with 3MB of RAM and a 40 Mb hard disk. It has an EGA colour monitor.

A PPG-Hellige Meddars 300 cardiac catheterisation monitoring computer is used in our catheterisation laboratory. This is equipped with a serial port and Kermit (Columbia University Center for Computing Activities, New York) communications software to download ASCII files. All data from catheterisation, including all derived data such as calculated flows, shunts and resistances, can be downloaded via the serial interface to a PC. PPG-Hellige do provide a 'workstation' system which allows downloading of data to a PC. However this system is orientated towards adult catheterisation with coronary angiography, and was not suitable for our use. We therefore used the download facility and in-house written interface software to incorporate the down-loaded data into our system. The hardware for this connection uses the RS232 port on the Meddars and a similar port on one of the terminals on the LAN.

Figure 2 is a diagrammatic representation of our system.



*Fig. 2.* A diagrammatic representation of the local area network, showing the file server, 3 terminals, peripherals and the RS232 interface with the Meddars catheterisation computer.

### 6.2. The software

The software was written by ourselves (KJS) over a period of about one year. It has been modified, debugged, and cleaned up over the last few years and has now been in full operation since October 1985.

The software chosen for our application was dBase III plus (Ashton Tate). We considered a fourth generation data base management language to be preferable to other high level programming languages for reasons which include:

- i) The language can be used in the interactive mode for those who know and can use dBase III. This allows ad hoc access to the data for special searches and research.
- ii) It can also be used in the programmed mode in which programs are written to make the system entirely menu driven. This is used for the routine use of the system.
- iii) dBase III plus is a procedural language with built in tasks such as creating data bases, relating data bases, searching, indexing, and other routine tasks required for creating and maintaining a data base.

The programs have been written in a modular form with as much independence of each module as possible. This allows for additions and alterations to the system without having to make too many changes to other modules.

The system was not devised to replace the written patient notes, but to keep a summary of patient information. In this way cross-referencing and research is possible. It was also devised to facilitate and to speed up the production of cardiac catheterisation and echocardiographic reports.

The system is set up as a relational data base with different files being linked to one another by key fields. All the rules pertaining to the construction of data bases have been incorporated. For example there is only one copy of any particular field in the data base, except for the key fields, which link together various related files. The patient name and other demographic details only occur once in the data base in the demographic file. The hospital identification, an 8 digit number, is used to link this file with other files requiring patient identification. In the catheterisation suite of data base files there is a further level of linking. The demographic file is linked to a basic cath file by the hospital number. This file then has a cath number field which links it to all the other files in the cath suite. See Figure 3 for an illustration of the interrelation of files.

Editing of data is thus simplified. Only one file is updated for each change, except when altering key fields. These fields should seldom be changed, but the software attends to the alteration of this field in all linked files should it become necessary.

There are no repeating fields in any particular file. Thus the file which contains the diagnostic codes, has only 3 fields in each record: the patient hospital number, the code and the date of entry of the record. Any patient requiring the entry of more than one code, and this includes most patients, requires one record for each code entry.

### *6.3. The basic suites available on the system*

*6.3.1. The demographic suite.* This allows initial registration of the patient on the system and input of demographic data. Duplication is prevented, and validation of birth date is performed (for example birth date cannot be after the entry date).

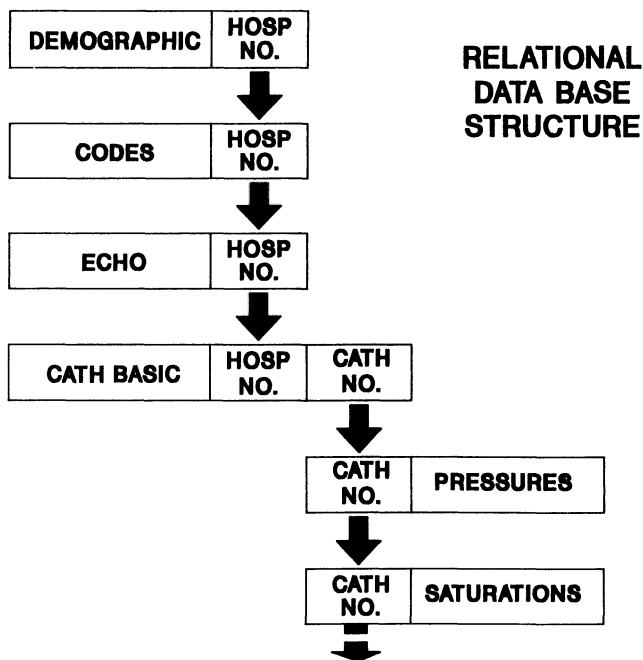


Fig. 3. A diagram indicating the linking of some of the files in the relational data base by hospital number. The cath basic data file is further linked to the other catheterisation files by the cath number.

This data can be edited or searched. For example, a patient can be found given any one piece information about that patient, for example surname, hospital number, or even just birth date.

**6.3.2. The catheterisation suite.** This is a collection of 9 data base files containing basic data (for example cath number, date, operator, weight, height), procedure narrative, pressure, oxygen data, output calculations, other calculations, angiographic report, comments and conclusions. All the data is automatically downloaded via the serial interface from the Meddars computer except for the angiographic report and the conclusions. These are keyboarded in by the physician on a pre-prepared 'mask'.

The system allows for editing, deleting, printing and searching. For example, given a patient's hospital number, or surname, a search can be performed which lists the catheterisations performed on that patient. These can be selected and printed out on a fast dot-matrix printer.

**6.3.3. The echocardiographic suite.** This allows input of both 2-dimensional (2D) and M-mode echocardiograms. The 2D echo is reported by the physician on a 'tick-sheet' (see Figure 4). This is keyboarded into the computer using a numeric input. Some free text is allowed for comment. An echo report is generated by the software from the input.

The M-mode section allows keyboarding of the data derived from the

Fig. 4. The tick sheet used by the clinical cardiologist to report the two-dimensional echocardiogram. This is subsequently keyboarded by the technologist, using simple numeric input screens.

	Solitus		IVC to		DATE:		WARD
Situs	Inversus		SVC to		Normal		None
	Isomeric		Pulm V to	RA	Small	LA	PFO
	Uncertain				Dilated		ASD
					Common		Secundum
					Isomeric		Primum
					Membrane		
A-V connexion	BIVENTRICULAR		Concordant			A-V connexion	2 A-V valves
	UNIVENTRICULAR		Discordant				Imperforate L or R
			Ambiguous				Common
			Absent L or R				Straddling
			Double inlet				Override % LV or RV
			Normal				
			Stenosed		Normal		
			Incompetent		Hypoplastic		
TRICUSPID VALVE	Cleft	MITRAL VALVE			Dilated		Present
	Ebstein's		RV		Hypertrophied	LV	IV
	Parachute				Function normal		Function normal
	Prolapse				Function impaired		Function impaired
	Prosthetic						
			Concordant				None
V-A connexion			Discordant				Perimembranous
			Absent PA	V-A connexion	2 V-A valves		Muscular
			Absent AO		Imperforate L or R		Subarterial
			Double outlet		Override PA % LV	VSD	Inlet
					Override AO %RV		Trabecular
							Outlet
R VA Junction	Tretalogy		Discrete SAS		PULMONARY VALVE	Normal	AORTIC VALVE
	Inf obstruction	L VA Junction	Septal obstruction			Stenosed	
	Other obstruction		Other obstruction			Incompetent	
	Lies A or P		Lies A or P			Dysplastic/Bicuspid	
	Lies L or R		Lies L or R			Prolapse	
						Prosthetic	
MPA	Normal	AORTA	Not assessed			L or R	
	Hypoplastic		Normal			Double	
	Dilated					Coarctation	
						Interruption	
RPA	Normal	LPA	RCA Anomalous	LCA	AORTIC ARCH	PDA	
	Hypoplastic		Aneurysm			Truncus	
	Dilated		Fistula			AP window	
	Anomalous						
COMMENT:							
CODES:							

measurements. Reference values are automatically calculated from the patient's body surface area and abnormal values are flagged. Echo reports can be edited, deleted, printed or searched.

6.3.4. *The diagnostic code suite.* This allows the entry, editing, deleting or searching of diagnostic codes, operation codes or complication codes. This facility has been extremely useful in research and in audit of the department.



**6.3.5. The rheumatic fever suite.** Rheumatic fever and rheumatic heart disease is still seen relatively commonly at our hospital, and constitutes one of our major research interests. This suite of programs and data bases captures data relating to clinic visits of all children with rheumatic heart disease. In addition data on all children who have prosthetic valves inserted is entered into the data base.

## **7. Some observations and considerations**

### ***7.1. Data capture methods: on-line vs keyboarding***

Keyboard data capture should be minimised as this is an important source of invalid data. Our system therefore uses automatic data trapping at source as much as possible. The data interface with the cardiac catheterisation computer in our environment allows down-loading of data directly into the data base.

### ***7.2. Data completeness – methods of ensuring trapping all data***

Unless data is complete, it is not useful. Completeness is usually very difficult to ascertain. The flow of data in the system and the boundaries of data extension should be carefully assessed. Procedures then need to be adopted which preclude any data not being trapped and captured. This is a human-based activity and as such, it is one of the biggest weaknesses in computerisation of a previously manual system.

Manual checking should be done regularly in an attempt to determine the completeness of the data. For example, the computer generated figures for a particular operation, complication or code should be compared to records obtained from other sources.

### ***7.3. Data validation***

All data must be computer validated and regular manual checks should be performed to ensure data validity. Dates are easy to validate, but other figures such as height, weight, catheterisation pressures, saturations, outputs and others must also be checked from time to time. Validation is essential and much of the program coding should revolve around this activity.

The data base is only as useful as the accuracy of the data.

### ***7.4. Data security and privacy***

Medical data is confidential and as such must be secure from unauthorised access. Restricted access to the computer and a password system are essential. Medical data is also important and therefore all data must be backed up on a reliable medium on a regular basis.

Our security uses a password to gain access to the network. Once on the

network the user has access to the patient data base. Accidental or malicious corruption of the data base is avoided by the use of 'access' codes. In order to edit or delete data the user has to enter the correct access code. All requests for a deletion are prompted a second time and the user has to enter 'S' to indicate that he is sure.

The data is backed up every day onto a 60 Mb streaming tape. Every third day the entire hard disk is backed up onto the tape.

### *7.5. Implementation*

The change over from a manual to a computerised system must be carefully planned. There must be staff involvement from the very first concept of the system. In order for the system to be acceptable to the staff they must be given an opportunity to contribute towards its development. Suggestions from them must be seriously considered and implemented wherever possible. If they are allowed to feel that it is 'their' system it is much more likely to succeed, than if it is forced on them.

Ideally the system should be phased in gradually. The implementation must be planned to take over from the existing manual system, without loss of data, or confusion. A decision must be made about capturing data retrospectively.

Manual methods should be retained until it is certain that the computer system is totally acceptable and operating without major bugs. This may take some years.

*7.5.1. Training.* Careful thought must be given to staff training and motivation. Most staff members are frightened of computers and extensive education is necessary in order for the system to be successful.

### *7.6. Specific problems*

We now have 4½ years experience of using this system in our unit. To date there are 4755 patients recorded in the data base, with 16590 codes. There are 1276 cardiac catheterisations, 3699 echocardiograms, 2495 visits by patients with rheumatic heart disease, and records of the implantation of 221 mechanical heart valves.

There have been some problems, most of which were never foreseen. The most important problems are discussed here as a guide to others who may be contemplating setting up a similar data base system.

*7.6.1. Non-capture of data.* As has been discussed earlier, it is often extremely difficult to ensure that data is complete. Unless some procedure exists to ensure that, for example, all postoperative complications are entered for all patients, it is possible that important codes are never entered. It is our belief that this can be minimised by being cognizant of keeping the data complete, and entering codes as soon as the condition arises. This can only be effective if there is a network

terminal readily accessible, in other words, in the Intensive Care Unit. Alternatively all codes must be recorded on a card attached to the patient's bed-letter as they arise. These are later captured, before the patient is discharged.

*7.6.2. Definitions of the codes.* Precise definitions and use of the codes are essential. For example, in our system, repair of tetralogy of fallot is coded as RVPAT (RVOFT patch without septal repair) and VSDR (VSD: patch repair). This must be used consistently. Another user entering RVRES (RVOFT resection without septal repair), in cases in which an outflow patch has been used, would effectively corrupt the data base.

Sometimes alternative codes exist and care must be taken that the usage is consistent. For example a dilated cardiomyopathy could be coded as LVCMO (LV dysfunction as a 'complication') or EMD (endomyocardial disease, a 'diagnosis'). A policy regarding usage must be established and adhered to.

Some codes must only be used in certain circumstances. For example the code PAB (pulmonary artery banding) is an operation code and is only entered for a patient once when the banding is performed. At a subsequent cardiac catheterisation, when the band is noted, the code PAB must not be re-entered as part of the catheterisation diagnosis, as this would imply that the patient was re-banded on that date.

*7.6.3. Retrospective data entry.* It is important to make a policy regarding retrospective data entry when implementing the system. The retrospective entry of codes for pre-existing patients is the best option if this can be easily achieved. However this is usually impractical and codes are usually entered when the patient next presents. If this is the procedure followed it must be remembered that the date entered for the code will not necessarily be the date of first diagnosis of the pathology, and searches for incidences of that pathology during a particular time interval may not be accurate.

*7.6.4. Interfacing with other computers or equipment.* Traditionally interfacing different computers and hardware can be problematic, both from the hardware and the software point of view. In our system the interface with the Meddars 300 cardiac catheterisation computer and the downloading of the data became a problem after an update to the software on the Meddars. The manufacturer failed to supply us with an updated version of the download file structure. As a result, data is occasionally downloaded which is unintelligible to the data base interface software.

*7.6.5. Patient duplication.* Despite diligent attention to this problem it is sometimes not detected. In our hospital a central mainframe computer maintains records of all patients seen at any of three major hospitals in the Cape Peninsula. This central hospital system allocates hospital numbers to new patients. If this system fails to detect duplication then patients may be entered into our data base with more than one hospital number.

*7.6.6. The 'garbage-in-garbage-out' syndrome.* It is often extremely easy to punch a few keys on the computer and obtain a long print-out loaded with information. The ignorant user often assumes that 'the computer is correct' and that the information so obtained is complete and accurate. The user may even use this information for policy making or for publication, unaware that it is incorrect.

This anomaly might easily occur if the user is not absolutely certain about what data the computer software is extracting when it searches the data base. For example the search program might have been written to extract all occurrences of a particular code, for example MBTS (modified Blalock-Taussig shunt). If some patients have had more than one shunt, they would have more than one occurrence of this code. A search for all codes MBTS would then count the number of shunts performed and not the number of patients who have had shunts.

All data extracted by the computer must be carefully evaluated as to its accuracy, and whenever possible a manual check done on a sample of the data. All users must be aware of exactly what the search programs do, and how the data is obtained. A careful check and audit must be kept on all data extracted from the data base.

The user would also be unaware that this extracted information is incorrect because of incomplete data entry. For example if the complications INFP (pulmonary infection) is omitted from the data base for some patients in error, then a search on the incidence of pulmonary infection would give erroneous results. The user extracting this information would be unaware of the error.

## **8. Conclusions**

This computerised data base has revolutionised the reporting of cardiac catheterisations and echocardiograms. The results of catheterisations are now available the same day as the procedure, and are easier to produce.

It has enabled us for the first time to extract information from the experience of the cardiac unit, and to audit the results of the unit. It has resulted in research and publications [10] not previously possible.

We do not believe that in the 1990s a busy cardiac unit can function adequately without such a computerised data base.

## **9. Acknowledgement**

We wish to acknowledge the important contribution made by Dr Derek Human (present address: Izaak Walton Killam Hospital for Children, Halifax, Nova Scotia) to the design and development of the coding system. He was also instrumental in the initiation of the project.

## References

1. Lodewyk HS van Mierop (1986) Diagnostic Code for Congenital Heart Disease, Supplement. *Pediatr Cardiol* 7: 31-34.
2. Miller GAH, Anderson RH, Rigby ML (1985) *The Diagnosis of Congenital Heart Disease, incorporating the Brompton Hospital Diagnostic Code*. Castle House Publications Ltd
3. Weinberg PM (1986) Systematic approach to diagnosis and coding of pediatric cardiac disease. *Pediatr Cardiol* 7: 35-48
4. Cooper IC, Carey C, Crowther A, Webb-Peploe MM (1988) A Microcomputer Multi-User Relational Database in a Department of Cardiology. In *Medical Applications of Microcomputers*, Coleridge Smith PD, Scurr JH (eds), Springer-Verlag, London
5. Brower RW, Ten Katen HJ, Bar FWHM, Koster RW, Meester GT (1987) The Role of the Personal Computer in the Cardiac Catheterisation Laboratory: an integrated approach to information management. *Comput Programs Biomed* 24: 87-96
6. Shepard RB, Blum RI (1986) Cardiology office computer use: primer, pointers, pitfalls. *J Am Coll Cardiol* 8: 933-40
7. Morin JE, Symes JF, Ralphs-Thibodeau S *et al.* (1986) A microcomputer system for the practising surgeon. *Can J Surg* 29: 364-66
8. Anderson RH, Becker AE, Freedom RM *et al.* (1984) Sequential segmental analysis of congenital heart disease. *Pediatr Cardiol* 5: 281-288
9. Shrire V, Vogelpoel L, Beck W, Nellen M, Swanepoel A (1965) Ventricular septal defect – the clinical spectrum. *Brit Heart J* 27: 813
10. Walker K, Sprenger KJ, De Moor MMA (1989) Long term penicillin in Sydenhams Chorea (letter). *S Afr Med J* 76: 582

## **PART FOUR**

### **Departmental applications**

# Patient documentation for the ultrasound laboratory

RÜDIGER BRENNECKE, FRANK SCHÖN, RAIMUND ERBEL

*2. Medical Clinic, Johannes Gutenberg – University, Mainz, West Germany*

## Summary

The development of a PC based system for the storage of patient data in the echocardiography laboratory is described. Special design objectives were low cost and user friendliness, including an integrated special query language. After 3 years of use, the system stores 80000 diagnoses from 12000 patients without any relevant slowdown of response. Future developments concerning the communication between the database system and an echocardiographic sector scanner are discussed.

## Introduction

Following the rapid development of personal computer hardware and software it has been realized, that the documentation of patient data can be performed by small, decentralized computer systems. Here we report on a documentation system for the ultrasonic laboratory, that helps in the management of patient data. More specifically the following fields are covered by the system:

- a) Most patient examinations have to be scheduled on account of telephone calls to the echo laboratory. Here, the instant availability of all previous patient information is of high value.
- b) If all important data on the results of previous examinations are stored in computer files, the retrieval for clinical evaluation of these results can be much easier in a computerized data base system than it is in an organization based on a conventional documentation on paper.
- c) For scientific reports, one often needs a list of cases with a certain group of echocardiographic findings. Such a list can be e.g. the basis for a detailed evaluation of all patient data including a review of the videotape recordings. A search for these patient groups based on the conventional documentation methods is extremely time consuming and error prone.
- d) It is practically impossible to provide a detailed periodic survey on the statistics of certain diagnostic findings, if the search is to be performed in

conventional paper based documents.

Software systems that solve the problems listed above can be implemented in very different computer languages and with different man- machine interfaces [1]. The system described here has been in practical use for more than three years [2]. During this time it has been refined several times. Therefore we can report here on experience with a small data base system that has already been tested by daily use over a relatively long period of time and by people without previous computer knowledge.

### **Hardware components**

The computer used is a standard personal computer IBM PC/AT. The operating system is MS-DOS. The computer system is equipped with 512 KByte main memory, a floppy drive with 1.2 MByte and a hard disc (20 MByte).

### **Database software**

The primary program has been written using the relational database system dBase III Plus from Ashton Tate. In the meantime, it has been transferred to the Clipper compiler. In this way, we could largely increase the speed of data entry and retrieval.

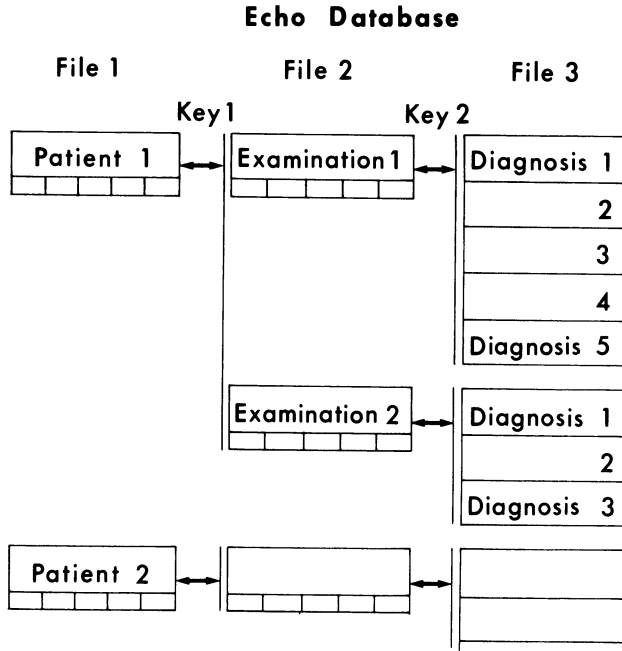
### **Application software**

At the basis of the relational organization of the data structure, three database files (Figure 1) are involved:

1. Patient data file. This file contains one record for each patient.
2. Examination file. Here we find one record for each examination of a patient. This file contains e.g. the date of the examination, the name of the doctor carrying out the examination and the number of the video tape where the images acquired during this examination are stored.
3. Diagnosis file. Here, the diagnosis for all patients and all examinations are stored. Each record of this file contains just one diagnosis. However, each patient has generally many records stored in the diagnosis file.

The three data base files mentioned above are connected according to the relational model by using keys (Figure 1). For instance, each diagnosis record contains a key pointing to a certain examination record. This examination record, in turn, has a pointer (key) to the data of a patient. These keys make it possible, that for each patient (first file) the examination parameters (second file) can be found and these again be coupled with the diagnosis (third file). The diagnosis is entered into the computer using codes consisting of a letter and two or three numbers. The codes are contained in a catalogue storing at present about





*Fig. 1.* The structure of the ultrasound data base. Three data base files are needed to store the data using the relational data base scheme.

600 diagnoses. It can be easily updated using a special option of the main menu.

During data entry, the full name of the diagnosis is visible on a pull down menu (Figure 2). Each diagnosis selected from the menu is translated by the computer into the short code mentioned in order to prevent errors in the coding process. Similarly, when the diagnoses for one patient are retrieved, the codes stored in the system are translated back to the full name of the diagnosis for better readability.

## Modules of the program

### *Patient data*

This part of the program implements the steps of retrieval of data of patients that have been stored earlier and the entry of new patient data. In each case, the name and the data of birth can be entered. Different types of wildcards are allowed to simplify this search. The system then shows either one patient, a list of patients with similar data or no patient. In the latter case, the new patient data has to be entered. For patients with earlier examinations, the most important data from the examination file are displayed. Then, the data for the new examination are entered. In a similar way, patient data and examination data can also be edited.

There is another way to access the diagnosis data. If the number of the

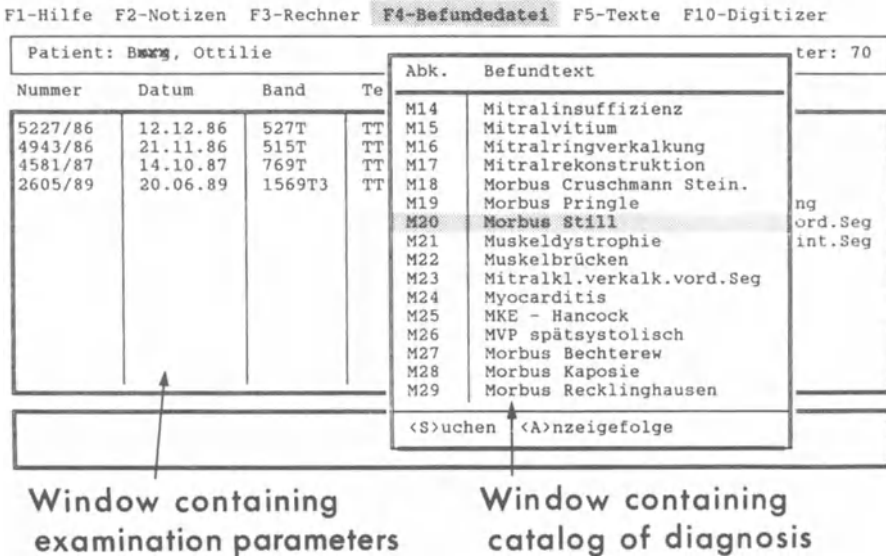


Fig. 2. Data entry screen. Entering the codes of diagnosis is aided by the contents of the window showing a catalog.

examination is known, one can directly address all data referring to this examination by entering the examination code.

*Data retrieval for clinical purposes*

For routine clinical use it is important to have a quick overview on all data previously acquired for a given patient. Therefore, it is possible to enter the name and date of birth of a patient, and to retrieve all information on previous examinations (file 2) and diagnosis (file 3). Figure 3 gives an example of the display of data from one patient. In the upper part of the screen, the demoscopic or patient data are shown (file 1). In the central region of the screen, we see at the left a list of all previous examinations. In this case, we have 4 examinations. At the right side, the central part of the screen shows all diagnosis from *one* examination. The examination parameters are highlighted in the left part of the central screen. By moving this cursor over this area (examinations), one switches the right part of the screen between the lists of diagnosis obtained during the respective examinations. The lower part of the screen shows options for further steps.

*Statistical evaluation*

While the retrieval options described above are centered on the patient or the examination, scientific evaluations are generally centered on combinations of diagnosis. For this purpose, we use a special retrieval routine. This program

**Patient data (file 1)**

F1-Hilfe F2-Notizen F3-Rechner F4-Befunddatei F5-Texte F10-Digitizer

Patient: **Baumg, Otilie** Geb.: 01.06.88 Alter: 70

Nummer	Datum	Band	Tech.	Abk.	Kl.	Befund
5227/86	12.12.86	527T	TTE/D	M14	3	Mitralinsuffizienz
4943/86	21.11.86	515T	TTE/D	L18	1	LV - Funktion
4581/87	14.10.87	769T	TTE/D	L11	0	LV - Hypertrophie
2605/89	20.06.89	1569T3	TTE/CD	A34	3	Aortenstenose
				M16	0	Mitralringverkalkung
				M23	0	MitralKl.verkalk.vord.Seg
				M30	0	MitralKl.verkalk.hint.Seg
				P06	0	Foto(Farbdia)

<N>eue Untersuchung <B>earbeiten  
Druck<p>rotokoll <I>nfo zur Untersuchung <L>öschen

Window containing  
examination data  
(file 2)

Window containing  
diagnosis  
(file 3)

Fig. 3. Overview on the data for a given patient is provided by a simultaneous display of 3 windows.

makes it possible to refine the parameters of the retrieval in an interactive way. It can also be used by people that are not familiar with computer systems.

The program provides 2 screens for the definition of the parameters of the retrieval. The first mask allows for the selection of parameters or parameter ranges. The second screen allows to refine the contents of the report. Thus, in the second selection step, we define the names of the fields of the database that have to be printed for the record defined in the first step.

The first mask can be considered as a sort of a form, where one fills out higher and lower limits for the most important parameters. The four last lines allow one to enter codes of four diagnosis. A 'filter' for the data can be formed by entering operators (<, =, >) and values into the different field provided on the mask before each parameter. In a second step, logical operators (AND, OR) are entered to connect these fields (last line). In this way, the filter condition for the database system is generated interactively.

After selecting the records from the three database files according to the filter condition, the resulting 'filtered' file can be stored on disc or selected data from this new file can be printed. A second screen, already mentioned above, allows to define the names of the database field to be printed in a report.

### *Archival storage*

After more than three years of permanent use, over 80000 diagnosis from over 20000 examinations in 12000 patients are stored on the hard disc of the simple

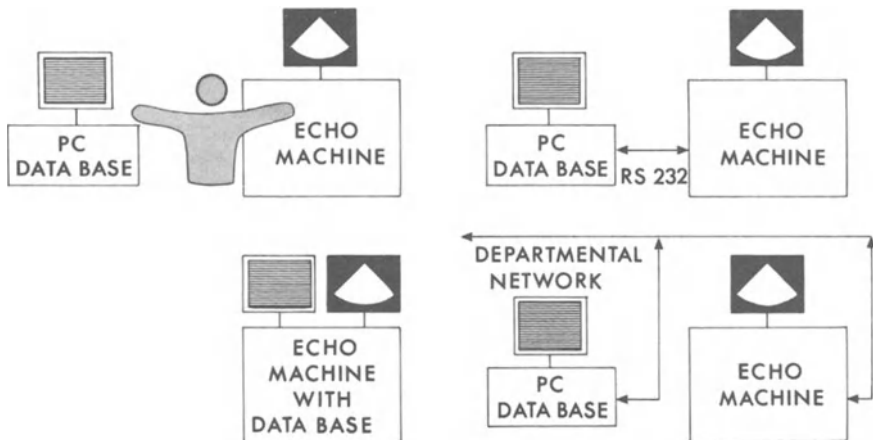
personal computer system described above. We have not noticed an essential increase in access time when searching for patient data. There has not yet been the necessity to transfer some of these data to a separate archival storage medium. Probably, later on data of patients that have not been seen for more than 5 years will be transferred to tape and all but the demoscopic data would then be deleted from the data base. Backup storage is still on floppy disc (3 diskettes for all relevant data).

*Future developments*

In the near future we will also acquire numerical data that are measured during the examination such as enddiastolic volume, ejection fraction and others. Fields for the storage of numerical values are being added to the structure of the examination file (Figure 1).

An additional extension should in the future allow for the incorporation into the data base of images. In a first stage, it should be possible to transfer images from the echo system to computer disc. All data overlays including the outlines of regions (for instance the LV contour) should be stored. The main question still to be discussed is, if it is necessary to store cine loops in addition to still frames. Even if the storage of single frames is sufficient, the inclusion of images into the data stored means a significant increase in the amount of data to be stored and in the complexity of the programs to manage these data. One of the aspects of such a system is the fact, that not all data of patients that go back for instance 2 years can be held in permanent storage.

Another possibility for future development is to transfer numerical data acquired during evaluation of the echocardiogram in the echo machine to the database computer (Figure 4). Presently, these data have to be read from the



*Fig. 4.* Four ways of transferring alphanumeric data between an echo machine and a data base system.

display of the echo machine and then to be typed into a mask of the data base. Some of the present machines, however, allow already for the transfer of data over serial link (RS 232). In this way, the manual transfer of data from the echo machine to the database and the errors occurring in this process can be avoided. However, in the long run, this link has to be bidirectional. In this way, the patient data that are normally entered using the keyboard of the echo machine could be entered into the database computer, then transferred to the echo machine. In reverse, the numerical data acquired by the machine could be transferred back to the personal computer at the end of the examination.

It is clear, however, that the relatively large effort required to implement this data transfer will only be worthwhile if a certain amount of standardization between the data formats of the machines of different manufacturers has been achieved. A different approach to the goal of integrating the data base system and the echo machine is to include the data base in the machine [3] (Figure 4). The final stage may be a departmental network that connects echo machines, PC based data entry stations and a larger central data base as shown at the lower right in Figure 4.

## **Conclusion**

A low cost data base system installed on a standard PC is sufficient to help efficiently in the management of patient data in the echocardiography laboratory, even if the data of about 5000 patients have to be entered per year. The user interface was changed several times in order to optimize user friendliness. The resulting structure of the data base and its user interface seems to be easily adaptable to many other applications such as the catheterization laboratory. Further progress requires one of the described interfaces between the data base system and the echocardiographic sector scanner.

## **Acknowledgement**

This work was supported in part by the Deutsche Forschungsgemeinschaft.

## **References**

1. Ott WJ (1986) The design and implementation of a computer-based ultrasound data system. *J Ultrasound Med* 5: 25-32
2. Jung D, Brennecke R, Erbel R, Meyer J (1986) Dokumentationssystem für das Ultraschall-Labor. *Ultraschall Klin Prax* 1: 49-51
3. Hagelund JE, Schipper KP (1989) The Echo Base. University of Trondheim, Department Biomedical Engineering. Internal report.

# A research-oriented database management system for Holter data

PIERFRANCO RAVIZZA

*Department of Cardiology, General Hospital, Lecco, Italy*

## Summary

Computerized management of clinical data may be very helpful for physicians: in particular, personal computers allow safe and friendly use for processing of biological signals and alphanumeric data in clinical practice.

Holter monitoring consists of a continuous recording of electrocardiographic signals during 24 hours. The analysis of the recording gives rise to a report of normal and arrhythmic events. Classification of the events is automatically determined by means of a dedicated algorithm for wave recognition and only visual control of the procedure is required by the operator.

Counts, minute and hourly rates of arrhythmic events, length and frequency of major arrhythmias (mainly tachycardias), allow precise classification of the recording, in order to establish its prognostic value in clinical practice.

Personal computers may be helpful in:

- storage and retrieval of the reports;
- joining of general and clinical information to the rough data;
- automatic calculation of parameters derived from rough data, produced by the analyzer;
- classification of the recording with a prognostic score based on the quality and quantity of arrhythmic events (Lown's class).

When correct recognition of the events is ascertained by the supervision of a skilled operator, all the subsequent passages may be automatically performed by the computer. From the clinical point of view, three main advantages may be underlined:

- high computational speed of the computer;
- correct prognostic classification, by means of an automatic algorithm;
- storage and retrieval of data with simple procedures.

Moreover, collected data may be processed, beyond standard clinical analysis, with complex algorithms for research purposes. Also such procedures may be automatically performed by the computer in a very short time and allow fast collection of a lot of data, without supplementary charge of work for the operator.

In conclusion, computerized database management systems for Holter data on personal computers warrant safe and fast collection of clinical parameters, both for standard and research purposes, with very limited medical supervision.

## **1. Introduction**

No doubt still persists about the usefulness of computers for the management of data in medical practice [1, 2]. Physicians use a great deal of diagnostic tools, increasing in complexity because of powerful hardware. Due to extended times of analysis, many clinical parameters are considered. For all these reasons, medical procedures are greatly dependent on computerized machineries.

The new diagnostic devices can usually be linked to an external micro- or mini-computer. In other cases, such an informatic support is contained just inside the machine. Anyway, the medical tools usually allow transfer and analysis of data by means of computers.

Moreover, medical operators show an increasing interest in personal computers, particularly because of their friendly use: in fact, general purpose software can easily be managed by 'ordinary' operators and is now widely used in clinical practice.

Finally, medical practice is based on a very complex and time consuming activity: computerized management of clinical data saves physicians a lot of time, by performing those problems that can easily be reduced to a well defined algorithm [3, 4].

A department of cardiology is usually equipped with a great number of diagnostic tools, based on the analysis of mono- or two-dimensional signals. Each of these devices gives rise to a great amount of data, that cannot be easily analyzed by direct reading. Personal computers allow a powerful solution for a large part of this problem.

Computerized applications usually deal with a great amount of data that can be analyzed by well defined algorithms. Holter monitoring seems a very good example of these conditions, because of the great quantity of events recorded, their simple classification, and the great clinical usefulness of their prognostic score (Lown's classification).

The aim of this paper is to summarize the characters of a database management system for Holter data, based on general purpose hardware and software.

A dedicated system for collection, storage and analysis of standard clinical data was developed in our laboratory. Moreover, special attention was paid to the automatic collection of research-oriented parameters.

## **2. Holter method**

Holter monitoring consists of twenty four hours electrocardiographic

recording, obtained by means of a portable device, in order to collect data on cardiac rhythm, during normal daily activity, in patients with cardiac disorders. The recording device has an internal clock that allows timing of the events. At least two electrocardiographic leads are recorded, in order to make their recognition unequivocal.

The recording is firstly analyzed by means of a dedicated device. Analogic-to-digital conversion of the signal is a necessary prerequisite. The sampling rate must be sufficiently high (usually 100 Hz or more) to allow a precise definition of the 'QRS complex' of the electrocardiogram. Threshold adjustments of the internal algorithm must be adapted case by case, in order to make optimal the recognition of normal and abnormal beats, and of intervals between each couple of beats, if shorter or longer than normal.

The counts of these events can be obtained both by full automatic analysis or by direct reading. The two systems are generally used together, in order to join the speed of the automatic analysis and the safety of the physician's control. The resultant procedure is quite free of errors in the recognition of the events and very complete about their quantization, in spite of a time just a little longer than that required for the automatic analysis.

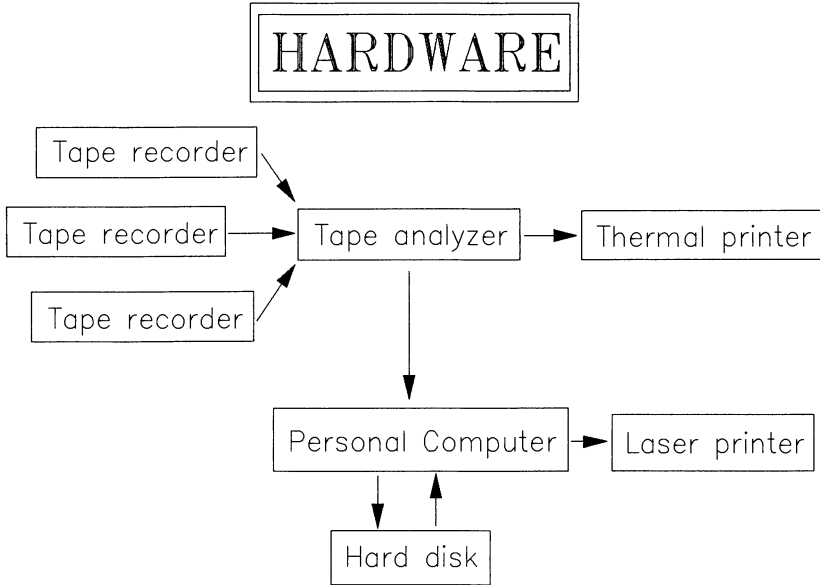
The report of the counts of normal and arrhythmic events, at fixed time intervals (usually hour by hour or shorter), can be automatically produced and printed.

### **3. General characters of the database management system**

The report of the events can be transferred to a personal computer, linked to the analyzer (Figure 1), and processed with statistical algorithms, in order to increase the completeness of the results and fully disclose the clinical potential of Holter monitoring. A dedicated program, designed and realized in the laboratory, allows the following functions:

- a) storage of the rough report of arrhythmic events;
- b) storage of a number of general, clinical and electrocardiographic data, in order to better qualify the patient's clinical background and the type of major arrhythmias (namely ventricular tachycardias);
- c) calculation of numeric parameters, derived from the report of the events (counts, minute and hourly rates);
- d) automatic definition of the severity of the arrhythmias, according to the well known Lown's classification (calculated from data of points a, b, c);
- e) automatic storage and printing of the complete report of clinical and Holter data;
- f) statistical analysis of stored data, both for clinical and research purposes;
- 9) easy management of directories and data files.





*Fig. 1.* The structure of the system is schematically represented. Data from patients are collected by means of a number of analogic tape recorders. Counts of the arrhythmic events are automatically performed by the tape analyzer and can be printed, with significant electrocardiographic strips. The reports of counts are then transferred to a personal computer, through a serial port. These rough data are joined on the computer with other clinical and electrocardiographic parameters. Final calculations and a comprehensive report are then created and printed by a laser device.

### 3.1. Choice of hardware

The power of microcomputers has increased so much in the last few years, that their use has become advantageous both for the management of alphanumeric data and of mono-dimensional signals. In fact, personal computers based on the microprocessors of the Intel Corporation (80286, 80386 and 80486) or of the Motorola Company (the '68000' family) became in some way competitive with some minicomputers, still appearing as complex machines, far less user-friendly and usually requiring a system specialist. On the contrary, personal computers are now based on widespread, well documented and easy to manage hardware, so that their use can be governed by unskilled operators.

### 3.2. Choice of software

This point deserves similar observations. Although a third generation language, like 'C' or 'Pascal', allows complete programmability, a lot of general purpose fourth generation languages, usually oriented to the database management (with various extensions towards other fields of data management, like spread-sheets or word-processing) may be user-friendly, in spite of only a short training.

The system analyzed was built by means of the dBase III (Ashton-Tate Inc.) [5-7], coupled with the Clipper compiler (Nantucket Co.) [8], but could certainly be created with a lot of similar database management systems, now available in the area of general purpose programs for personal computers. However, we want to underline that this couple of programs allows a very effective treatment of alphanumerical data, although it does not completely match the requisites for being considered a relational database.

In fact, our system warrants a number of functions such as:

- a) specific database oriented commands;
- b) friendly use;
- c) standard connectivity towards other working areas, by generation of data files in the standard ASCII code;
- d) widespread availability on the majority of personal computers used as host;
- e) complete and satisfying documentation, due to the great diffusion of the languages.

All these aspects allow physicians to build database management systems well adapted to their own needs, without depending on specialized operators, rarely present in general hospitals of the area of public health care, in this country.

#### **4. Structure of the database management system**

Two different flow-charts were identified, relating to the needs of a 'single purpose' or of a 'multi-purpose' system. The single-purpose system is only focused on the collection of data from Holter monitoring, and all clinical and general patient's data, necessary to appropriately define the clinical ground, are directly inserted into a unique database.

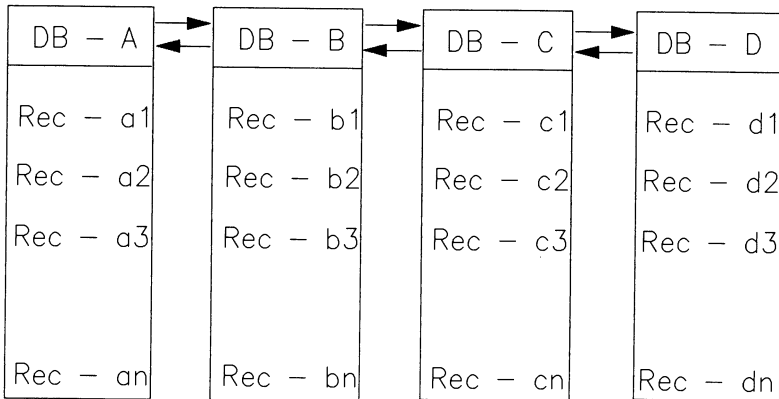
In this case, a 'parallel' structure (Figure 2) can be simply used, in which every new examination is related to its own set of general and clinical data, stored in one or more files, even when some of those general data have already been introduced, because of previous examinations of the same patient. This organization brings some degree of redundancy, but makes their analysis easier, particularly when it will be performed in the interactive mode on the computer.

On the other hand, in a multi-task system, general and clinical data comprise information used to identify the patient and his clinical background, in respect to a number of diagnostic procedures (for example echocardiography, stress test, catheterization procedures).

In this latter case, a hierarchical structure (Figure 3) is preferable, in which some general or clinical information can be shared, by a number of 'special' records, containing different diagnostic sets of data. From this point of view, it is not convenient to replicate data in every 'special' section of the database, but groups of data of general interest must be collected in common files, and all the diagnostic databases can easily be related to them.

### PARALLEL STRUCTURE

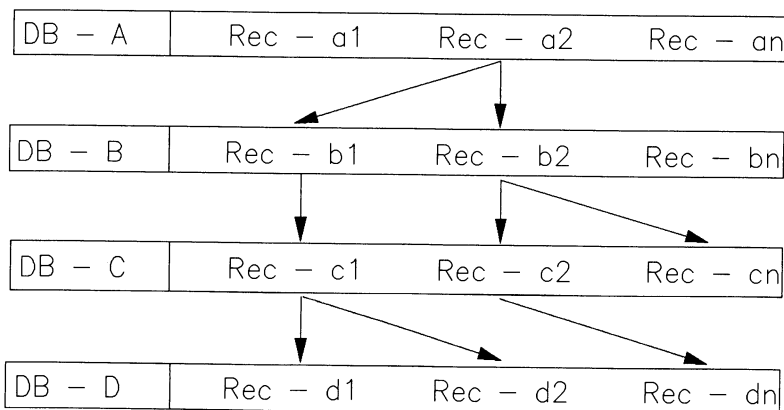
One-to-one relationship between the records



*Fig. 2.* The so called parallel structure of the DBMS consists of a number of files of data, each with the same number of records. Information about one tape is collected in all the records with the same record number, through the parallel files. Conversely, each file comprises only a partial sequence of data about all the cases of the DBMS. Such a kind of structure gives rise to some redundancy of data, but makes very easy all the retrieval functions along the system.

### HIERARCHICAL STRUCTURE

One-to-many relationship between the records



*Fig. 3.* The so called hierarchical structure of the DBMS consists of a number of files with a different number of records. Each file can comprise a partial sequence of data about all cases, but no duplication is allowed, so that one particular set of data may be related to more than one set in another file. On the contrary, no information about that case may be contained in another different file. This structure is highly efficient in sparing memory, but creates some difficulty in using the query functions of the database by unskilled operators.

At present, a single purpose version of the database management system is routinely used in the laboratory, but a more complete multi-purpose release is under testing for future application in clinical practice.

## **5. General rules for software development**

A number of rules have been identified, that seem necessary to allow a safe and user-friendly management of the database in clinical practice. Some of these structural principles are dependent on the general theory of databases, while others are more directly derived from the particular medical application. We are going to analyze these topics by a general point of view: in fact, the practical way of application is dependent from the specific language used to build the database management system. We do not think it important to show the details of the system, but it seems necessary to have these conditions satisfied, in order to allow a safe use of the database by unskilled operators (that is to say also technicians and nurses).

### *5.1. Check against unintentional doubling of data*

This is an important source of problems in computerized systems. In fact, each record can be linked to many others, by means of indexed codes. The use of indexed fields to relate database files is the core of the relational structure, and consists in building 'fast' connections between different files, containing data about the same topic. Using this artifact, it is very simple to link, for example, data of all the examinations performed during a hospitalization period to a unique record of the 'admission' database or to a unique record of the patient's general information database. Unfortunately, the codes used to build such relations are usually complex (consisting of quite casual sequences of digits and letters), so that they are very difficult to remember, and represent a great source of error in data collection. One suggested solution consists in using an automatically determined progressive numeric code (that in the case of the previously described parallel structure should simply be the record number) as the main relational field: this trick allows effective reduction of the indexed relations and accelerates the procedures of localization of the records along the database.

Automatic characterization of each record by means of a progressive number is certainly a more simple solution, in respect to the use of complex codes, like the date of birth or the hospitalization code. On the other hand, if these complex codes have to be used as indexing fields, a facilitating choice should consist in implementing procedures of selection of data with the so called 'pop-up' or 'bar' menus, that assure a simple choice of complex data, without writing them more than once on the keyboard.

Moreover, a very crucial point in a lot of database management systems is the possibility to choose between the entry of a new record and the review of an old

one. Unfortunately, this procedure is not easily protected against errors, such as input into a new record of something that has been previously inserted, or doubling of an old record, because of the misspelling of a name or of a code. One trick to avoid, as far as possible, unintentional doubling of records consists in using a unique procedure, both to insert new data and to review the old ones. The search for a record always begins from the name of the patients, that is usually the best known available data. The specification of the first name or of only a significant part of it, activates listing of all the cases that match the condition, in which the searched case, if already present, could be easily recognized and selected, avoiding erroneous opening of a new record for an old case.

### *5.2. Pre-defined coding of the descriptive data*

A lot of qualitative definitions are frequently used in medical practice, that is to say data not easily reducible to a precise numerical measurement. Coding of the qualitative aspect of the data is a necessary step for making their analysis user-friendly. In this field, a lot of solutions have been proposed, but no one can be suggested as completely satisfying and each programmer spends a lot of time in searching for the more practical one. The solution here proposed consists of coding of the content of the fields, with triplets of letters or short words, previously chosen as univocal markers of the events and possibly well related to the word represented. The names of pharmacological compounds, for example, some aspects of the diagnostic procedures, some events of the patient's history and a lot of similar conditions, provide a good outlet for the use of such a method of coding.

The main advantage seems to be the readability from the point of view of the medical operator: the use of pre-defined short, but easily readable, codes represents a way of storing data in a more comprehensible and functional way, in comparison to the rough numerical one, whose uncoding is rarely immediate.

The storage of qualitative data in such a short pre-coded form, requires just a little more time from the operator (and usually more memory on the hard disk), in respect to the numerical coding, but warrants an easier availability of data, in the retrieval procedures. Moreover, all major available languages, dedicated to database function, allow procedures for windowing of the screen and for inserting lists of available codes for each field that has to be coded. This procedure avoids using external lists of codes (on paper for example) that usually bring a high error rate by the operator.

### *5.3. Automatic transfer of data and calculation of derived parameters whenever possible*

Doubling of data in different files of the database management system must be automatically done. This statement seems to be controversial from the theoretical point of view, because very little doubling of data could actually be allowed in a true relational database management system. In fact, a system built

in respect of such a severe rule, is frequently difficult to analyze from the interactive point of view. But the interactive analysis (from the 'dot' presentation) is a very useful way to check data in clinical ground, particularly when frequent control is necessary.

The right compromise seems to provide the minimal doubling of data sufficient to make possible fast interactive analysis of the contents: transfer of data must be performed automatically, strictly avoiding double input of data and possible secondary errors.

Direct input of data should be limited to the so called 'minimal set': in fact, data collection can comprehend both essential measurements and an often larger number of derived parameters, calculated from one or more basic measurements. A very general principle is that every derived parameter should not be directly stored, but derived when needed, using the powerful computational speed of the microprocessor. Nevertheless, when the number of derived parameters is so high and the time for its calculation becomes cumbersome, it could be convenient to store almost selected sets of these calculated data in a permanent form. This is particularly true when derived data are obtained from sets of measures, stored in large files outside the database, whose continuous storage on the mass memory could be very 'byte-expensive' and/or time consuming. In this case, it should be convenient to summarize derived data in one or more related databases for at least two reasons: first, significant data are collected in a structure of little dimensions, that can be easily managed, by means of the query functions of the database (on the contrary, large amounts of data are usually stored in 'text' files, that cannot really be treated with a friendly approach). Second, this procedure allows discharge of more large quantity of rough data from the mass memory of the computer: the capacity of the hard disk should not really be a problem nowadays, because of the availability of large magnetical or magneto-optical supports, anyway, the larger the space of memory used, the heavier becomes the problem of time-by-time saving such amounts of data.

In this sense, it is recommended to collect the significant derived data in a database form and free the storage memory from the large files of rough data, whose analysis can be limited to the first approach to the case. In fact, data stored in large 'text' files may be easily transferred into a database form and more and more simply treated and analyzed for obtaining derived parameters, and then can be newly back-transferred, modified or not, to 'text' form. Anyway, this procedure is rarely requested, other than during the first examination of the data, and does not justify the maintenance of the original data on the hard disk. On the contrary, all the sets of data that are frequently retrieved for review or for new calculations must be carefully kept in an easy to handle database form, for every day needs.

In our Holter system the rough data consist of the counts of the electrocardiographic events, collected time-by-time during the analysis of the recording. Mean minute or hour rates of premature ventricular contractions, couplets or triplets are automatically calculated. More complex statistical

procedures obtain parameters related to the variability of normal and abnormal beats in the 24 hours or in definite subsets of time. All these data represent the set of derived data and are permanently stored into the database management system for further retrieval, printing and analysis.

#### *5.4. Full disclosure of the data*

The operator must be able to retrieve the stored data at every moment with a simple procedure. This rule is very important in order to make fully acceptable (and desired indeed) the computerized system. In fact, one of the more surprising characteristics of the computers is the availability of the results in a very short time, in spite of the great amount of data treated and of the complexity of calculations required to obtain them. Nevertheless, it is not easy to obtain this result in database management systems with strictly relational structure, unless an effective retrieval procedure is available: each programmer must clearly keep in mind the wishes of the users and realize all the tricks, in order to warrant a satisfying accessibility of the data. On the contrary, a too frequent need for treatment of the data by skilled operators is considered a source of problems by the ordinary users, for a lot of reasons such as intermittent availability of the analysis, the relatively long time necessary for it and something like a feeling of impotence in respect to the computer. So, it is very important to provide the system with various degrees of retrieval facilities, by which the operator could verify all the data, in order both to check the status of the system and to directly obtain immediate results from his personal fatigue.

In fact, in a strictly relational architecture, the analysis of the data can be very difficult and time wasting, unless they have been previously collected into a single database. Predefined automatic or semiautomatic analysis cannot completely satisfy some complex needs of the clinical users of the system, if the interactive approach to the database cannot be used. On the other hand, the interactive mode of analysis allows a very complete processing of the data, but may require a deep knowledge of the host language, that is rarely present in all the users of the system. A good compromise seems to be that of supplying different degrees of analysis: first of all, the program must warrant to all the users the possibility to retrieve all the data in a record-by-record sequence; second, a semiautomatic query system must allow selected listing or statistical analysis of data, based on the more common needs: this step should be made as easy as possible by the programmer, in order to make successfully available the procedure to each user, even in spite of little comprehension of the structure of the database; finally, a free interactive mode of analysis must always be allowed for complete research, but considering its potentially high degree of complexity and the low protection towards loss of data, it must be delegated to an expert user. In fact, it is important to remember that general purpose database management systems contain a lot of powerful query facilities that can be used with great satisfaction, but cannot easily be used by unskilled operators.

Our Holter system was equipped with a number of statistical facilities,

dedicated to the automatic or semiautomatic analysis of large amounts of data, when previously grouped in temporary databases. The use of such procedures seemed very advantageous almost as a 'first step' statistical approach to clinical data.

### *5.5. Friendly use of the DBMS for on-line collection of data*

We have already discussed at length this aspect of the interaction between the operator and the machine. In fact, personal computers can now be equipped with very easy to handle software, in comparison to that available in the working area of minicomputers or mainframes. Nevertheless, the main characteristic of such systems is that they must be used by untrained or 'quickly' trained operators, with a range varying from physicians very interested in informatics, to people with a very low interest in such procedures. Anyway, it is clear that the actual success of informatics in medicine, as in all other fields, is based on its widespread use, as a new and substantially more convenient way of treating data.

An effective use of database management systems greatly depends on the direct collection of data, just at the time they are produced, eliminating the time-wasting and error-creating two step procedure of first producing data and only later inserting them into the computer, perhaps by different operators. Such a procedure may be disappointing for the responsible of the latter passage and does not give a satisfying and immediate feedback of results to the data producer.

The best way to obtain interest in personal computers seems to be to demonstrate that they can be very useful, in an 'on-line' way, with an immediate consistent result. This is the reason why every computerized system must be used in an interactive mode, during the day-to-day activity, and, most of all, with the production of minor or significant, but consistent, results in terms of computation, reporting, improvement of work and so on. This is the only effective way to motivate users in spending the supplementary time necessary to store data into the computer (that is often a little bit longer than the minimal amount necessary to pursue the standard 'on paper' collection). Such a result does not appear very easy to obtain, but made more simple if the developer of the system is a member of the medical staff or is very familiar with it.

In fact, we spent a lot of time in developing easy to handle and safe procedures for input, retrieval, printing and listing of data, taking in great evidence the need of the operators in their day-to-day use of the system. As a result, almost all the above mentioned functions can be accessed by each authorized operator in no more than a few seconds, even in very large database files.

### *5.6. Help in accessing the operating system commands*

A secondary, but important part of the facilities available in the system is dedicated to help the operator in managing a few fundamental commands of the



operating system of the computer. In fact, the access to these functions from the host database management system is quite unusual. Nevertheless, use of some DOS commands may be necessary, when the program requests to handle external files, imported from or exported to other environments. Fortunately, each general purpose package usually enables management of such functions from inside the program, and these DOS commands can easily be inserted in the compiled code, just like any other function of the database management system. Guided procedures have consequently been installed, in our system, in order to transfer portions of data to external worlds, such as packages dedicated to special statistical processing or to graphical presentation of data, i.e., the two fields of which the host database seemed quite lacking. These functions may require some degree of accomodation, in relation to the size of the dataset and to the program they are directed to, but fortunately these procedures are only seldom employed in the standard use of the database management system.

## **6. Compliance of the users towards the system**

Two physicians and two nurses have been using the database for Holter data for about one year in our laboratory. The system was built with respect to the principles reported above. It seems important to show some general results of the interaction between medical operators and the system, in order to better exemplify the rules exposed.

### *6.1. Use of the DBMS*

Nurses and physicians could use the system after a very short training period (only a few days were necessary for the informatic support, provided that they had good experience in electrocardiographic monitoring). In fact, the input of data and their analysis are fully guided and controlled with great care, particularly towards doubling of data. Some experience is needed of the use of the keyboard and its general functions. A lot of minor problems were solved day-to-day in the first weeks of use of the database, by means of a very close interaction between the programmer and the ordinary users, more directly involved in the input of data. A facilitating point was that the author is a physician, particularly familiar with the management of clinical data. This condition probably prolonged the solution of some problems on the informatic side, but, in contrast, brought satisfying results on the clinical side. This condition is very paradigmatic and reflects the reality of our country, in which only a few institutions have already realized the actual need for people expert in informatics being involved in the work of clinical units, in the area of public health care. This approach could certainly produce some disadvantages in terms of time spent in solving the problems. Nevertheless, many centers developed their own experience in the computerized management of clinical data, guided by specialists in informatics, but they did not always obtain satisfying results

(possibly because of the poor comprehension of the computer functions by the physicians).

### *6.2. Doubling of data*

When the program is first built, it is not easy to anticipate what errors could be performed in the use of the system by unskilled operators. We experienced the method of controlling the identifying code of the case, whenever input or retrieval of data must be performed, in order to avoid incorrect insertion of data. Such a procedure makes it quite impossible to mislead the previous input of the actual data. The experience with such a checking system is very satisfying to date, both because of the short time requested to verify data (due to the great efficiency of the management of indexed databases), and of the rare interventions needed for the correction of mistakes, mainly related to misspelling of the examination codes. Intervention rate of the system programmer was however no greater than one over about one hundred cases inserted and it seems very low for a 'self made' system.

## **7. Clinical applications**

The following aspects of the system must be first underscored, due to the prominent clinical function of the database on a cardiological unit, mainly dedicated to delivery of social health care.

### *7.1. Automatic calculations*

Calculations performed over the report of events, created by the tape analyzer, have a very high degree of accuracy. Some problems could arise in the treatment of reports from recordings with a high degree of noise. In this case, the report might contain a lot of false recognitions of events and their correction could be required, by means of a dedicated procedure, before processing for derived calculations. Experience with the first examinations identified the commonest errors due to the noise of the tape, usually leading to numeric overflow in some steps of calculation. At the present time, however, no more problems are appearing, provided that tapes with large amounts of noise are excluded from the analysis.

### *7.2. Automatic calculation of Lown's class*

One particular aspect of the analysis of the tape is the quantization of the arrhythmias according to the well known classification proposed by Lown [9] and to its modified Italian version [10]. Previous storage of all qualitative parameters and of a number of calculations needed for such classification, made simple its automatic calculation, presentation for check and final insertion into

the database. This seemed an important facility, especially as it was to be the system managed by the nurses: in fact the calculation of Lown's class and of the related score [10] is quite difficult for untrained people. The automatic calculation of these parameters was always correct and safe, and reduced the need for checking of the final report by the physician. On the other hand, the calculation is a very simple challenge for the computer, being based on a sequence of very precise data, both directly inserted or calculated from Holter data.

### *7.3. High quality of clinical report*

The computerized Holter system allows a significant improvement of the quality of the final report. The tape analyzer is able to create a rough report of the counts of the events, but these counts cannot be modified, no Lown's class can be automatically calculated and a high degree of interaction is required to give a complete report of the examination. Using our system the data can be analyzed in a more complete way, corrections are allowed at each level and, finally, a more complete representation of the report is performed, by means of a good quality laser printer. All these facilities bring a more efficient method of data presentation and increase the quality of the work of the laboratory.

### *7.4. Listing of the data*

Full disclosure of data allows retrieval of all the cases at each time, both for checking their correctness and for reviewing information. Pre-programmed listing of selected data was also allowed. Both selective and non-selective listing are available and every user can easily perform such procedures in a user-friendly way. Moreover, a few statistical reports are comprised into the standard application, in order to make always possible some sets of calculations of frequent interest about the stored data. The use of such facilities by the ordinary operators was very satisfying and limited the intervention of the programmer only to some rare special interactive treatment of data.

## **8. Research applications**

This paper does not seem the appropriate place to analyze in detail the research data resulting from the use of our system [11, 12]. Furthermore, we want to underline some general applications of the database management system, that seem very important in terms of efficiency and availability.

### *8.1. Special analysis of data*

The above mentioned retrieval facilities are related to the day-to-day clinical use of the database. Nevertheless, the system was built both for ordinary clinical

practice and for research purposes. In fact, Holter data represent a very good challenge for informatics, due to the great amount of events considered for each examination and to the possibility of precisely defining these events. However, it could be quite 'expensive' to handle data from a large number of cases, to obtain, for example, well selected results for congress presentation. This is a procedure not easily available to the majority of the users. In fact, the complete analysis of the data requires thorough knowledge of all the functions of the host language, in order to avoid unintentional deletion or alteration of data, and it is consequently only performed by an expert operator. From this point of view, autonomous self-building of the system is certainly a great advantage in terms of flexibility and adaptability to the physician's needs.

Moreover, public health care institutions are charged by a very large amount of work, so that it would be quite impossible to conveniently manage all the data created by the new diagnostic techniques without informatics. An intensive use of the computers allows a satisfying control of the data and an interesting fall-back in terms of paper publishing or congress participation.

### *8.2. Calculation power*

The computational power of a microcomputer allows the physician to manage the great majority of the alphanumeric data considered in his medical practice, with surprising speed and usually great satisfaction. When Holter data are analyzed by a computerized system, a lot of parameters can be easily examined and compared, new mathematical approaches can be tested, and, finally, complex statistical analysis can be easily performed. These possibilities allow the physicians to become confident with the diagnostic tool, to expand their knowledge of its clinical application and to select the more significant data to be used for research purposes.

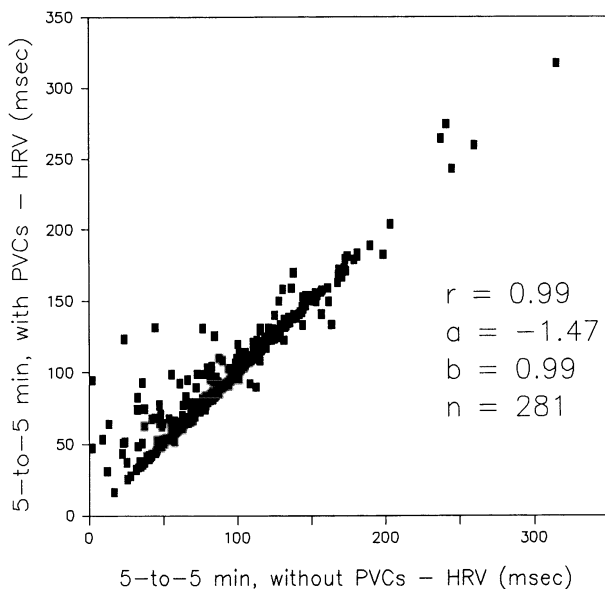
This way, computers can solve, at least in part, the problem of the increase of clinical data. We work every day with new machines, able to produce large amounts of data, but sometimes we do not have enough time and space to store and analyze all of them. A well oriented use of personal computers can help the physicians to solve this problem, avoiding a useless collection of data and allowing their correct analysis.

### *8.3. Example of application*

A specific example of such great flexibility is the calculation of some complex parameters related to the heart rate variability during Holter recordings. Based on the observation that the standard deviation of the intervals measured between each couple of subsequent beats, during Holter monitoring, is a good marker of sympathovagal interaction on the cardio-circulatory system, great attention was paid to the analysis of such a parameter. Moreover, a lot of other related parameters could be analyzed to check their relative clinical significance. The calculation criteria [11] are not fully standardized and were first established at

# HRV – METHOD

(5-to-5 min, with vs without PVCs)



*Fig. 4.* Heart rate variability is usually calculated from the mean values of R-R interval, excluding those obtained in presence of more than one premature ventricular contraction per minute. Calculations could be repeated for some hundreds of cases, including also the values related to periods with every number of premature ventricular contractions. A dedicated program allowed a quick recalculation of all cases, during less than two hours of automatic work on the computer. In this way, the two methods could be easily compared, with great satisfaction for medical staff, in spite of a very short computational time.

the beginning of data collection. Based on the ‘on-going’ clinical experience, a further review of those criteria was performed, and a complete recalculation of the previously analyzed examinations was requested. Computerized management of the database allowed a very fast and satisfying performance in this direction, that could not have been easily approached without such support. In fact, a parallel database was created for the new set of data, some hundreds of reports were examined again, all the derived parameters were calculated according to the upgraded criteria and results stored in the new database. Statistical correlation was then verified by means of a dedicated procedure (Figure 4). The computer could automatically perform these recalculations by means of a quite simple routine, in a very short time (about two hours for more than 500 cases, about 120 counts per case). Similar results could only be obtained with days or weeks of work in the pre-computer era. This kind of approach allows a very simple analysis of data and increases the potential of clinical research, particularly considering the small amount of time available to physicians during clinical practice, at least in general hospitals.

## 9. Conclusions

Holter data are a very good challenge for computerized applications in clinical medicine. In fact, a lot of data are usually considered during the analysis of each examination, and only the use of computers allowed this new diagnostic tool to be introduced. Moreover, it soon became clear that the true potential of Holter monitoring could have been disclosed only by complex treatment of data by means of computers, in order to avoid misinterpretation of significant data. The use of a database management system for the handling of clinical information could effectively complete the computerized analysis of the Holter monitoring.

### *9.1. Interfacing between hardware supports*

A lot of programs allow efficient management of data, provided that they are stored into the hard disk of a computer. Nevertheless, it is not always easy to obtain the storage of data in such a manageable support. In fact, only in recent years have the producers of medical instrumentation clearly realized the importance of giving to the physicians a complete and easy to handle way of managing the data, deriving from diagnostic tools. At the present time, a number of problems regarding the interfacement of the diagnostic machines with the storing computers have not yet been satisfyingly solved. Only a good standard for the transfer of data could warrant a really open and user-friendly computerized system. Unfortunately, the competition between the producers of medical devices is a great obstacle to a clear solution of such problems.

### *9.2. Use of general purpose database*

This is one of the more satisfying aspects of the system. In fact, in recent years, a lot of general purpose databases became available in the area of personal computers, needing a low training and giving a high degree of standardization. These possibilities could be improved by the latest advances in the use of standardized languages (like the Structured Query Language). Everyone should be able to handle data into his own personal computer, with a speed and a degree of complexity, that could have never been imagined until a few years ago.

### *9.3. Clinical applications*

The use of such systems in the medical area allowed physicians to obtain a clear improvement of clinical work. First of all, the computerized analysis can be performed by a large number of operators, easily obtaining a lot of complex calculations in a safe way. Reduction of complexity of the procedures seems a very important point.

A second one is the reduced need for checking of computational procedures.

The use of computers allows complete control over computational sequences: they can be performed with great safety, even by nurses or technicians, without necessary intervention of physicians.

Finally, the use of computers improves the appearance of the clinical reports, that is a secondary, but equally important side of medical procedures.

#### *9.4. Research advantages*

Some of the previous considerations clarify how and why clinical research can be better performed by means of computerized systems. In fact, such management of data shortens the time needed for statistical analysis and their graphical presentation, that are considered a worldwide standard for scientific production.

Moreover, it became clear in recent years that great trials represent the only way to really solve a lot of clinical questions. Even in this perspective, computers can actually help in collecting data from a lot of clinical units and make easier the performance of such large studies. Unfortunately, something has yet to be done to make completely safe the use of personal computers and to introduce standardized ways for data transfer and storage.

However, our experience in the treatment of Holter data by means of a computerized database is very satisfying and encourages further pursuit of such a method of managing medical data, both for clinical and research purposes.

#### **Acknowledgments**

I wish to thank the nurses Mrs. Paola Secomandi and Mrs. Lorenza Dell'Oro and particularly my colleague Camillo Gerosa for the very careful attention paid in testing the system.

#### **References**

1. Milazzotto F, Martiny W, Matta F, Pincirolì F (1987) Banche di dati integrate in cardiologia. Centro Scientifico Torinese, Torino
2. Ravizza P, Addamiano P, Gerosa C, Knippel M (1983) Going towards informatics into a cardiology department: a four years experience with self programmed tools. In Banche di Dati Integrate in Cardiologia, Centro Scientifico Editore, Torino, p. 3
3. Ravizza P, Dacquino G, Divieti L, Pissarello M (1983) Left ventricular performance evaluation with computer aided infrared chest thermography in left ventricular hypertrophy. In Proceedings of Computer in Cardiology 1983. IEEE Computer Society, Silver Spring, p 445
4. Dacquino G, Ravizza P (1988) Introduzione alla elaborazione di immagini. In AIIM '89, Franco Angeli (ed), Milano, p 70
5. Hergert D (1988) dBase III Plus. Microsoft Press, Redmont, Washington
6. Robbins J, Braly K (1987) Expert dBase III Plus. Sybex, San Francisco
7. Tsu-der-Chou G (1987) dBase III Plus Tips, Tricks and Traps. Que Corporation, Indianapolis
8. Straley SJ (1988) Programming in Clipper. Addison-Wesley Publishing Company, New York

9. Lown B, Wolf M (1971) Approaches to sudden death from coronary artery disease. *Circulation* 44: 130
10. Brusoni B, Sanna G, Bettini R, Fonda F, Pagnoni F, Santoro GM, Sartori F (1983) Aritmie ipercinetiche ventricolari maligne. Criteri classificativi: revisione della classificazione di Lown. Studio policentrico. *G Ital Cardiol* 13: 253
11. Knippel M, Ravizza P, Gerosa C, Bana G, Locatelli V, Mauri L (1989) Heart rate variability and ventricular tachycardia. 4th European Symposium on Cardiac Pacing, Stockholm, May 28-31, p 141
12. Ravizza P, Gerosa C, Locatelli V, Bana G, Ferrari A, Mauri L, Knippel M (1990) Heart rate variability and ventricular arrhythmias. *New trends in Arrhythmias*, 6(1/2): 655



# A database management system of coronary care unit data

MARCO TUBARO, CARLO IACOANGELI, ENRICO NATALE,  
ROBERTO RICCI, FILIPPO MILAZZOTTO

*Coronary Care Unit, S. Camillo Hospital, Rome, Italy*

## **Abstract**

The development of a database management system is an important task to improve decision making in a medical environment.

Main goals of our system are: establishment of a computerized database; generation of a discharge form by the computer; statistical analysis of coronary care unit data; and achievement of an informatic support to decision making. The initial version of our programme utilizes memory variables to store data, speeding up analysis of patient records. A newer, recently developed, release is based on a completely coded system: advantages are ease of use by inexperienced operators combined with great flexibility and velocity of data retrieval expediting.

Patient management and clinical research are immensely supported by computerized database systems of this kind.

## **Introduction**

Management of the clinical data of a modern Coronary Care Unit (CCU) is a complex task of the utmost importance as regards both patient care and scientific research. The creation of a database management system (DBMS) is expedient to the development of information technology, informatics and decision making in the cardiologic field. The intention is to increase the efficacy of the physician's knowledge and to improve clinical research and patient care.

The aim of our work was to develop a DBMS of the clinical data of patients admitted to our CCU, with the following objectives:

- establishment of a computerized database from medical records of CCU patients;
- creation of a discharge form, automatically generated by the computer, which could be attached to the official medical record;
- possibility of creating general statistical data of CCU patients;
- achievement of an informatic support to clinical decision making in CCU.

## 2. Materials and methods

A personal computer 'PC Bit' of Bit Computers Co. (Rome, Italy), with 8086 microprocessor, random access memory (RAM) of 640 kilobytes (KB), 20 megabytes (MB) hard disk, color graphic adaptor (CGA) monitor and dBase III Plus software (Ashton Tate) was used.

Each patient's record is organized in 52 fields, alphanumeric and numeric, with a total of 3186 bytes. The first fields relate to demographical and historical data and to diagnosis; further fields contain clinical complications (including arrhythmias) and therapy; data of cardiac enzymes and the results of the diagnostic examinations (Holter tape analysis, echocardiography, coronary angiography and ventriculography) may also be introduced. Finally, there are fields for recording deaths and causes of death.

The records of the various patients are entered after collection of all the data in a specially designed record-chart, which facilitates the secretarial work.

Each record in the file can be visualized and a discharge form which summarizes all clinical, laboratory and diagnostic findings of the patient at issue can be printed. To obtain a graphically acceptable printout, a programme for screen presentation and fields denomination was designed, for easier interpretation of the discharge form by external users (cardiologists in the ward, or general practitioners).

A primary aim of our DBMS was the possibility of obtaining information from the entire patient population admitted to our CCU over a number of years. Two different systems of data entry and data retrieval were developed: the first system is presently in operation whilst the second represents the immediate future of our DBMS.

With the present system, uncoded data are entered in clear language, indicating historical data, clinical complications, therapy and diagnostic investigation results. This method of data entry offers two advantages:

- ease of use;
- no coding, which eliminates the problem of having to anticipate all the various features of clinical and therapeutic parameters.

A drawback of the uncoded system is that it is easier to make mistakes when entering words than when keying codes of only one or two letters; moreover, one has to be careful not to use synonyms for the same clinical event.

A discharge form containing all the information relative to a patient's stay in the CCU can be automatically generated by entering the name in the first option of the menu. The form provides demographical, historical, clinical and therapeutic data, as well as laboratory findings and the results of diagnostic investigations (Figure 1).

A complex series of data retrieval commands was created, which evaluates the most common demographical, clinical and therapeutic parameters. The whole series begins with a single command ('update statistics') in the menu: data of all patient records are examined and calculation results go to update a series of memory variables: the data retrieval is performed through options

UTIC S. CAMILLO - Roma  
 Director: F.Milazzotto,M.D.

record chart: V. GIANNI  
 age: 50 sex: M  
 days in CCU: 7

diagnosis: ANTERIOR (Q) AMI  
 history: Hypertension

events 1 day: ANGINA,KILLIP 1,HYPERGLICEMIA  
 therapy: UROKINASE2MUI,ivPROCICLIDE3.6g,iv DILTIAZEM200mg  
 events 2 day:  
 therapy: ivNG50mg,ivDILTIAZEM200mg,ivPROCICLIDE3.6g  
 events 3 day: arrhythmias: VEBs  
 therapy: tdNG20mg,ivNG50mg,ivDILTIAZEM200mg,GALLOPAMIL100mg,  
 ivPROCICLIDE2.4g  
 events 4 day:  
 therapy: tdNG20mg,GALLOPAMIL150mg,ivPROCICLIDE2.4g  
 events 5 day:  
 therapy: tdNG20mg,GALLOPAMIL150mg,ivPROCICLIDE2.4g  
 events 6 day:  
 therapy: tdNG20mg,GALLOPAMIL150mg,ivPROCICLIDE2.4g  
 events 7 day: DISCHARGE FROM CCU  
 therapy: tdNG20mg,GALLOPAMIL150mg.

daily CK: 153 1794 468 158 134 117 74  
 daily CKMB: 34 67 52 31  
 coronary angiography: 2VESSELS:60%RC,100%LAD,40%CX;  
 ventriculography: ANTEROLATERAL and APICAL AKINESIA  
 echocardiography: ASINERGY SCORE=10; LVWM SCORE=45%  
 holter: VPBs=248,LOWN=1B

*Fig. 1.* Discharge form of a patient from CCU.

(Figure 2) which scan these variables.

The main advantage of this system is that the data retrieval activity takes very little time even when hundreds or thousands of records are examined: this is a consequence of the fact that the memory variables read by the programme consist of few bytes only and that they have been updated in advance with calculation results (through the command 'update statistics' in the menu). It is always possible, however, to perform data retrieval with non-preset clinical parameters by using the command 'set filter to', granted a longer time for processing.

### 3. Results

By suitable commands (Figure 2) we may access to several preset data retrieval options: the choice of one of these options (e.g.: 'clinical complications': '4') presents the results shown in Figure 3, quickly and independently from the number of records scanned (in this case, 50 patients).

To retrieve data with non-preset parameters, one has simply to enter key words linked together by 'and' or 'or' relations. An example of this possibility is shown in Figure 4, where the parameter investigated refers to two vessel

COMMANDS FILE  
=====

NUMBER OF RECORDS, AGE, SEX.....1  
SITE OF AMI.....2  
HISTORY.....3  
CLINICAL COMPLICATIONS.....4  
LIFE THREATENING VENTRICULAR ARRHYTHMIAS.....5  
SUPRAVENTRICULAR ARRHYTHMIAS.....6  
DISTURBANCES OF RHYTHM AND CONDUCTION.....7  
THERAPY.....8  
PATIENTS WITH POST-INFARCTION ANGINA.....9  
CARDIAC FAILURE.....A  
THERAPY OF VT OR VF.....B  
GO TO MENU.....X

Fig. 2. Possibility of choice of preset data retrieval options through memory variables.

CLINICAL COMPLICATIONS  
=====

KILLIP 1-2 = 27  
ACUTE PULMONARY OEDEMA = 1  
SHOCK = 6  
POST AMI ANGINA = 9

Fig. 3. Statistical data on some clinical complications (50 patients in the file).

coronary disease: in this example, the number of patients' records and the description of the site and entity of coronary disease are shown.

Graphical presentation of the data included in the memory variables is also supported: these variables are read by the Lotus 1-2-3 software (Lotus Development Corp.) and transferred to a graphical programme such as Concorde (Visual Communication Network, Inc.). Figures 5 and 6 show two examples of graphical presentation with histograms of historical data and clinical complications of all the patients in the database.

```
SET FILTER TO 2VESSELS$CORONARO
LIST CORONARO
```

```
04 coronary angiography=2VESSELS:RC=80%,CX=100%
18 coronary angiography=2VESSELS:RC75%,LAD85%
25 coronary angiography=2VESSELS:LAD70%,CX=60%
40 coronary angiography=2VESSELS:CX60%,RX60%
```

Fig. 4. Analysis of coronary angiographic data of patients with two vessels disease (50 patients in the file).

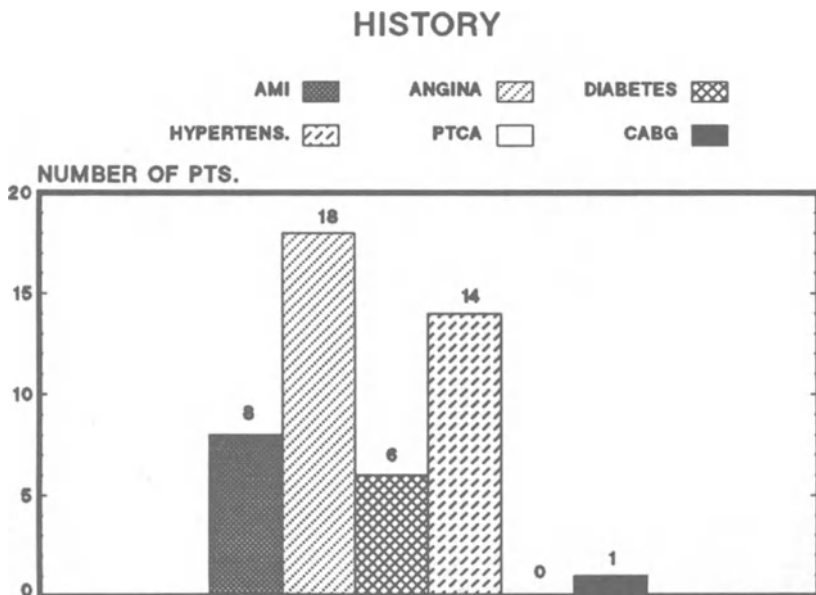


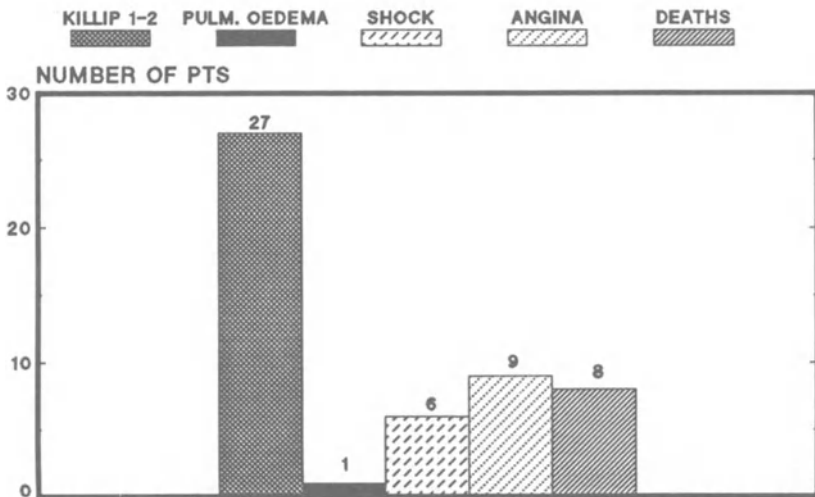
Fig. 5. Graphical presentation of data related to history of the patients (50 patients in the file).

#### 4. Developments

To further improve data entry and data retrieval activities in our DBMS, we decided to switch to a different, completely coded system, having the following features:

- data entry system based on lists of parameters (Figure 7): the operator has only to key the symbol indicated in brackets. For therapy data, it is necessary to choose the drug and key the dosage per days of treatment;
- two or three blank spaces are provided to enter events or drugs or parameters not present in the list;
- the use of symbols instead of words ensures a quick data retrieval, through fields of one or two bytes, without any predetermined query (unlike the system of 'commands file').

## CLINICAL COMPLICATIONS AND DEATHS



*Fig. 6.* Graphical presentation of data related to clinical parameters, complications and death (50 patients in the file).

This new DBMS architecture is easier for inexperienced operators to use and more flexible in data retrieval: in fact, it combines the speed of research through very small fields with the flexibility of non predetermined data retrieval queries. It is possible to retrieve data from hundreds or even thousands of records in a few seconds, with no limit to the number of investigated parameters and their relations.

### 5. Conclusions

The DBMS developed by our group is based on the dBase III Plus software, operating in the MS-DOS environment: this programme is quite easy to use and can be expanded and modified to meet user requirements. The forthcoming edition of our programme is, for instance, designed with only two large fields for clinical complications and therapy: this permits the recording of an unlimited number of days of hospitalization and simplifies the series of instructions to create a data retrieval option.

In our opinion, the main advantage of this programme is the great speed in data retrieval through memory variables. For all the parameters connected to a memory variable, data retrieval covering thousands of records may take only 1-2 seconds; this time could be further reduced with faster microprocessors.

## HISTORY

PREVIOUS AMI (A)	<input type="checkbox"/>	SUPRAVENT. ARRHYTHMIAS (L)	<input type="checkbox"/>
ANGINA (B)	<input type="checkbox"/>	VENTRICUL. ARRHYTHMIAS (M)	<input type="checkbox"/>
PERIPHERAL VASC. DISEASE (C)	<input type="checkbox"/>	A-V BLOCK (N)	<input type="checkbox"/>
CEREBRO-VASC. DISEASE (D)	<input type="checkbox"/>	RBBB/LBBB/LAH/LPH (O/P/Q/R)	<input type="checkbox"/>
HYPERTENSION (E)	<input type="checkbox"/>	COPD (S)	<input type="checkbox"/>
CABG (F)	<input type="checkbox"/>	NEPHROPATHY (T)	<input type="checkbox"/>
PTCA (G)	<input type="checkbox"/>	G-I ULCER (U)	<input type="checkbox"/>
PMK (H)	<input type="checkbox"/>	DIABETES (V)	<input type="checkbox"/>
CONGEN. HEART DISEASE (I)	<input type="checkbox"/>	HYPERURICAEMIA (W)	<input type="checkbox"/>
VALVUL. HEART DISEASE (J)	<input type="checkbox"/>	ALCOHOL ABUSE (X)	<input type="checkbox"/>
CARDIOMYOPATHY (K)	<input type="checkbox"/>	..... (Y)	<input type="checkbox"/>

*Fig. 7.* Data entry coded system for historical data.

The new structure of our DBMS, with its completely coded data entry, certainly offers great advantages in terms of flexibility and data retrieval.

In conclusion, we believe that the development of a DBMS for a CCU with dBase III Plus software is feasible and supplies very satisfactory results. It is necessary to obtain a close collaboration between the clinical cardiologists and the computer experts during the phases of design, creation and installation of the programme.

### Bibliography

- Cox JR, Zeelenberg C (1987) Computer technology: state of the art and future trends. *J Am Coll Cardiol* 9: 204-214
- Reichert PL (1982) The challenge of medical informatics: delusion and new perspectives. *Med Info* 7: 57-66
- Sakabe N, Nakizaki K (1985) Review of medical information processing applications by personal computer. *J Acoust Jpn* 41: 637-644
- Teichholtz L, Leblanc D (1984) *Cardios: clinical and research database information system*. Computers in cardiology, Salt Lake City, Long Beach, CA, USA, IEEE 233-236
- Yasnoff WA, Brewster PS, Demain-MC Greevey K, Leighton RF, Fraker TD (1984) A database system for research, teaching and clinical decision making in cardiology. *Computers in cardiology*, Salt Lake City, Long Beach, CA, USA, IEEE 489-492
- Citone C, Carelli M, Di Marcotullio G *et al.* (1984) Programma di gestione giornaliera computerizzata dei pazienti ricoverati in UTIC. *Atti XV Congresso Nazionale ANMCO*, Firenze
- Citone C, Tubaro M, Carelli M, Di Marcotullio G, Malinconico U, Milazzotto F (1986) Computerized daily management of a CCU. *X World Congress of Cardiology Abstract Book*, Washington, DC, USA 244
- Citone C, Iacoangeli C, Tubaro M, Milazzotto F (1988) Use of a data base programme for computerized management of CCU data. Presentation at X Congress of ESC, Wien, Austria

Tubaro M, Iacoangeli C, Milazzotto F (1989) Computerized management of Coronary Care Unit data: use of a data base programme. Pincioli F, Milazzotto F, Martini W, Matta F (eds) (1989) Banche di dati integrate in cardiologia. Torino (Italy), Centro Scientifico Torinese 103-109



# Information management for decision making by critical care physicians and nurses

THOMAS W. ROSENAL\*

*Director, Department of Intensive Care, Calgary General Hospital and  
Associate Professor, Department of Medicine, Faculty of Medicine, The University of Calgary*

MARGARET WHITE

*Nursing Unit Manager, Department of Intensive Care, Calgary General Hospital*

KATHRYN J. HANNAH

*Director of Nursing, Systems, Research and Education, Calgary General Hospital and  
Professor, Faculty of Nursing, The University of Calgary*

## Abstract

This chapter summarises our experience in selecting and evaluating three intensive care monitoring systems on the basis of information needs for clinical decision making by physicians and nurses. It describes how, in the intensive care unit, large volumes of information must be acquired accurately, stored in databases, manipulated correctly and displayed in a useful format. It discusses the importance of validation and integration of data. It points out the strengths and weaknesses of currently available North American monitoring systems and patient data management systems. Finally, this chapter discusses the future of integrated intensive care patient data monitoring systems.

## Introduction

The increasing demands for the management of information in critical care have provided both an impetus to develop a working computer system in the intensive care unit environment and a useful model for the more general management of medical information [1]. The reasons that critical care provides such a good model are several fold. First, it emphasizes the multidisciplinary aspects of the care of the sick by a coordinated, efficient team of health care workers which includes physicians and nurses. Secondly, critical care relies on large volumes of information gathered by both humans and machines. This information is subjected to both simple calculations (eg. cardiac index from cardiac output and body surface area) and complex analysis (eg. ST segment elevation) and must be presented back to the team in a digestible form so they

\*Address correspondence to: Dr. T.W. Rosenal, Director, Department of Intensive Care, Calgary General Hospital, 841 Centre Avenue, East, Calgary, Alberta, Canada, T2E 0A1.

can use it to make clinical decisions. Critical care information must also be handled in real time (e.g. the alarms sounded by automated arrhythmia detectors). Thirdly, critical care requires multi-lateral communication with areas outside the unit such as the clinical laboratory and the pharmacy. This chapter will discuss the history of information management in critical care; experience in selecting and evaluating three systems for the Intensive Care Unit, Calgary General Hospital, Calgary, Canada: the strengths and weaknesses of the North American state-of-the-art and important issues for the future.

## **Background**

Information has been acquired electronically from patients since the development of the electrocardiograph by Einthoven in 1903 [2]. Intensive care units in North America date from the postoperative neurosurgical recovery room established by Dandy at Johns Hopkins Hospital in 1930 [3]. The spread of postoperative intensive care units over the next three decades was driven by recognition of the need for concentrating special nursing and medical skills in one geographic area. During the 1950's, respiratory intensive care units arose in response to the development of ventilatory support for poliomyelitis epidemic victims [4]. In the same decade, after the development of AC defibrillation by Zoll *et al.* [5] in 1956, coronary care units came into being to allow continuous EKG monitoring for detection of fatal arrhythmias. At that time, continuous EKG monitoring was accomplished by specially educated nurses continuously watching sets of monitors.

In the 1960's, the complexity of intensive care was increased greatly by three developments. First, reliable and rapid clinical laboratory tests (such as blood gases and serum electrolytes) became generally available to assist in diagnostics. Simultaneously, the chemical physiology and pathophysiology became better understood and new techniques of patient management evolved. The number of tests available from clinical laboratories has continued to increase exponentially to the present day.

The second development responsible for increasing the complexity of intensive care medicine was that of invasive monitoring. Techniques once limited to a physiology laboratory were brought to the patient's bedside. Physicians, nurses and technicians gained expertise in the clinical use of machines which could measure physiologic parameters such as pressure, blood flow, and temperature. This equipment required such technical manipulation as zeroing, electrical balancing and calibration. The classic example is the pulmonary artery catheter developed by Swan and Ganz [6] which was followed quickly by an enhanced catheter which could also perform thermodilution cardiac outputs. A spin-off from these activities was the need to create desk top calculator programs to determine derived variables such as systemic vascular resistance from known mathematical formulae applied to the measured variables.

Finally, the third development contributing to the complexity of intensive care was the ability to provide organ support. This began with haemodialysis and positive pressure ventilation in the 1960's and has been extended to cardiopulmonary bypass and intra-aortic balloon pump support. Each new support machine required monitoring of the machine itself as well as increased monitoring of the patient to measure the effectiveness of the support system.

Throughout the 1960's and '70's, the information from the laboratory, invasive monitoring and organ support devices was recorded, almost always by nurses, on specially designed paper forms called flow sheets. The need for this data recording redefined the role of hospital-based nurses in a manner that was unlikely to have been anticipated by their colleagues in the first half of the century. In centres with innovative computing science departments, such as the Latter Day Saints (LDS) Hospital in Salt Lake City, Utah and University Hospital in Birmingham, Alabama, there was a sense that computers might be able to relieve nurses of this task and perhaps perform it even more efficiently, accurately and legibly [7]. The majority of ICU's did not have such computing resources and, therefore depended on off-the-shelf systems.

Toward the end of the 1970's, the rapidly increasing volume of information virtually cried out for automated information acquisition and manipulation. Although the technology available at that time was barely adequate to allow such automation, one form of automation that was successfully initiated at that time was arrhythmia analysis. Several algorithms were developed [8] to analyze on-line electrocardiograms in order to detect fatal arrhythmias, thus replacing the drudgery for intensive care nurses of observing cardiac monitors for hours at a time. Given the repetitiveness of this task, it is not surprising that studies done at the time showed that a properly functioning arrhythmia analysis program would outperform an experienced nurse [9]. These early arrhythmia detection systems provided a lower false negative rate, but a greater false positive rate than did nurses. This latter problem has improved with subsequent developments.

On the other hand, attempts at developing patient data management systems (PDMS) were unsuccessful. A PDMS is a computer system designed to collect information into a database both from the monitoring system and from nurses (eg. vital signs, urine output, fluid and medication administration) and physicians (eg. diagnoses, medical procedures). Even less successful were the attempts to integrate a PDMS with the laboratory, pharmacy and hospital information systems. To illustrate some of these problems, there follows a description of our evaluation of three different patient data management systems at the Calgary General Hospital between 1978 and 1989.

### **The Calgary experience**

A new 31 bed multidisciplinary intensive care unit (ICU) was opened in 1978. The medical director and nursing unit manager at the time sought an off-the-

shelf monitoring system which would provide automated arrhythmia analysis. Since that time, on three different occasions, we have been involved in the selection and evaluation of ICU patient monitoring systems. The first, purchased in 1979, was a General Electric 3500 (GE 3500) system consisting of bedside monitors interfaced in a complex manner to a PDP 11/34 minicomputer and accessible through two central stations (video terminals, keyboards and EKG strip chart recorder) with a single graphics-incapable printer. Up to four modules designed to perform specific functions such as EKG and pressure monitoring were 'plugged into' each bedside as required. Modules measuring respiratory rate and temperature were less reliable. The system provided automated arrhythmia analysis which was appropriately sensitive, but which frequently misinterpreted electrical noise and atrial fibrillation as ventricular fibrillation. It also provided drug dosage calculations, haemodynamic and pulmonary profiles and would calculate 24 hour fluid balances. The system could trend monitored parameters.

In an attempt to automate charting, we engaged in development work with the manufacturer and Swedish Hospital in Seattle to upgrade the GE 3500 to a true PDMS which would allow nurses to enter intravenous fluid type and rate information as well as drugs and dosages. The intent was to integrate fluid and drug information with the trended data to ease making medical and nursing decisions. For example, we hoped it would automatically notify us of drug incompatibility or interaction, or allow a patient's nurse to quickly recognize after a few hours that the intended 24 hour fluid balance would not be achieved or help the physician assess the effect of an antiarrhythmic on the frequency of premature ventricular contractions (PVC's). The PDMS component of the system did not work for a variety of reasons. With only two terminals, and these not at the bedside, the nurses recorded their information first on paper and then entered it into the system resulting in the inefficiencies (and the potential for errors) of dual data entry. There was a tendency for all the information to be entered at the end of a 12 hour shift resulting in a line-up of nurses in front of each terminal for the last hour of each shift, clearly a poor use of their time. The screens provided poor character resolution and access to various functions was through a menu structure which was several levels too deep. These inefficiencies were further exacerbated by a relatively slow response time. The software was not ergonomically designed for the management of intensive care patients: for example, common tasks such as starting five medications at once required reentering several submenus five times. Since the bedside monitors did not have a built-in cardiac output system and measurements needed to be performed on a separate machine, these outputs had to be entered manually into the system rather than automatically. All this resulted in the nurses spending inordinate amounts of time entering information through a poor user interface. Further, the system was not flexible enough to provide truly useful reports based on the information entered at so great a human expense.

By 1982, we recognized that the arrhythmia detection system worked well but that the technology for patient data management was too primitive. We,

therefore, changed to the next generation General Electric central terminals which had just been developed. The bedside monitors were left unchanged and the new terminals were driven by three PDP 11/23 minicomputers. This GE 3600 system was successful because its scope was limited to arrhythmia detection and the calculator function for cardiac and respiratory profiles. The screens provided better resolution and the keyboard, being highly specialized, avoided the problem of paging through submenus to execute commands. It easily provided trending information with some graphics that were printed on a rather temperamental thermal printer. The system functioned as advertised and was well accepted. However, it did almost nothing to reduce the shock of ever increasing volumes of information flowing from the patient, the nurses, the lab or pharmacy, all of which needed to be collated in order to make clinical decisions.

The GE 3600 system was orphaned in 1985 when General Electric sold its patient monitoring systems to Marquette. By 1987, it was impossible to obtain spare parts for a system whose bedside monitors were nine years old and whose video terminals were four years old. We began to search for our next system and were surprised to find that only two (Hewlett Packard and Mennen-Greatbach) of the five vendors examined in 1978 had survived to 1987. E for M, General Electric and Roche had disappeared. In the meantime, Nihon Kohden, Siemens, Space Labs and Marquette had appeared in the ICU monitoring field.

### **State-of-the-art**

We recently surveyed the market for ‘off-the shelf’ intensive care patient monitoring systems. To facilitate this process, we developed the following checklist. The checklist was based on established concepts of system selection and our belief that we had to meet the needs of all clinical information users [10].

1. *Analyse current system to generate a Request for Proposal (RFP)*
  - Establish a small selection committee with representatives from at least medicine and nursing.
  - Consult users (nurses, physicians, respiratory therapists etc.) about needs.
  - Describe the current system (paper and/or computer) in detail. Highlight those features that help the users and those that hinder.
  - The selection committee should categorize the requests from users by priority.
  
2. *Consider future needs (Average system lifetime seven to ten years)*
  - Will the new system necessitate more personnel or new roles?
  - Is a patient data management system of value in your setting?
  - With which other systems (eg. laboratory, hospital information system) should the ICU system be integrated?

### 3. *Review the submitted proposals*

- Have each system demonstrated under simulated, but realistic conditions at a site where the majority of users can inspect it and offer candid comments. Remember to test as much of the system as possible. For example, if the ability to pull in data from one monitor to a monitor in a different room is important, ensure that the vendor demonstrates this capability.
- Consult the institutional biomedical engineering department or computing systems department to ensure that the vendor can meet technical requirements. However, you should make clear to this support department that the system is being chosen to answer your needs, not those of biomedical engineering. This will avoid acquiring a system whose hardware is closer to the cutting edge than other vendors, but has poorer clinical utility.
- Ensure that the pressure monitoring capabilities are accurate.
- Ensure that the rhythm interpreter has been tested against a standard database of arrhythmias [11].
- After developing a shortlist, travel to ICU's already using these systems. Choose units similar in size, personnel and administrative structure and patient population to your own. Ensure that the vendor is not present when seeking the opinions of these users.
- Ensure that the vendor is viable for the length of time you expect to have the system.

### 4. *Check hidden costs*

- Installation.
- Hardware and software upgrades.
- New disposables (eg. paper, ink ribbon).
- Maintenance and repair costs. This will depend in part on the level of involvement by in-house biomedical engineering department. The number of systems by the same vendor in the local institution, city or area will have a large influence on both cost and the availability of material and repair personnel. Spare components are always worthwhile.
- Ensure that user training is on-site. Check cost of trainer and need for extra caregivers during training.

### 5. *Establish a timetable*

- Installation
- Training

From our survey, it is clear that during the 1980's both the new and the established companies have made good use of the advances in microcomputer chips to significantly improve the monitoring system's abilities in arrhythmia detection and haemodynamic profile construction.

The systems currently available are all modular and robust and provide more functions and information directly via the bedside monitor including:

1. High resolution tracings and numerical displays of four or more parameters.
2. Flexible labelling of each parameter by input through the bedside monitor.
3. A compact printer attached to each bedside monitor. Optionally, in a less intensive nursing care setting, printers can be located centrally, away from the bedside.
4. Arrhythmia detection and alarming at multiple programmable levels that relate to the severity of the arrhythmia. There are some differences from one manufacturer to the next in the arrhythmia algorithms and multilead analysis so that some claim better signal-to-noise ratio. There are also differences in the ways in which arrhythmias are stored: in some systems, they are stored by comparison to a 'template' of arrhythmias and in others, each arrhythmia is considered unique.
5. Simplified zeroing and calibration through the monitor.
6. Built-in thermodilution cardiac output capability including the ability to select and average several measurements.
7. Special cursor control to allow fine tuning of pulmonary capillary wedge pressure measurements *vis-à-vis* the respiratory cycle and other factors.
8. Built-in haemodynamic and pulmonary calculations.
9. Trending of all captured information.
10. Monitor networking so that a monitor in one room can 'pull in' the data from any other monitor in the unit.

Several other new capabilities are available from subsets of manufacturers. For our purposes, the most important of these were:

1. The ability to continue patient monitoring when the patient temporarily leaves the intensive care unit for tests or procedures. Some systems do this by transmitting information from a separate module at the patient's bedside back to the intensive care unit through a telemetry system. Other manufacturers solve this problem by having a removable integrated cardiac output/EKG/pressure module. This module can be inserted into a transport monitor which travels with the patient. On the patient's return, the module is 'plugged back into' the bedside monitor without loss of data. In the first method, the patient information is displayed on the monitor in the ICU room that the patient has vacated temporarily; in the second method, the information is displayed at the patient's bedside wherever the patient travels.
2. ST segment analysis to assess for silent myocardial ischemia.
3. Holter monitoring integrated into the bedside monitor.
4. Pulse oximetry
5. Non invasive blood pressure.
6. Touch-sensitive screen using on-screen menu buttons.

Given the complex engineering needed to produce a bedside monitor, ICU's in virtually all hospitals use off-the shelf monitoring systems. It seems clear that the advances in the systems described above are the result of close collaboration with the major user of these monitors, the ICU nurse. The displayed information is now brought to the nurse rather than requiring the nurse to leave the

patient's bedside to find the information. The integration of cardiac output and haemodynamics into the monitor also avoids the need for the nurse to record this information on a flow sheet. The number of false positive arrhythmia alarms have been substantially reduced.

In contrast to the effective fine tuning of monitoring systems, the full patient data monitoring system remains an immature product. Some PDMS's have been developed in large hospitals with active computing departments. Some of these include high levels of integration to the laboratory and other systems. Commercial ICU patient data management systems have also been developed, but can also not be considered mature products. Virtually all leading vendors of ICU monitoring systems have partial solutions to patient data management. The approach varies greatly from company to company, the installed base is very narrow indeed and, therefore, the choice of among them is difficult. For example, Marquette is developing a PDMS built around a bedside keyboard which is impervious to body fluids and has a 'trim knob' loosely modelled on the track-ball device. This knob allows rapid selection of menu items for efficient recording of clinical interventions. Independent clinical trials have not yet been performed on this PDMS. Generally speaking, the PDMS is so closely tied with the monitoring system that at present, one must use the PDMS made by the same vendor as the monitoring system.

At the far end of the spectrum of patient monitoring is the closed loop system. Here, under well defined clinical circumstances, output (eg. urine rate) from a measuring device drives therapy (eg. intravenous fluid administration rate) with only intermittent caregiver intervention [12].

At the University of Calgary, we have developed TRACER, a microcomputer-based database system to assist in professional decision making by physicians and nurses. The system is installed in the adult multidisciplinary intensive care units of the two major affiliated teaching hospitals. TRACER manages information for direct patient care, for quality assurance, for teaching and for research.

To assist in patient care, TRACER organizes clinical information which has traditionally been widely scattered throughout the patient's chart and the ICU. The system's output includes active and resolved diagnoses, procedures done over the previous 24 hours, names of physicians performing procedures, and any complications of these procedures. TRACER also lists each patient's current invasive diagnostic and therapeutic devices and the times since their insertion. This allows the system to prompt caregivers to change invasive devices for infection control purposes. TRACER assists prognostication by calculating daily APACHE and TISS scores and by graphically trending this information. The system also tracks referring physicians and consultants so that they receive regular updates on patient status.

For quality assurance purposes, TRACER provides information which includes patient volume, diagnostic mix, and severity. It keeps track of iatrogenic diseases, critical incidents and unexpected patient readmissions. TRACER lists patient mortality corrected for severity on admission and allows



for interhospital comparison of data. It provides data to supplement nursing workload measurements through the more detailed information available from TISS and by tracking the time patients are transported outside the ICU (eg for imaging) accompanied by a nurse. The system provides information on resource utilization by tracking procedures and drug usage.

In the area of clinical education, TRACER provides information in the form of scheduled reports for each nurse and resident by listing the major diagnoses of patients they cared for. It also provides some measure of clinical competence via a list of all procedures performed by each resident and the complications thereof. TRACER supports research efforts by maintaining very detailed data on carefully identified subgroups of ICU patients. For example, in all patients who have a diagnosis of adult respiratory distress syndrome, the system records respiratory parameters such as static lung compliance and arteriovenous oxygen difference.

TRACER is written in Foxbase Plus and runs on an IBM-compatible 80386 microcomputer with a 100-megabyte hard disk. The database software was chosen for its relative simplicity and speed and to conform to hospital and university requirements for software standardization. The greatest effort was expended in developing the user interface. At both the bedside and at the computer, we worked with the ensure that the interface incorporated the principles of human ergonomics. The bedside paper entry forms emphasize the fact that both physicians and nurses record data and that both use the combined information that TRACER displays. One ergonomic principle was the minimizing of effort necessary for the bedside nurse and the resident to record data. As a result, data entry onto a paper form requires less than 3 minutes daily per patient. Information from the paper form is then entered into the microcomputer by a data entry clerk. Routines are in place for error checking, information security and patient confidentiality.

Two enhancements will improve the system by reducing unnecessary human work: (1) Direct data entry through bedside terminals and (2) Direct capture of laboratory and pharmacy information from the hospital information system. When we overcome fiscal constraints, the addition of the Marquette EPIC patient data management system to our monitoring system will allow development of these enhancements. The EPIC bedside terminals will capture all physiologic parameters and provide a gateway to the hospital information system. EPIC will then pass this information on to TRACER.

TRACER is accumulating 5 megabytes of information yearly. Data has been entered for the past nine months. We are developing several formal projects to validate the clinical, quality assurance and educational components of the system.

## **The Future**

Is an integrated patient information monitoring system a solution looking for

a problem? No. The problem of information overload has been described in the literature of both medicine and mathematics [13]. Only a nihilist would suggest we might do better ignoring much of our data. The solution to the problem of automatically integrating this information still largely eludes us.

The broad goals in developing and implementing such a system include:

1. Improved patient care including reduced morbidity and mortality
2. Improved environment for the caregiver.
  - a) Maximized time with patient and family.
  - b) Minimized frustration associated with a reduction in:
    - i) Unnecessary charting.
    - ii) Searching for scattered information.
    - iii) Performing repetitive simplistic non-patient care tasks.
    - iv) Learning time for new staff.
  - c) Optimized information presentation for clinical decision making.

Whereas several excellent clinical databases have improved patient management by illuminating the natural history of diseases and the effects of intervention, there is little evidence to date that monitoring systems or PDMS's have benefitted patient care, improved cost/benefit ratio or assisted the caregivers. On the negative side, our British colleagues have made a fair case against the mere existence of coronary units based on trials [14] showing no improvements in short-term mortality. However, others point out that such units may reduce infarct size and limit morbidity and long-term mortality [15]. Eugene Robin has captured the essence of the lack of data to support the use of Swan Ganz catheters in the title of his article 'Death by pulmonary artery catheter' [16]. On the neutral side, the LDS group has shown little difference to nurses in the use of a system developed on-site [17]. On the positive side, pharmacokinetic management of aminoglycosides has shown reduced incidence of toxic serum levels [18] and the clinical utility of integrated patient information has recently been well described [19, 20].

Based on the experience in implementing, using and evaluating three ICU monitoring systems as discussed above, we would echo George Santayana's comment that those who do not know their history are doomed to repeat it. Identifying current needs and anticipating future trends also helps to prepare for changes. Several identifiable factors will influence the development of integrated systems for professional decision making. These are discussed below.

In the foreseeable future, decisions about patient care will continue to be made by physicians and nurses. Although there will be an increase in the level of sophistication of computers used as decision support tools, some years will elapse before machines will be able to make non-trivial independent decisions about patient care. Technology and science will continue to provide us with completely unexpected surprises which will dramatically change our insight into diseases and therapy. Undoubtedly, the growth of biology and biotechnology will increase the number of tests and interventions available. Advances in mathematics and computing science such as the development neural nets,

parallel computers, fuzzy logic and decision support systems will affect the management of information in intensive care in ways which are difficult to anticipate.

Another factor which will influence the development of integrated systems is further movement towards standardized information transmission. Considerable gains have already occurred in Europe with the acceptance of EKG standards [21]. The IEEE is working towards the Medical Information Bus [22, 23].

Human ergonomics will play a large role in the development of future systems. Major improvements in this area have already occurred both in the ICU environment and in the broader world of microcomputers. The need to integrate ever-growing volumes of information is sure to become more desperate. The implementation of such advances depends not only on their own merit, but on the vagaries of market forces.

Another factor will be pressure from institutions and professional bodies to ensure that new clinical tools have been properly evaluated to ensure their validity and efficacy. Accreditation bodies and hospitals are placing increasing emphasis on quality assurance, a process which usually entails the acquisition of large volumes of valid and timely information. This requirement together with the increasing cost of both systems and labour will provide a strong impetus for institutions to integrate their computer systems. The 1980's could be called the decade of the clinical trial. Not only did the number of trials increase dramatically, but so did their quality based on a very productive interaction between clinicians, clinical epidemiologists and statisticians. Toward the end of this decade, trials were being proposed to assess the efficacy of existing technology. Data supporting the use of coronary care units was reexamined. The question arose whether the Swan Ganz catheter provided useful information and improved patient outcome [24]. Randomized, properly controlled trials of such technology were recognized to be more difficult than studies of new drugs because of the need to overcome strong but unfounded beliefs in the benefits of established technologies. Furthermore, blinding is nearly impossible. Nonetheless, the need to define outcome measures and prove efficacy will result in careful clinical analyses of technology, a process which will slow the rate at which technology is introduced into clinical practice, but should help prevent the dissemination of dangerous, or (as is more often the case) useless machinery.

A related area, still in its infancy, is the demonstration of validity and clinical efficacy of computer software. The mathematical approaches to and limitations on such validation will need further development [25]. The issue of software legal liability has heretofore rarely been tested [26]. The rapid growth of medical software will no doubt soon be noticed by the legal profession.

Finally, many societal and professional factors will influence the future of information management in critical care. There is currently a shortage of qualified nurses throughout North America. The factors (eg. poor working conditions, lack of respect, low wages) that are producing this shortage have

been resistant to solutions. This suggests that any technology which reduces nursing workload and consequently increases nurse productivity would be welcome. Also, changes in professional standards of care such as the possible movement in the United States toward exception reporting by nurses will influence information systems. There are two possible consequences of this change. First, if automated systems are also required only to perform 'exception reporting', this will lighten the information load significantly. The second possibility, however, is that automated systems would be expected to fill in routine information as a clinical and liability 'backup' to exception reporting by nurses and this would increase the information flow and the need to ensure validity.

## Conclusion

Over the past 30 years, electronic solutions to information management in the intensive care unit have developed rapidly. These developments have provided some definite benefits such as improved patient outcome using automated arrhythmia analysis. Health care worker environments have also been improved by the complete automation of some functions (eg. arrhythmia detection), through simplification of other tasks (eg. cardiac output determination, calculating derived haemodynamic variables) and by a better user interface (eg. to select and display information to the nurse and physician at the bedside).

In order to further improve access to clinical information, manufacturers must continue to consult with physician and nurse users. To ease integration, manufacturers must also come to recognize the benefits of standardized data storage and transmission. For their part, users must ensure that the information arising from their monitoring system is valid by (1) establishing effective controls on data entry and by (2) requiring manufacturers to demonstrate consistently correct results when testing the system on standardized input including that from databases. The next evolutionary step, the integration of multiple information sources and the provision, of easy and meaningful access to this information will be difficult, but not impossible.

## References

1. Gelernter D (1989) The metamorphosis of information management. *Scientific American* 261(2): 66-71
2. Einthoven W (1903) Ein neues galvanometer. *Ann Phys* 12 (Supp 4): 1059-1071
3. Hilberman M (1975) The evolution of intensive care units. *Crit Care Med* 3: 159-165
4. Astrup P, Gotzche H, Neukirch F (1954) Laboratory investigation during treatment of patients with poliomyelitis and respiratory paralysis. *Brit Med J* 1: 780-786
5. Zoll PM, Linenthal AJ, Gibson W *et al.* (1956) Termination of ventricular fibrillation in man by externally applied electric countershock. *N Engl J Med* 254: 727-32

6. Swan HJC, Ganz W, Forrester JS *et al.* (1970) Catheterization of the heart in man with the use of a flow-directed balloon-tipped catheter. *N Engl J Med* 283: 447-51
7. Warner HR, Oinsted CM and Rutherford BD (1972) HELP – A program for medical decision making. *Comp & Biomed Res* 5: 65-74
8. Cox JR, Nolle FM, Fozzard HA *et al.* (1968) Aztec, a preprocessing program for real-time ECG rhythm analysis. *IEEE Trans Biomed Eng* 15: 128-129
9. Holmberg S, Rydén L, Waldenström A (1977) Efficiency of arrhythmia detection by nurses in a coronary care unit using a decentralised monitoring system. *Brit Heart J* 39: 1019-1025
10. Ball MJ, Hanah KJ (1984) *Using Computers in Nursing*; Appleton-Century-Crofts; 92-97, 137-138, 195-198
11. Mark R, Moody G (1982) *MIT/BIH Arrhythmia Directory*
12. Stephen RL, Pryor TA, Jacobsen SE, Janata J, Kablitz C (1978) A computerized closed-loop feedback control of treatment for patients on dialysis. *MECO 1978 Athens Greece*, June 26-29
13. Ulam S (1976) *Adventures of a Mathematician* New York. Scribners
14. Hill JD, Hampton JR, Mitchell JRA (1978) A randomised trial of home-versus-hospital management for patients with suspected myocardial infarction. *Lancet* 1: 837-841
15. Gordis L, Naggan L, Tonascia J (1977) Pitfalls in evaluating the impact of coronary care units on mortality from myocardial infarction. *Johns Hopkins Med* 141: 287-295
16. Robin ED (1987) Death by pulmonary artery flow-directed catheter. *Chest* 92: 727-731
17. Bradshaw KE *et al.* (1988) Improving Efficiency and Quality in a Computerized ICU. *SCAMSI 12th Annual Proceedings* 763-767
18. Franson TR, Quebbeman EJ, Whipple J, Thomson R, Bubrick J, Rosenberger SL, Ausman RK (1988) Prospective comparison of traditional and pharmacokinetic aminoglycoside dosing methods. *Crit Care Med* 16: 840-843
19. Safran C *et al.* (1989) ClinQuery: A system for online searching of data in a teaching hospital. *Ann Int Med* 111: 751-756
20. Zimmerman JE (ed.) (1989) APACHE III study design: Analytic plan for evaluation of severity and outcome. *Crit Care Med* 17(Pt2 supp)S169-S221
21. Willems JL (1989) Computerised electrocardiographic: Setting the example for standardisation. *Medinfo* 538-542
22. Hawley WL, Tariq H, Gardner RM (1988) Clinical Implementation of an Automated Medical Information Bus in an Intensive Care Unit. *SCAMSI 12th Annual Proceedings* pp 621-624
23. Nolan-Avila LS, Paganelli BE, Norden-Paul RE (1988) The Medical Information Bus: An Automated Method for Capturing Patient Data at the Bedside. *Computers in Nursing* May/June; 115
24. Robin ED (1985) The cult of the Swan-Ganz catheter. Overuse and abuse of pulmonary flow catheters. *Ann Intern Med* 103: 445-449
25. Turhim S, Reggia J, Weinberger J, Mazlin S (1989) A computer aid for transient ischemic attacks. *Medinfo* 538-542
26. Young F (1987) Validation of medical soft-ware: present policy of the Food and Drug Administration. *Ann Intern Med* (1987) 106: 628-629.

# Practical data management in the Cardiology Department of a City Hospital

ALFREDO PIZZUTI

*Department of Cardiology, SS. Annunziata Hospital, I-12038 Savigliano, Italy*

## Abstract

This article describes the database system used in the Cardiology Department in which the writer is employed. A commercially available database management system was adapted to enable operation by personnel unfamiliar with the database interactive language. A straightforward menu structure provides for troublefree performance of basic operations, such as filing and retrieval of data relative to a given patient or examination. Operations involving more sophisticated research are performed in interactive mode which, however, is not difficult to master.

The advantages of a self-made system designed by the actual user/s are: 1) fast installation; 2) low cost; 3) flexibility in terms of future adaptation, data exchange and software maintenance.

## 1. Foreword

The following is an account of the writer's experience in the design and setup of a computer system for a cardiology department. As the problems involved are those commonly encountered in the design of systems of this sort, the solutions proposed here may, of course, be adapted accordingly to meet the requirements of similar institutions. The system itself is extremely straightforward and, to some extent, preliminary. The writer, in fact, is not a computer programmer by profession, but a doctor who, out of necessity and personal interest, was appointed to take charge of the project. This was therefore directed from a medical rather than strictly technical standpoint, with all the advantages and disadvantages this entails.

## 2. Description

### 2.1. Environment

The SS. Annunziata Hospital, with roughly 500 beds, caters to a predominantly rural district, with departments specializing in maternity, paediatric, emergency and practically all types of surgical services. The Cardiology Department (total 10 beds) provides for both intensive and post-intensive care, with an outpatients clinic equipped for non-invasive diagnosis: ambulatory and exercise ECG and echocardiography. Approximately 8000 patients are examined yearly, with a total of 700 ambulatory ECGs, 900 stress tests and 2000 echocardiograms. The department is also equipped with a receiver for ECGs transmitted telephonically by GPs and industrial infirmaries.

### 2.2. Requirements

The first point to bear in mind is that, having been set up *ex novo* in 1985, the department, as indeed the hospital as a whole, presented no existing structure to which the system had to be adapted. Though this, at least theoretically, afforded maximum freedom in the selection of hardware and software, various considerations concurred to make the final solution practically a foregone conclusion: the *de facto* standard status of the Microsoft MS-DOS operating system; hardware availability; cost; and the assistance network available to the world of Personal Computers (PC).

In view of the decision to launch the operation of the system concurrently with that of the department, installation time was of vital importance for ensuring complete data collection and, consequently, accurate statistical findings over the first few months of activity. As, to our knowledge, no

PC based system with 8086 and 80286 CPUs  
 Non-dedicated server (40 MB) + 2 workstations  
 MS-DOS ver 3.3

Ethernet LAN with Novell software  
 dBASE III and dBase III as DBMS

5 modules:

Patients file	(5 yrs.	7000 recs.)
exercise ECG	(5 yrs.	3500 recs.)
ambulatory ECG	(4 yrs.	2500 recs.)
echocardiography	(5 yrs.	7000 recs.)
I.C.V.	(3 yrs.	300 recs.)

Total disk space in use: approx. 9 MB

*Fig. 1* Hardware and software resources.

suitable, low-cost, readily available software existed at that time, it was decided to use a commercial database – Ashton-Tate dBASE III [1] – and adapt this to our requirements.

The hardware and software resources are outlined in Figure 1.

### 2.3. Database structure

The patients file consists of a relational database containing essential data relative to the patient's clinical (cardiological) history. The main file, containing the patient's registration data, refers to additional files relative to potential cardiovascular diseases (Figure 2). Each patient is assigned a record and, of course, an identification number. Optional subfiles, also with a personal record for each patient, may be provided for indicating: type of angina; myocardial infarction, PTCA, By-passes location and date; type of thrombolysis; valvular defect; type of prosthesis and date of surgery; anticoagulant therapy; congenital defect and date of correction; conduction and rhythm disturbance; pacemaker data; type of cardiomyopathy and possible pathogenesis; biopsy date and findings; presence of pericardial disease and/or effusion, etc. Only a limited amount of storage is therefore required: 170 bytes for the registration file and a total of about 800 bytes for the subfiles. On average, each patient occupies 400 bytes.

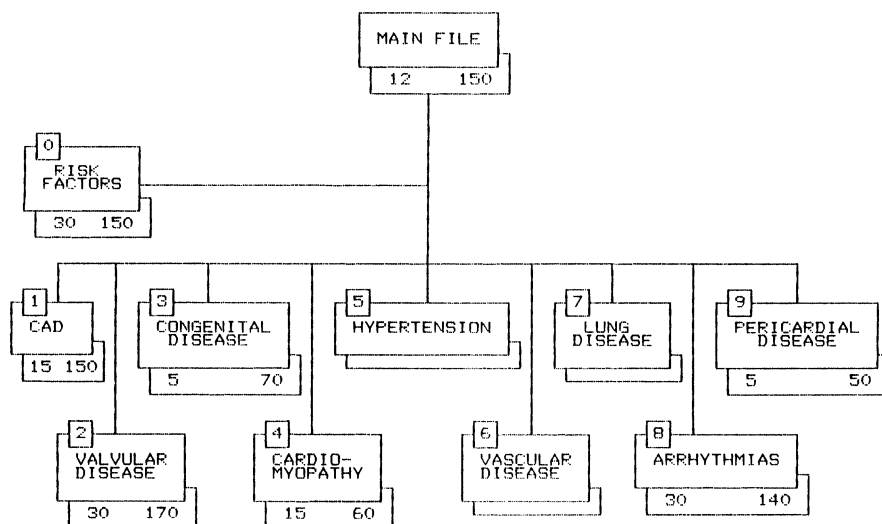


Fig. 2 Database structure. Numbers on the left refer to the fields to be filled in. Numbers on the right indicate the total number of bytes per record; an average of forty bytes per record are left for notes and comments. Files 5, 6 and 7 are not currently in use in our Department; however, the presence of vascular disease or lung disease is recorded in the main file. Data on hypertension are included in file 0 (Risk factors, Physical examination, Laboratory findings).



For Intensive Care Unit patients more detailed information is collected: course, therapy, complications, etc. A separate file has therefore been created, in which each record represents an event.

As regards the ambulatory or exercise ECG and echocardiogram files each examination is assigned a record, and each file contains the examinations conducted per year. The data entered at the time of the examination is used for drawing up the clinical report.

In the case of exercise tests and echocardiograms, the report routines contain analysis and processing algorithms, and automatically provide for any additional remarks. The exercise test report includes a Bayesian analysis of disease potential on a risk and test data basis [2]. The echocardiogram report supplies data relative to valve area, pulmonary hypertension, segmental wall motion and shunt magnitude [3, 4]. The file storage requirement for the above three non-invasive examinations is roughly 400 Kb per 1000 echocardiograms, 200 Kb per 1000 exercise tests, 500 Kb per 1000 ambulatory ECGs.

#### *2.4. Local area network*

The Cardiology Department occupies two floors, which means a small network was required to enable file access from the ward and outpatients clinic. An Ethernet network with a non-dedicated server, thin cable connection and Novell software was selected for reasons of economy, simplicity and, more importantly, upgradability [5]. The file management routine was divided into five modules (patients file, AMI file, exercise ECG, echocardiography, ambulatory ECG) to prevent conflicting multiuser operation.

### **3. Discussion**

#### *3.1. Time saving*

Staff members of Intensive Care Units do not like spending too long in front of a keyboard. An overcomplicated file involving a lot of boring, time consuming keywork will therefore most likely be left incomplete and of little use. Moreover, most situations (check-ups, emergencies, telephone calls, etc.) require not so much consultation of a complete clinical record as a rapid synthesis of the patient's cardiovascular condition. For this reason, provision has been made for a minimum database containing only essential data relative to the patient and his or her clinical history, i.e. an extremely simplified version of the Problem Oriented Clinical Record concept, whereby the record is updated solely in the event of significant occurrences in the patient's clinical history [6].

Access to the five databases is simplified by batch files using the customary DOS routines. File management routines are compiled and directly executable, so that knowledge of the dBASE language is not necessarily required for basic operation.

The patient's name is all that is required in the patient's file. Some routines provide for spelling corrections (upper/lower case letters, figures/characters); determine the existence of the patient's record or patients with the same name; and request, in the case of a new patient, whether a new file is to be created. Once the patient's registration data has been entered, the routine requests which subfile is required, and indicates those already active for the patient in question.

Data is coded wherever possible, each field comprises a help window complete with entry codes or instructions, each choice is flashed, together with the most likely reply, and wherever possible appropriate functions ('range', 'valid', etc;) ensure the input data is congruent. The codes employed were proposed by the medical staff on the basis of simplicity and troublefree operation. Changes may, however, be made to meet the requirements of approved standards.

The design of the main file has also proved realistic and effective for the databases relative to the three non-invasive examinations: minimum data and few options. The choices open to the doctor are essentially two: new examination or retrieval of an existing one. In the case of the former, dating and numbering are automatic; for the latter, only the year and the patient's name are requested. Additional service functions are only accessible when required.

The presence of a data input mask, most of which data is compulsory, provides for maximum uniformity in terms of data collection and report compilation. A certain amount of freedom is allowed, however, for entering personal remarks.

File management, list compilation and statistical analysis have been reduced to an absolute minimum.

A great effort has been made in improving the exterior appearance of the reports, to make them as readable as possible by referral physicians (Figure 3).

Main file data is processed interactively and requires a knowledge of the dBASE language and database structure. The database, however, provides for meeting strictly clinical requirements, and sophisticated or complex analyses involving all the data are rarely necessary (2-3 a year). It was not thought necessary to duplicate what is already successfully provided for in the dBASE software.

As homogeneous lists compiled according to diseases from the non-invasive examination files are more frequently requested by doctors (1-2 a month), provision has been made for special menu-driven routines. Direct printouts are available of lists based on one or more key words (e.g. mitral stenosis, pericardial effusion, ischemic ST in male patients, etc.). Though no more than rudimentary, these functions have proved more than sufficient for our requirements.

### *3.2. Algorithms*

Though straightforward 'IF...THEN' routines, the algorithms in the echo-

## ECHOCARDIOGRAPHIC REPORT

Patient Name

birth date: 11/12/53

BSA: 2.00 m<sup>2</sup>

Left atrial dimension	:	30 mm	(15.0 mm/m <sup>2</sup> )
Aortic root	:	20 mm	
Left ventricle (end-dyast)	:	50 mm	(25.0 mm/m <sup>2</sup> )
(end-syst)	:	25 mm	
(septum)	:	10 mm	
(free-wall)	:	10 mm	
(SF)	:	50%	

The right atrium is enlarged; the right ventricle is normal. The left ventricular wall thickness is normal. An echogenic structure is seen in the left atrium, possibly a thrombus. A moderate pericardial effusion is present.

## Segmental wall motion:

11. apex: aknetic

Left ventricular ejection fraction (Simpson) is approx.: 50%

Mitral valve: stenotic, with area (pressur half-time) of approx. 1.5 cm<sup>2</sup>, and a mean gradient of 12 mmHg.

Aortic valve: stenotic, with a maximal gradient of 50 mmHg, a mean gradient of 18 mmHg and area (continuity eq.) of approx. 2.0 cm<sup>2</sup> (1.0 cm<sup>2</sup>/m<sup>2</sup>)

Pulmonic valve: normal

Tricuspid valve: insufficient, with a systolic gradient between RV and RA of 64 mmHg. Pulmonary hypertension is present.

REMARKS: Examination of good technical quality. The heart rate is 120 bpm, some Doppler data may be unreliable. This underlined text is all what the examiner write down as a free text. The space for this comment is 240-360 byte, accordingly with the version in use and is sufficient in 90% of the examinations.

the cardiologist

Fig. 3 Example of an echicardiogram report. The free text written by the examiner is underlined.

cardiogram and exercise ECG report routines have proved highly effective in terms of time saving and more detailed evaluation of examination findings. As the processing involved would be performed anyway by the examiner, it would be absurd not to take advantage of the processing potential available on the computer, even without recourse to Artificial Intelligence techniques. Several diagnostic aid packages are in fact available on the market, most of the algorithms of which however are dedicated to a specific type of hardware – automatic ECG analysis, Holter monitoring and stress testing systems. Though

the effectiveness and reliability of such packages is sometimes excellent, these are unfortunately self-contained systems related to specific associated equipment. Interfacing with commonly used databases or, worse still, with packages of a different make poses quite a few difficulties. What this virtually amounts to, therefore, is a toss-up between a system of multiple non-dialoging databases, and a complete, high-cost, rigid, commercially self-contained system.

### *3.3. Results*

Staff acceptance of the system has been good. Undoubtedly, gradual implementation in blocks helped to eliminate any existing prejudice against computerized systems and brought about no substantial change in everyday working procedures. The major difficulty was in training personnel to operate the keyboard, especially any functions not found on an ordinary typewriter. All members of staff, however, are capable of operating the database within the scope of their duties: entering registration data (nurses) and clinical data (doctors). New employees need no external training.

Other factors which might possibly account for trouble free acceptance of the system are the low average age of the staff (36 for doctors and 22 for nurses) and the fact that 66% of the doctors possess a home computer.

All paper-printed records of echocardiograms, stress tests and ambulatory ECGs have been abolished. Examination records and reports are now drawn up in approximately 10-20% of the total time taken to perform the examination itself. If a comparison is required, copies of reports relative to any examinations conducted since 1986 are obtainable in the space of a few seconds by simply keying 3-5 straightforward instructions: [search][year][name]. This feature is particularly useful for telephone consultations with other departments or hospitals, with GPs, or in emergency cases. In 70-80% of the cases, the data available from the database makes it unnecessary to consult the complete clinical record, thus saving considerable amounts of time.

All the patients admitted or examined in the outpatient clinic since 1986 (over 7000) are recorded in the database together with any essential cardiological data. Installation of the LAN has greatly simplified updating of the database, which was formerly performed off-line once a week by a nurse who is now no longer required for this particular job. The availability of a workstation alongside the Intensive Care Unit and an Intensive Care Unit file as of 1988 immediately resulted in more detailed information concerning patients with this particular disease. So much so that retrospective analysis is now limited to the time frame in the file, examination of the data relative to the foregoing period being considered either unreliable or excessively time-consuming.

A lot of useful information, in terms of both patient care and efficient running of the department, would undoubtedly be lost or difficult to recover without routine use of the database. In this respect, the cost/benefit ratio is extremely good. Though difficult to quantify for lack of previous supporting

data, the system undoubtedly provides for considerable time saving – approximately 20%.

Since its introduction, the system has been adopted either totally or partially in a further 6 hospitals, with no special operator training required.

#### **4. Conclusion**

Flexibility is an essential feature for enabling future modifications and additions to be accommodated as required. As these are seldom predictable at the initial stage, a commonly used commercial database system providing for maximum flexibility and troublefree handling was selected, and which enabled the files and file management routines to be designed quickly, cheaply, and with a minimum of programming skill. The file management routines have since been updated further to simplify operation by hospital staff as far as possible.

Though rarely employed and involving a number of obvious disadvantages, self-made routines also provide for a lot of important advantages, not least of which is the cost factor. Nowadays, most hospital staff members have at least some training in computer programming [7], so that writing a program for a commercial database is no more difficult than writing a scientific report. Moreover, a self-made routine is often less time consuming than outlining what is needed to a non-medical programmer and checking out the results in terms of clarity and detail. Secondly, any modifications which might crop up later or in response to specific situations may be implemented more or less immediately. This applies to both the data structure and the routines and data analysis algorithms. The present database structure, in fact, is the result of continual additions and pruning based on day-to-day experience over the past five years. Finally, using standard, multipurpose application software facilitates data exchange between different structures and institutions. This latter feature is particularly important in view of the current tendency towards greater cooperation, and is already implemented by our center for exchanging data with the cardiology department of another hospital.

Though a self-made database may not, of course, be the answer to everyone's problems, it goes without saying that greater knowledge and experience of computers on the part of medical staff in general cannot fail to improve the products offered by software house specialists.

#### **Acknowledgement**

I am indebted to Mrs. Sandra Jones for her assistance in the preparation of the manuscript.

## References

1. Ashton-Tate dBASE III technical information book (1984)
2. Diamond GA, Forrester JS (1978) Analysis of probability as aid in the clinical diagnosis of coronary-artery disease. *N Engl J Med*, 300: 1350-58 (1985)
3. Hatle L, Angelsen B Doppler (1985) *Ultrasound in cardiology*. Philadelphia: Lea & Febiger
4. Edwards WD, Tajik AJ, Seward JB (1981) Standardized nomenclature and anatomic basis for regional tomographic analysis of the heart. *Mayo Clin Proc* 56: 479
5. Callahan P, Bradley B (1989) New Token Ring versus Ethernet: counterpoint. *Data Comm* 18: 127
6. Hurst JW (1978) The problem-oriented system: the heart and blood vessels. In Hurst JW (ed) *The Heart*, pp 120-31, New York: McGraw Hill
7. Pincioli F, Milazzotto F, Martiny W, Matta F (ed) (1989) *Banche di dati integrate in cardiologia*. Torino: Centro Scientifico Torinese

# A PC-based implementation of a multi-service software for cardiology

F. PINCIROLI (<sup>1, 2</sup>), C. COMBI (<sup>4</sup>), G. POZZI (<sup>1, 3</sup>), R. ROSSI (<sup>1, 3</sup>)

<sup>1</sup> *Dipartimento di Bioingegneria, Politecnico di Milano, Pza L.da Vinci 32, 20133 Milano Italy*

<sup>2</sup> *Centro di Teoria dei Sistemi of the Consiglio Nazionale delle Ricerche*

<sup>3</sup> *Dottorato di Ricerca in Bioingegneria*

<sup>4</sup> *Cardiology Department of the Ospedali Riuniti, Bergamo, and Fondazione Sestini, Bergamo, Italy*

## Abstract

In recent years some data-management needs in Cardiology Departments seem to find effective software-based solutions. Nevertheless solutions often remain independent of each other. The clinical community is still waiting for more comprehensive proposals, in such a way to make really convenient the use of medical software in everyday practice. One possible solution is to consider performances separately as well as gaining additional marginal value from their integration into a single multiple-service package. A time oriented DBMS to manage medical records for follow-up of cardiological patients and a drug handbook, including adverse drug interactions and dosage information on cardiovascular and related drugs, have been considered basic performances able to provide additional benefits to users because of their integration. The prototype of such software is described in this chapter. The software runs on standard, even portable, IBM compatible PCs. Evaluations about its degree of acceptance by the clinical community are still to be made.

## 1. Introduction

Experience in using electronic computers for the management of medical data has been recorded since the early 60s. Computers seemed to be the ideal instruments to satisfy many different needs: to store/retrieve information, to collect documentation, to analyze the history of a single patient, to analyze a population by multiple statistical analysis, to provide help in clinical decisions, to organize complex information and so on.

Reaching the 90s, we observe that we are still far away from the goals that 30 years ago seemed easily achievable. Many reasons may be considered. Among them the number and complexity of problems that can be solved with the aid of computers.

In the sixties, the goal was to find general methods and to implement general-purpose softwares. Use of general-purpose computer programs did not suffice

in solving broad classes of problems. In fact, the more classes of problems a single program could handle, the more poorly it seemed to do on any individual problem.

Faced with the failure of this approach, the goal is now to develop special-purpose programs, devoted to very narrow classes of problems. Narrow problems can be easily and satisfactorily managed by special-purpose programs. The more narrow the class of problem is, the more satisfactory the programs are.

Today, computer programs may provide doctors with different kinds of services. There are programs for any need: storage and management of clinical data, statistics, graphics, local data-bank consultation, decision aids, diagnostic parameter computation, signal analysis and/or diagnosis, expert consultation in some very specific fields [1, 2], and so on. Each single program is designed to solve a specific problem. Sometimes it does very well, much better than a doctor can do, but it completely ignores the presence of other related problems, and doctors generally do not use it. In fact doctors are generally faced with all the different aspects of a problem at the same time, and they need comprehensive tools. To make a simple operation, like a prescription, doctors could need help from expert systems, decision makers, drug data-bank, and statistical analysis.

To effectively use computers, doctors need programs capable of switching between all the different services they need at the same time. Today's programs cannot do this. A multi-tasking environment gives only limited advantages. In fact, even if simple keystrokes are needed to switch from one program to another, any program needs its own data entry. Very often different programs require the same data. For example, to check for recommended posology of a certain drug before prescribing it, the drug name has to be entered both in the drug consultant and in the medical record program. Consulting a program elicits a need to remember results and re-enter them somewhere else. For instance, if the drug consultant program computes posology, the doctor has to remember and re-type it into the medical record.

Switching from different programs, re-entering data, remembering or writing out intermediate results is annoying, time consuming and may generate mistakes. While using computers implies such complications, they will be resorted to limited and more complex applications.

Merging into a single package the services doctors need at the same time makes things easier.

Merging takes place by integrating different programs and by making possible data exchange among them. Data entered in any unit should be available to all the others; results produced by a unit should be (reasonably) available to all the others. Each single unit should satisfactorily solve the problems it was designed for. Integration should not limit performances achieved by any stand-alone program.

Time needed to switch from one program to another is reduced; time needed to enter data in each unit is reduced, too; the integrated service provides a wide



set of utilities. In such a way the package should prove to be more useful than the simple collection of stand-alone programs.

Full integration may mean automatic activation of some services when some conditions occur. For example, a multiple drug prescription may automatically activate checks on drug interactions: prescription of critical drugs may automatically activate a drug consultant for recommended dosages.

In this chapter we describe an integrated package giving multi-service performances (MS)<sup>2</sup>. The multi-service word comes from the inclusion in the same package of different services. Integration is by now limited only to three units, and only first experimental testing was done. Probably we are still far away from the Ideal Solution, but this is a first step towards it.

## **2. Materials and methods**

The main task of a multi-service software should be a medical record management system (MRMS). Many different units may be integrated into the package: e.g. programs computing special diagnostic/prognostic indexes, aiding in decision making, expert consultation, data-banks.

We started integration merging a MRMS and two different programs of drug consultation: a Drug to Drug Interaction Program (DDIP) and a Drug Information Manager (DIM). These two services, together with a MRMS, do not make the full integrated environment doctors need; they make much more convenient the use of all three units. In fact a simple keystroke is needed to activate any unit from the others. Each satellite unit may use most of the information already entered in the MRMS. Under some conditions the DIM may automatically give MRMS back recommended dosages of drugs as a first suggestion of prescription. During multiple drug prescriptions DDIP is automatically activated to check for interactions.

### *2.1. Hardware and operating system*

Hardware used is standard IBM-compatible PC with a hard disk. The hardware choice comes from two different considerations: firstly, last generation PCs may offer performances typical of yesterday's mini-computers at very low cost. Secondly, PCs are becoming widespread in medical environments: today almost every doctor has access to a PC. This makes the software more accessible.

The price to pay is the operating system. The standard operating system on IBM-compatible PCs is MS-DOS: while today PCs may easily achieve the required performances, MS-DOS does not: it does not support multi-tasking, multi-using, virtual memory, big amount of memory, and so on.

## 2.2. Software – the three different units

The package is made up of three different units. The main unit is a MRMS. It records/retrieves patient information. It also acts as master unit of the package. The two satellite units deal with drugs prescription. The first one is DDIP. It helps doctors to avoid adverse interactions among presently prescribed drugs. The other unit is DIM, whose main purpose is to retrieve recommended dosages of drugs. All the three services may work independently, but they are made stronger by integration into a single multi-service package. In fact to use a stand-alone version of all these programs, doctors need to exit the current program, load another one, use it, exit, restart the first one, and go back to the previous situation. This operation takes too much time to be used in everyday practice: it is resorted to in exceptional cases.

This calls for a multitasking environment, allowing to run more at the same time. In a single-tasking operating system, if a computer already runs a MRMS, DIM or DDIP will probably never be used. An integrated package makes it possible to use these services directly without switching from different programs. Integration allows the satellite units to take information directly from the main unit, without retyping in common data. This makes use of satellite programs easier and faster.

As integration time needed to consult services like DDIP or DIM is cut down: doctors will use these services more often, reducing mistakes, and improving quality of care.

In the next sections the three different units are examined, showing some of the main features of each of them.

*2.2.1. The cardiological time oriented medical record – CarTOMR.* Many different kinds of MRMS are possible. (MS)<sup>2</sup> uses a Time-Oriented Medical Record (TOMR) [3, 4, 5]. A TOMR is a medical record specifically designed to manage patients belonging to a population undergoing periodical visits. The set of parameters checked during every visit is pre-defined and not variable. It mostly depends on pathology and patients' group rather than the single patient. TOMRs are suited to manage many different kinds of patients: hyperthesis, post-infarctuated, diabetics, post-transplanted, and so on. Almost all patients undergoing long or chronic care may be managed by a TOMR.

A TOMR manages data from periodical visits. It also manages other data, like demographics, diagnosis, risk factors and so on. These data differ from visits because they have a less strict temporal relation.

A TOMR should store clinical data in such a way as to allow an easy statistical analysis. Today, the main reason to use computers to manage medical records is ease of data retrieval. Time, money and organization needed to use computers are considerably greater than that needed for traditional hand recording. On the other hand, computers make data retrieval incomparably easier and faster. Normally fast data retrieval is used only to find a required patient. The features of TOMR data structure make it easy to perform

statistical analysis on the whole population of patient [6].

MRMS we used in (MS)<sup>2</sup> is a generic TOMR designed to manage cardiological patients. The main feature of the program, called CPM (Cardio-Patients Manager), is customizability. CPM is designed in such a way to make it easily adaptable to many kinds of different patients that can be managed by a TOMR. Using different customization, it may manage generic cardiological patients, hypertensive patients and post-infarctuated patients. Customization to change data stored by the system. This capability makes it possible to adapt customization to the needs of any specific center where CPM is installed. In this chapter we refer to a generic cardiological customization as CarTOMR.

The open design structure of CPM makes integration with other programs very easy, providing that CPM is the master unit. Acting on configuration data only it is possible to associate events, like introduction of a datum or special keystrokes, to execution of external programs or internal procedures. Activation mechanism is context sensitive: i.e. the same event may give different effects depending on the context.

*2.2.2. The drug-to-drug interaction program-DDIP.* Drug interaction problems may arise when a patient is given two different kinds of active principles. In most cases, the global effect of two drugs given at the same time is an enhanced effect of one of the drugs; in some other cases, the effect may be totally unrelated to that of either drug. It should be known beforehand whether a pair of active principles could be safely prescribed.

Information on drug interaction is available in bulky, unwieldy handbooks. Doctors should check these handbooks every time an adverse drug interaction is suspected. Normally they do not: consulting manuals takes too much time. A computer program may be a good solution: consultation time is reduced to a few seconds. Today many solutions are available [7, 8] but generally are not used. To run a special program checking for drug interactions takes too long.

In an integrated environment, activation of DDIP takes only one keystroke. On activation, DDIP receives from CPM all drug names entered into MRMS. Time needed to consult DDIP is really reduced to a minimum.

Besides fast consultation, DDIP may offer other advantages over manuals: for example, it may perform an effect-keyed retrieval, finding all drugs generating an observed interaction. This kind of retrieval is quite impossible with handbooks.

Like handbooks, DDIP drug-keyed retrieval allows one to find all drugs interacting with a given one and all interactions among a couple of drugs. Moreover, DDIP may also check for interactions among many different drugs.

Information used by DDIP comes from standard literature on drug interactions [9, 10]. We mainly considered drugs affecting the cardiovascular system. Some generally used drugs and substances were included, too. For all drugs names of active principles were used because, in most cases, interactions do not depend on pharmaceutical form or preparation.

**2.2.3. The drug information manager – DIM.** Choice of correct posology is a very difficult task. Doctors may find help in many different bulky handbooks describing drug features, posology, contraindications and other information.

In Italy, many handbooks of this type are published and annually updated. Every doctor owns one or more and really uses it.

The number of drugs listed in handbooks is very large. Computers may really help: fast data retrieval cut down consultation time from minutes to seconds and database update is easier and cheaper [7, 11].

DIM is a typical consultation program for drug information. With only a few keystrokes it gives commercial names, pharmaceutical forms, posologies, clinical information, active principles, effects and therapeutic subgroup of a drug. When needed average, emergency and maintenance dosages and age specifications are given. Posology is detailed in drug quantity, unit of measure, repetition frequency and length of therapy.

Search keys are commercial name, active principles, main effects, therapeutic subgroup. Main source of data is *L'Informatore Farmaceutico* [12]: it collects technical features about drugs sold in Italy and it is officially recognized by the Health Care Ministry. DIM mainly consider drugs affecting cardiovascular system.

DIM can work as a stand-alone program. Nevertheless, if a doctor is using a MRMS and he needs to consult DIM, he needs to exit the MRMS, start DIM, consult it, exit DIM and then start the MRMS again and go back to the required patient. This takes too long and DIM will be seldom used.

Integrated environment of (MS)<sup>2</sup> makes things easier: only a keystroke is needed to start DIM. Some typing is avoided, too. In fact the MRMS gives DIM information about the patient (age, weight, sex). DIM will also give back to CarTOMR names and recommended dosages of the prescribed drug.

### **3. Results**

This section describes main features of the three units and of the global package.

#### **3.1. The medical record management system – MRMS**

The main unit of the package is CPM, a customizable TOMR. CPM is the most complex one. It is divided into different modules, providing data entry, data analysis, system maintenance, and customization. Integration with other units allows one, during data entry, to consult satellite units; units may automatically exchange data.

The main CPM feature is customizability. The program may be customized to manage all data needed for a certain class of patient.

Customizability depends on the design of data structure, that is based on the use of data-dictionary and on data fragmentation in paragraphs. Data-



dictionaries are configuration files containing all information needed to describe any datum used by CPM. Any data file has a related data-dictionary, describing all the features of any field of that file. Most important features described in data dictionary are: field names, types, descriptions, default and legal values, validation procedures, automatic conversion procedures, positions on the screen when the field is entered or displayed, printing and editing options. CPM reads from data-dictionary all the information it needs to know about a datum. In such a way maximum flexibility is warranted. Flexibility is paid in terms of system overall performances. In fact a double seek onto the disk is needed to access any datum.

To gain maximum flexibility data are fragmented in small homogeneous paragraphs. Access to paragraphs is provided by special indexes. Those indexes are configuration files, too. To store data with different features, different kinds of paragraph are provided. There are paragraphs with strict time relation, well suited for periodical visits; paragraphs suitable for demographics data or 'not variable' clinical data; paragraphs more suitable for diagnosis, risk factors and some kind of therapies. (Figure 1)

Number and contents of paragraphs depend on configuration. CarTOMR, our generic cardiological configuration, contains 21 paragraphs, and the user can add or delete how many he wants.

To reduce input data time, no paragraph is mandatory, except identification. Paragraphs not compiled are simply not considered.

On data entry the user has to select on the paragraph index the section he wants to compile. Then CPM looks at the data dictionary to build up the input mask. Field descriptions are painted on the screen. If the record is new, default values are assigned to any field. Then field values are displayed.

When the user modifies a field value, CPM checks in the data-dictionary for possible automatic conversion routines or validation procedures. Automatic conversions (e.g. lower-to-upper-case) are performed while data are typed-in; validation is performed as the field value has been confirmed. Generally validation is a simple comparison with allowed values or ranges, but it can also be a complex routine or an external program defined in the data-dictionary.

*Vocabularies.* Some data, like diagnosis, are very important when statistical analysis is performed. Typing errors or use of synonyms may be critical. A solution is encoding them by vocabularies containing all possible entries. Rather than typing a value, the user selects it in a menu. This makes data entry faster and easier, avoids typing errors and speeds up retrieval.

Vocabularies have a hierarchical structure (Figure 2), allowing faster data selection. The first level contains general descriptions (for example of diseases). The next level contains descriptions relating the previous level choice. This structure is used during data retrieval, allowing selection of patients using general descriptions, too.

Some of the most important data of the CarTOMR are encoded in vocabularies: CarTOMR has 7 different vocabularies, some of them including

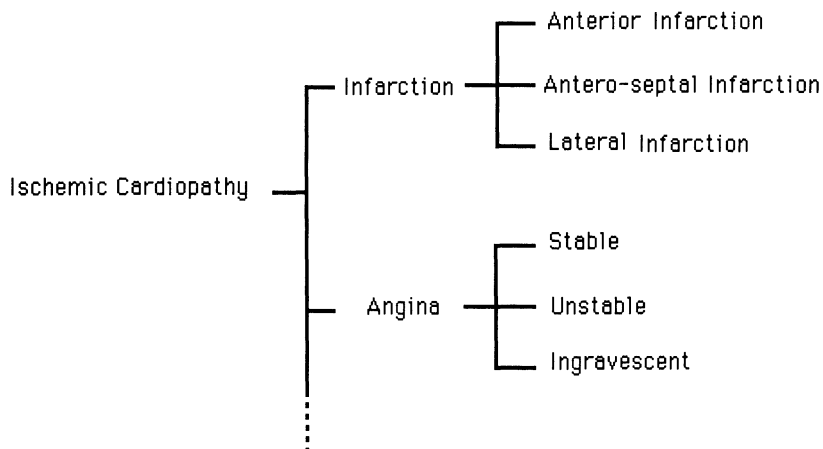


Fig. 2. Example of hierarchical structure of CPM vocabularies. Each item includes more detailed descriptions. Entering data, user selects in a menu the item Ischemic Cardiopathy, then he chooses among Infarction and Angina, and then among the next items. During data retrieval he can choose the item Infarction, meaning all patients with Anterior, Lateral or Antero-septal Infarction. (A real vocabulary may include much more items).

some hundred items. The same vocabulary is used in different paragraphs.

*Data retrieval.* CPM is completed by data retrieval and statistical analysis sections. Data retrieval allows selections of patients-lists starting from almost arbitrary criteria. The user can build up selecting conditions using any possible field or combination of more fields (e.g. age, residency, diagnosis, risk factor). If needed, a condition may consist of two different conditions evaluated after a certain time-interval.

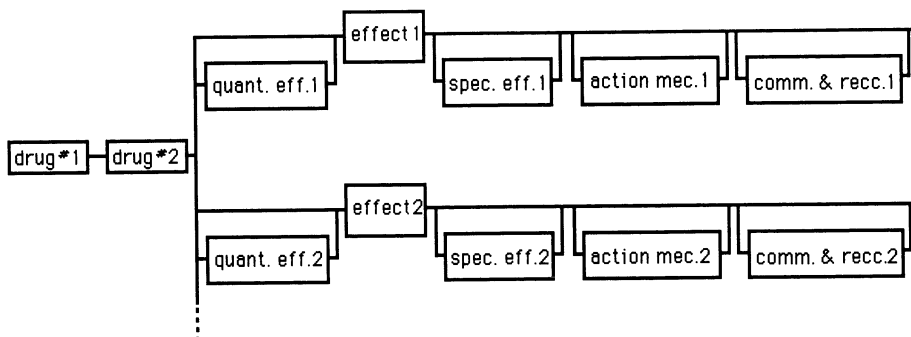
Simple operations among lists like sum, intersection, and logical AND and OR, are allowed, too. A list of patients may be printed, or used for statistical analysis.

The program can build up tables with data from patients in a list, and perform simple descriptive statistics.

An advanced statistical section is not provided. When needed most statistical packages available on MS-DOS PCs can read CPM data. This is not a pitfall in the concept of integration: deep statistical analysis is not performed everyday, and advanced statistical packages give much better performances than a generic MRMS can do.

### 3.2. The drug-to-drug interaction program

DDIP originated as an independent program. After a primary validation in some Clinical Departments, it was distributed as stand-alone program to 1500 clinical departments and family doctors by the accomplishments of the National Hospital Cardiologist Society (ANMCO).



*Fig. 3.* DDIP data structure. The couple drug # 1 and drug #2 triggers effect 1 and effect 2. Any effect description is specified by the name of the effect and four optional fields: a quantifier (mild, severe), a specifier (in case of alcohol abuse), the action mechanism (reduced metabolism), and a field of comments or recommendations.

The database is made of 259 different drugs generating 634 drug-couples, and triggering 198 different effects (Figure 3). The database will be periodically updated. In fact the normal user will prefer an update service by periodical releases rather than a self-update.

The database considers only interactions coming from couples of drugs. This is because very little is known for interactions among three or more drugs. Interactions among more drugs are reduced to the interaction among all pairs of drugs it is possible to build.

DDIP may check for interactions among two or more given drugs, for all drugs interacting with a given one, and for all drug couples triggering a certain effect. To check for the contents of the database some listings are possible, too.

The stand-alone program was considered very useful by most doctors. User interface is so simple that no training at all is required, even for completely inexperienced users.

The problem is the use in everyday practice. Use of DDIP as a stand-alone program implies, every time an interaction is suspected, finding out on the medical record all assumed drugs, activating the program, entering drug names into computer and checking for results. Use of MRMS and DDIP at the same time is needed. Either a multi-tasking environment or an integrated system are necessary.

In (MS)<sup>2</sup>, DDIP may be activated directly while entering therapy into CarTOMR. All prescribed drugs are automatically used by DDIP. This allows consultation with a minimal overhead. Dead times are reduced, and doctors may consult the program every time they need to.

Optionally an early warning procedure may be activated: as a new drug is prescribed, DDIP data-base is automatically consulted to check for possible interactions. This completely avoids unwanted interactions, but slows down data entry.



### 3.2. The drug information manager

DIM originated as a stand-alone program, too. After a primary validation it is now being distributed in 500 copies to doctors.

DIM database consists of 1724 drug-names, 1922 different posologies, 484 active principles, and 2414 descriptions of effects (Figure 4). Updating will be made by periodical releases.

The main purpose of DIM is to inform about recommended posologies. It also gives drugs composition and effects. Main search keys are drug commercial names and active principles. Given an active principle, all drugs containing it are listed. Given a drug name, all features of that drug are displayed.

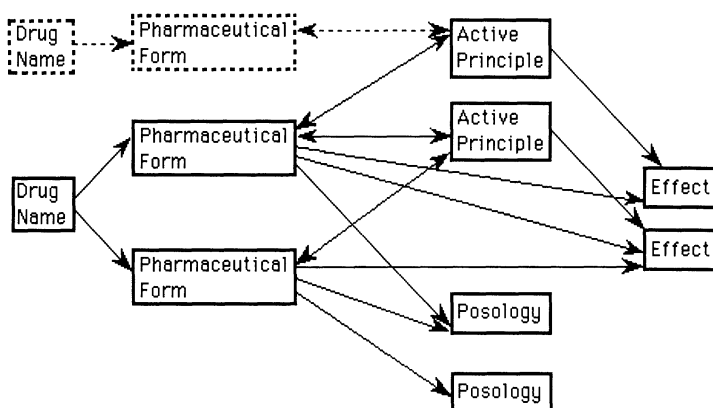


Fig. 4. DIM data structure. Commercial drug names may have different pharmaceutical forms. Any form may have many effects, many different posologies, and it may contain different active principles. The same active principle may be present in many different drugs. Different forms of the same drug may contain different active principles. Active principles are generally responsible for drug effects. Sometimes a pharmaceutical form may have effects unrelated to effects generally triggered by included active principles.

DIM is considered a very useful instrument. User interface is a little more complex than DDIP's, but still very simple. The program is frequently consulted.

The same problem of DDIP arises when both a MRMS and DIM have to be used on the same computer: they are needed at the same time. To consult DIM a doctor must exit from the MRMS, consult DIM, restart MRMS, and restore the situation.

Integration allows one to start DIM from CarTOMR by a single keystroke. Exiting from DIM, the CarTOMR status is automatically restored. Very often patient-dependent data are required to compute recommended posology. All these data are automatically read from the CarTOMR. It will also be possible to give back to CarTOMR names and recommended dosages of drugs to be prescribed. This makes the use of DIM more advantageous.

### 3.4. Integration

The three units described above were initially developed as independent applications. Most features of any single application were already described in previous works [13, 14].

Any unit was initially developed using the same language: Clipper (TM), a dBase (TM) compiler. The master program, CPM, was designed in such a way as to make easy (in principle) a connection with another system. In fact, using configuration files, it is possible to define names of internal routines or external programs to be executed when some specific conditions occur. For example, an external program can be executed when a function key is pressed or when an illegal value is entered on a specific field.

Some problems may arise from program dimensions. CPM is a 450 kbytes long executable file. Minimal run-time memory requirements, even using a sophisticated overlay technique, is over 500 kbytes, and it is nearly close to the upper limit that DOS can manage.

Clipper version of DDIP is an executable file 230 kbytes long, with a run-time occupation of about 300 kbytes. DIM is even larger. Use of these programs as external is not feasible.

A second change is to link together the 3 units, making a unique (very large) executable file. Space saved in this way is considerable, because all run-time support is shared. This way DDIP code is about 80 kbytes long. If DIM and DDIP are never called at the same time and they are used only during data entry procedure, it will be possible to fit both of them in memory using a sophisticated overlay technique. More problems come from the maximum number of files opened at the same time. MS-DOS puts a limit to this number: only in the most recent releases is it possible to overcome it.

To solve this problem and to make consultation faster, a new version of DDIP was developed using C language. It is considerably smaller and faster, and may control very easily the number of opened files. C-version is so fast as to make it possible to check for interactions while entering drugs into the CarTOMR.

Probably the same solution will be applied also to DIM. By now, integration is made using the Clipper version. This implies a complex interface routine to save the current state of CPM, close all open files, start DIM and restore CPM status. This routine limits communications among the two units and slows down calling procedure.

Another problem is to translate all vocabularies from any unit to the others. Since all the three units were initially developed as independent applications, drug vocabularies are different and not compatible. CPM capabilities of customization allows one to store, inside the CarTOMR vocabularies, translation tables providing compatibility.

DDIP and DIM can be activated by pressing special keys; activation is almost instantaneous for DDIP; it takes a few seconds for DIM. Therapies prescribed for the current patient are automatically given to the activated unit. Since DDIP

activation is so fast, it is possible to set-up an automatic activation of the program when a multiple drug prescription is done: if any interaction is found, DDIP emits a warning.

#### 4. Conclusions

We implemented a Multi-Service Medical Software (MS)<sup>2</sup>, integrating in the same package as MRMS a Drug Interaction Program and a Drug Information Manager. The MRMS is a customizable TOMR, allowing traditional data collection, multiple-keyed data retrieval and analysis. Integration allows one, by simple keystrokes, to activate the other two units. Data already entered into the MRMS are automatically given to the activated unit. When some conditions occur, automatic activation of satellite units is possible, too.

The system, providing integrated services, allows one to switch from one unit to another without having to exit the current program, run the other one, restart the first one and go back to the previous situation. This way time wasting in low level operation is avoided, promoting the use of all three units.

The system is of real advantage only if used directly by doctors themselves. In fact all services provided are designed to make doctors' tasks easier and less exposed to mistakes.

#### Acknowledgements

This work was funded by Italian CNR's grants No. 87.02637.07, 88.02870.07, 88.02729.07, 88.02323.07 and by the following: CNR's Centro Studi per la Teoria dei Sistemi; MPI, National Project for Cardiovascular System Bio-engineering; MPI 60%, Engineering School, Politecnico di Milano.

#### References

1. Cimino JJ, Hupp JA, Hoffer EP, Famiglietti KT, Barnett GO (1987)DXPLAIN: An Interactive Knowledge Base for Assistance in Medical Diagnostic Decision. In Proc. of Annual Int. Conf. of IEEE Engineering in Medicine and Biology Society, Boston, MA, USA, pp 1528-1529
2. Quaglini S, Stefanelli M, Barosi G, Berzuini A (1986) ANEMIA: An expert consultation system. Computers and Biomedical Research 19: 13-27
3. Jung D, Brennecke R, Kottmeyer M, Hering R, Clas W, Erbel R, Meyer J (1986) Tree-Base: A Query Language for a Time-Oriented Cardiology Data Base System. In Proc. of Computers in Cardiology, Int. Conf. pp 381-384.
4. Langer RD (1987) TALMRS: Time-Affinity Linked Medical Record System. In proc. of Symposium on computer Application in Medical Care
5. Stead WW, Hammond WE (1987) Demand-Oriented Medical Records: Toward a Physician Work Station. In proc. of Symposium on Computer Application in Medical Care, pp 275-280
6. Mr. Tod User's Guide, Seattle (WA) (1987) Retriever Data System

7. Physicians' Desk Reference on CD-Rom, Oradell (NJ) (1988) Medical Economics Company
8. Hogan RW (1986) Drug Interactions Advisor, Williams & Wilkins, Baltimore, MD, USA
9. Rizack MA, Hillman C (1985) The Medical Letter Handbook of Adverse Drug Interactions. The Medical Letter Inc., New Rochelle
10. Hansten PD (1985) Drug Interactions. Lea & Febiger, Philadelphia PA, USA
11. Drugdex, Micromedex Inc., Denver (CO) (1989)
12. L'INFORMATORE FARMACEUTICO (1988) Annuario Italiano dei Medicamenti e dei Laboratori, Organizzazione Editoriale Medico-Farmaceutica, Milano
13. Pincioli F, Rossi R (1988) GPC: A Time Oriented Medical Record. Poster in 'Experiences of Computer Based Management of Cardiac Units', Satellite meeting to the 1st World Pediatric Cardiac Surgery Congress, Bergamo, Italy
14. Pincioli F, Pozzi G (1989) A drug-to-drug interaction package. Computer Methods & Programs in Biomedicine 28: 183-189

# An out-patient clinic data base management system

G. PINELLI, G. CARINI, G.L. MANINI, S. URBINATI, M. RUFFINI.

*Cardiology Dept. Bellaria Hospital, Bologna, Italy*

L. GUAZZOTTI

*S.G.A.S.r.l., Software Gestionale Applicativo, Bologna, Italy*

## Summary

We describe the main features of our computerized Database for the management of an out-patient cardiology clinic.

The hardware is based on the mod. 5000/95 minocomputer by Unisys, with a 32 Mb central memory and 1 Gb hard disk, connected to a network of personal computers and terminals serving 5 extensions within the Bellaria Hospital (Administration, Admissions Office, Pharmacy, Cardiology, Gastroenterology).

The operating systems are Unix and MS-DOS. The operating language is the relational DataBase Oracle and the query language is SQL-PLUS.

The program manages patients' records, lab test reports and appointments.

The patient's record is made of several forms containing the identification data, the history, the lab exam results and the therapy.

Lab exam (ECG, Echocardiography, Holter, Stress Test) reports can be stored and printed by means of a coding system.

The appointments program is connected to the City Unified Appointment Center that manages the appointments for all the Hospitals and Health Care Centers of the metropolitan area.

The memorized data can be easily retrieved using the SQL-PLUS query language.

The main problem we are facing is to educate doctors, nurses and technicians to work 'on-line' with the computers.

We are confident that our system will make the routine work easier and will improve the clinical and research activities.

## Introduction

The rapid development of computer technologies has led to a wide diffusion, in the medical field, of computerized systems for the purpose of rationalizing and optimizing everyday activities. In cardiology, computers were initially

introduced for limited tasks (ECG, catheterization laboratory, Echocardiography, and nuclear medicine) and later were employed for database management of Divisions and Departments [1-6].

Our experience in this field dates back to 1985 and the features of our earlier system have been previously described [7-8].

Recently, our Administration has undertaken a project that aims to create an integrated computer network connecting all the Services of our Health District, both inside and outside the hospital. This project has implied a technological updating of all the existing hardware and software, including those pertaining to our Department.

However, we have not changed the structure of our program, that consists in putting together in a 'Clinical record' all the patient's data, gathered from history, physical examination and lab tests, from the first examination to the subsequent controls.

## **Materials and methods**

### *Hardware*

The central unit consists of a UNISYS Model 5000/95 Mini Computer, with high computing and memory capabilities (1 gigabyte hard-disk and an initial central memory of 32 Mb), located in the computing center of the Bellaria Hospital, connected to all the hospital extensions. The services already connected in this early phase of the project are:

- Administration
- Admissions office
- Pharmacy
- Cardiology Department
- Gastroenterology Department

For the Cardiology Department the hardware configuration is:

- 1 PC with a 132 chr. printer in the Secretaries office
- 1 PC with a Laser printer in the Doctors room
- 1 terminal with a 80 chr. printer in the ECG room
- 1 terminal with a 80 chr. printer in the Echo room
- 1 terminal with a 80 chr. printer in the Stress-test room.

The PCs are UNISYS-PW2 series 500, with 4 Mb RAM memory, 1 40 Mb hard disk, 1 disk drive for 3 1/2 inches floppy disks and color screen.

### *Software*

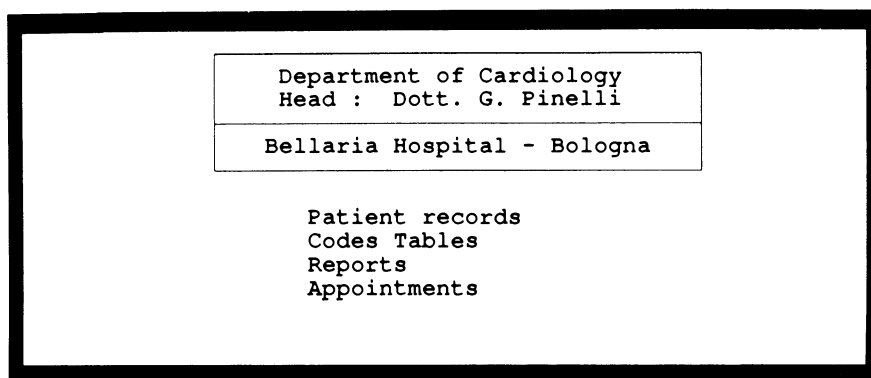
The operating system is UNIX, chosen because of its standardization characteristics. The PCs can, of course, also act as independent workstations, using MS-DOS operating system and commercially available office automation software (word processing, spreadsheet, graphics, statistics).

The operating language is the relational DataBase 'ORACLE', whereas for data analysis and retrieval the SQL-PLUS language is used.

### *Structure of the program*

There are 4 main subprograms (Figure 1):

- CODES TABLES
- PATIENT RECORDS
- REPORTS
- APPOINTMENTS



*Fig. 1.* DataBase main subprograms.

*Codes tables.* All the names, terms and descriptions used to define the data fields are coded and listed in these tables. The use of codes (that have a maximum of 4 digits) means an optimal use of mass memory, a more correct data input and an easier field selection. Moreover, it implies the adoption of a more uniform terminology.

There are three groups of tables:

- laboratory examinations tables, in which the codes of the definitions used for the ECG, Holter, Echo and Stress test reports are listed;
- therapy tables, where the cardiovascular drugs (both active substances and tradenames) are coded;
- general tables, containing the codes of all the remaining fields.

With the 'tables' program, the user can visualize codes and decodifications on the screen, modify them or create new ones.

*Patient records.* The patient records program allows 4 choices (Figure 2):

- input of new patient's data
- research of stored data using patients' names
- research of stored data using patients' code numbers
- general selections and research

```

Patient Records Management

New Patient's Data

Research By Name

Research By Code

Selections and Researches

```

Fig. 2. Patient records management subprograms.

The patients' data are organized in 5 forms:

- patient's identification data
- patient's history
- obitus
- laboratory exams
- therapy

The *patient's identification* form contains several fields as name, address, telephone number, Health District and so on, as shown in figure 3. Currently, every patient is identified by 2 numbers: the first is a progressive code number automatically given by the computer to every new patient at the time of the first data input; the second number refers to the traditional 'papers' file where the folders with all the medical documents (physical examinations reports, ECG

```

PACIENT'S IDENTIFICATION FORM

Input Code : Fiscal Code :
Date Date : File Number:
Date 1' visit :

Family Name: First Name:
Date of birth: Sex:
Status : Job :
Address :
Postal Code: City :
Telephone :
Place of birth :
Region :
Patient :
USL : Ward :
NHS number :

Note: Code :

```

```

PgDn=Obitus F1=Pt.history F5=Therapy sF5=Exams
sF2=Programmed exams

```

Fig. 3. Patient's identification form.





CARDIAC DIAGNOSIS	
Patient :	Code :
Pathology	Date
_____	_____
_____	_____
_____	_____
_____	_____
_____	_____
_____	_____
_____	_____
_____	_____
Notes:	
_____	_____
_____	_____
_____	_____
_____	_____

Fig. 6. Cardiac diagnosis form.

The 'Obitus' (Figure 7) form contains the date and cause of death. Currently it is difficult to gather information about patient death. We carry out periodical inquiries, but only for those patients included in special clinical studies. Later on we plan to connect our computer (as a part of the 'CUP' project, that will be explained later) to the population General Registry of the Bologna City Hall, making possible an effective 'on-line' updating of the 'Obitus' form, at least for the residents in Bologna.

OBITUS FORM	
Patient :	Code
Date of death :	
Cause of death :	
Post-mortem examination :	
_____	_____
_____	_____
_____	_____
_____	_____
_____	_____

Fig. 7. Obitus form.

The *therapy* form contains the list of the cardiovascular drugs, with the name of the active substances, the tradenames and the daily dosage.

The forms of the *lab exams* performed in our Clinic include: ECG, Holter, Echocardiography and Stress test. Every test form is divided into two parts: description and conclusions; the data fields include digits (heart rate, ventricular dimension, valvular gradients) and texts (descriptions of cardiac structures or of electrocardiographic features). All the terms and definitions are coded. The codes directory can be consulted using the 'tables' program or can

be visualized on the screen in a quadrant within the exam's form. For an easier search the long list of codes is divided into 'chapters'. For example, the chapters of the Echocardiographic descriptions are 'left atrium', 'right atrium', 'left ventricle', 'mitral valve' etc. If, for instance, the 'mitral valve' chapter is chosen, the definitions pertaining to the mitral valve ('normal mitral valve', 'leaflets thickening', 'leaflets calcifications' etc.) appear on another quadrant and can be selected using the selection function. The same procedure applies to the other exams.

An example of the Echocardiographic forms is shown in Figures 8, 9, 10, 11.

Patient		ECHOCARDIOGRAM		doctor :		Code	
Date		Indications to the exam					
H.R.	Rythm	LV ts	RV td	Weight	High.	Body	Surface
LV td		IVS ts	LA	PW td	AO	PW ts	Delta %
IVS td		IVS td/PW td		LA/AO			
VCF		LEFT VENTRICULAR VOLUME - Method :		STROKE VOLUME		-	
		Vol. telediastole		CARDIAC INDEX		EF %	
		Vol. telesistole					
DOPPLER							
		max.grad.	ave.grad.	valve area	insufficiency		
Mitral Valve.....		—	—	—	—		
Aortic Valve.....		—	—	—	—		
Tricuspid Valve..		—	—	—	—		
Pulm. Valve.....		—	—	—	—		
LV Pressure		RV pressure		Notes			
sF2=Description		F5=Conclusions					

Fig. 8. Echocardiographic form - I part.

The text can be completed adding coded 'attributes' (as: 'mild' or 'severe') and 40 characters of free text that appear on the screen under the heading 'various'.

Combining definitions, attributes and free text, a complete exam report can be assembled. The final cardiological examination report is prepared in the same way, with a combination of coded sentences and free text. The patient's code number, which appears on every form, is the logical link which unifies and allows a correct retrieval of all the data concerning the patient.

*Reports.* This program manages all the printing functions in order to obtain a copy of all the tests results and of the final report.

*Appointments.* This program allows an accurate scheduling of the clinical activities. The number and the names of the patients who have an exam scheduled at our clinic and the number of free places left are recorded for each day, week or month.

Our system is connected to the Unified Appointments Center (CUP) of Bologna. In this program, realized by the City Administration, a network of





In our previous system, data input followed the 'batch' modality: paper forms were filled in with the data and then entered into the memory.

This procedure caused many difficulties (loss of data, input of incorrect information) and therefore in the new system the data input is 'on-line'.

However, the optimal management of our program still requires an adequate training of personnel in the use of the computers. It appears, at this early stage, that personnel are quite willing to learn these new procedures, especially the younger staff, who seem attracted by computer technologies.

The choice of appropriate codes and definitions for the cardiological lab tests reports required a long phase of research in order to find uniform terminologies and data collecting methods that could be accepted by all cardiologists. This task was carried out in conjunction with the Working Group on Biostatistics and Computer Methods (MBI) of the Italian Association of Hospital Cardiologists (ANMCO). The Working Group has published protocols with the guidelines for the standardization of terms and data collecting methodology concerning ECG, Holter recording, Echocardiography, Stress test and coronary angiography [9]. The use of these protocols will favour the adoption of a common language among the various clinical centers, and it will make it easier to collect uniform data for clinical and statistical purposes.

## Conclusions

We have outlined the most important features of our Database Management System for an out-patient Cardiology Clinic. The aim of the system is to gather all the data concerning the patient, from the first examination to the subsequent controls. We think that our main problem will be the training of doctors, nurses, secretaries and technicians. It will certainly not be an easy task, but we are confident that the difficulties of the early stage will be followed by unquestionable advantages: quicker and more complete test reports, a data base more complete and easier to handle, and a more efficient scheduling of all the clinical activities, for the sake of better patient care.

## References

1. Brodaty D, Montgermont P, Dubois C, Schlum-Berger S, Tawil N, Bachet J, Guodot B, Lebrault M, Guilmet D (1984) La gestion informisée du dossier médical en chirurgie cardiaque. *Arch Mal Coeur* 77: 1344-1350
2. Castellano B, Ricciardello V, Chieffo C, Corsini G, Correale F, Tritto C (1984) Archiviazione dati di una Divisione di cardiologia: Progetto di un sistema con minicomputer ad espansione graduale. *Atti XV Congresso Nazionale ANMCO, Firenze*, p 78
3. Martiny W, Matta F, De Bernardi A, Cerrone G, Doronzo B, Angelino PF (1984) Protocollo computerizzato per i test ergometrici. *Atti XV Congresso Nazionale ANMCO, Firenze*, p 75
4. Fiorista F, Marzegalli M, Barbaresco L, Bana R, Orvieni C, Vincenti A, Lazzari M, Bonini W,

- Broffoni T, Bertolone G, Pinna G, Bertellino A (1987) Sistema informatico coordinato multicentrico per Divisione di Cardiologia. Da: 'Banche di dati integrate in Cardiologia', Edizioni CST, Torino, pp 1-10
5. Matta F, Martiny w (1987) Architettura di un sistema informativo in cardiologia: proposta di un modello operativo. Da: Banche e dati integrate in Cardiologia'. Edizione CST, Torino, pp 27-32
  6. Bonelli R, Maslowsky F, Caico S, Alberio M, Refetto S, Binaghi G (1987) Primo anno di esperienza di gestione di una Divisione di cardiologia con programma di archiviazione dati: difficoltà, evoluzioni, prospettive. Da: 'Banche di dati integrate in Cardiologia'. Edizioni CST, Torino pp 105-110
  7. Pinelli G, Guazzotti L, Manini GL, Di Pasquale G, Dominici P, Carini GC, Grazi P, Tencati R (1987) Gestione automatizzata di un Servizio di Cardiologia senza posti letto. Da: 'Banche di dati integrate in Cardiologia'. Edizioni CST, Torino pp 145-152
  8. Carini GC, Manini GL, Urbinati S, Ruffini M, Pinelli G, Guazzotti L (1989) Considerazioni, valutazione dei risultati, proposte di aggiornamento e potenziamento del programma: Gestione di un Servizio di Cardiologia senza posti letto. da: 'Banche di dati integrate in Cardiologia'. Edizione CST, Torino pp 9-16
  9. Pinelli G, Martiny W, Marchesini A, Matta F (1990) Protocolli informatici in cardiologia. Edizione CST, Torino (in press).

# Databases for network systems

L. GONCALVES and R. DOMINGOS

*Department of Cardiology, University Hospital of Coimbra, Portugal*

## Introduction

Since ancient times men experienced the need to improve their capacities of calculation and storage of data. Successively, they used the abacus, slide rule or an electromechanical calculator for calculation, and mud tablets, tally sticks or hand written accounts for storing data. However, as the complexity of calculation and storage of data increased, these systems became inadequate.

In 1886, when the results of the 1880 census in USA were still being processed, Hollerith realized that it wouldn't be possible for the clerks to cope with the flood of data arising from the 1890 census. Thus, he suggested that the data could be coded on punched cards and that the encoded data could then be analysed by an electromechanical sorter which sensed the position of holes [1]. This idea had already been used by Jacquard in the beginning of the century to program patterns of weaving machines.

With the discovery of the transistor by Shockley, Bratain and Bardeen in 1947 at the Bell Research Laboratories [2], a new era has begun: the era of communication and information. In the next four decades two more, equally important discoveries were made in the world of electronic components [3]. One was the bipolar integrated circuit, in the beginning of the 1960s, followed ten years later, by the integrated circuit CMOS. This astonishing evolution allowed the appearance of the first personal computer (PC) in 1975, produced by a company almost forgotten: Altair. Soon afterwards the worldwide distribution of PCs became a reality.

In many areas of knowledge, such as Medicine, the computer is a powerful tool which can be used in many ways. The European Economic Community (EEC) project Advanced Informatics in Medicine (AIM) has just started financing approved research proposals which represent a landmark in international cooperation in many interdisciplinary areas [4]. Such cooperation owes much to the stimulus of a large number of international bodies working in the field like the Working Group on Computers in Cardiology of the European Society of Cardiology.

Computers' applications in Cardiology are many, and include fundamental



parts of complex machines in Echocardiography [5] or Digital Angiography [6], and other electronic devices [7, 8, 9], such as simulators of physiologic or pathologic phenomena [10, 11, 12], in education [13], diagnostic [14, 15] and therapeutic [16] decision, as database systems [17], or simply in word processing, graphic facilities, statistical calculation, communication and others.

The Cardiology Department of today is a rather complex unit formed by several sectors (Coronary Care Unit, Wards, Consultation Room, Echocardiography, Catherisation Lab etc.) which, if computerized, would gain immensely from the communication between them. This is possible by the use of computer network systems. Networks enable all computerized sectors to take advantage of the organization's total computing capabilities, through information exchange. Networks encourage the free flow of information throughout the Department by giving individuals (with a proper password) ready access to data. Any sector can become a resource to the entire Department simply by connecting it to the network.

### **What kind of network to choose?**

The evolution of computer network systems has led to the appearance of several kinds of networks.

The choice of the right one is highly dependent on the problem we wish to solve, this includes:

- how many, who, in what manner and how often are users going to request information from the databases?;
- how much information is to be kept in the system for real-time use?;
- how much do you intend to spend on the system (hardware and software)?;
- what do you plan to do in the future?;

all these questions must be answered in order to choose the best solution.

Currently there are two major approaches in computer network systems: central processing [18, 19, 20, 21] (see Figure 1) and distributed processing (see Figure 2), each one having its advantages and disadvantages.

On a central processing network system, we have a computer system that acts like a host, and several dumb terminals (without processing capability). In such a system, all the information is stored on the host in one or several large databases, all the processing is also made on the host computer.

On a distributed processing network system, we have several computer systems linked through a Local Area Network. In each one there exists a host with or without a local database with specific information and several dumb terminals, each local database may be accessed by the other hosts for consulting and updating information, depending on a system security scheme. For example: each major sector in a Cardiology Department (such as a cath. lab.) could have a local database with specific information, and a central system could contain the information common to all the Departments (such as patient information).

### The Central Processing approach

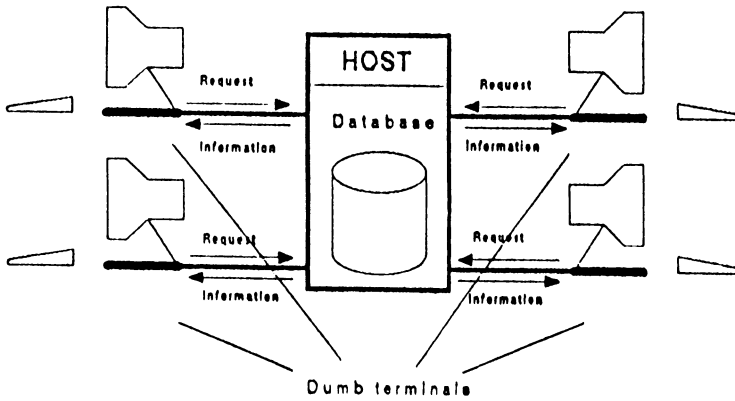


Fig. 1. Central processing network system.

### The Distributed Processing approach

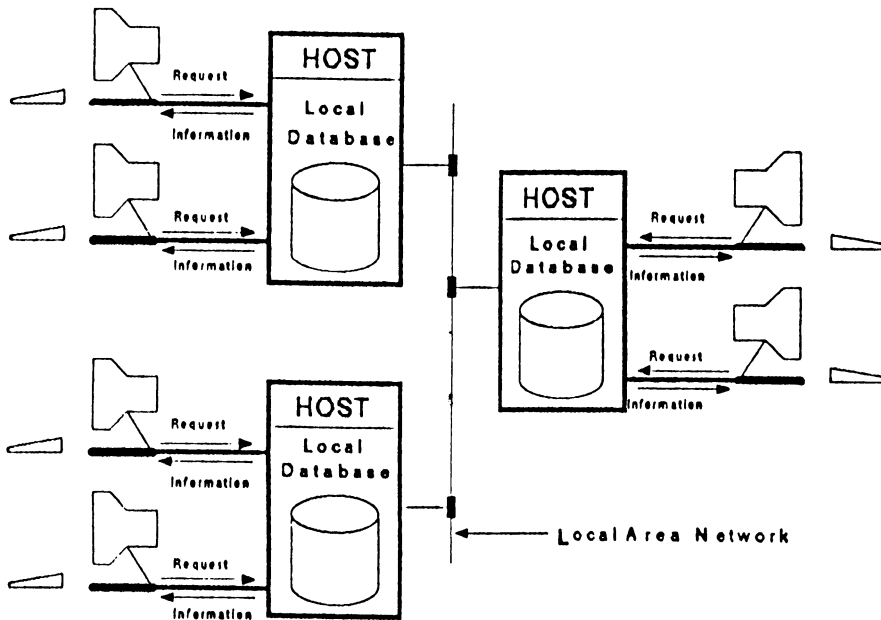


Fig. 2. Distributed processing network system.

### *a) The central processing approach*

If you want to have all the information centred in one big database for the purpose of better administration of the information and keeping the costs low, you should consider the central processing approach (see Figure 1).

Like everything else, central processing has disadvantages and advantages, which we try to explain in the next two paragraphs.

One of the greatest disadvantages is the heavy load on the host derived from intense requests for information. As we can see from Figure 1, all the requests for database information are processed by the host that retrieves the information requested, so, if a large number of users constantly request large amounts of information (the situation one may encounter in a Cardiology Department) the system's response time becomes a problem and the usefulness of the computer network is diminished.

However, the central processing approach has several advantages, the most important being the administration of the system's resources. The security and secrecy of the database information is much simpler, since the database resides physically in one place. Another very important advantage is its low initial and maintenance costs, the hardware and software are much simpler reducing thus the cost of the whole system.

In spite of all these advantages and disadvantages, it is always possible to upgrade a central processing system to a distributed processing system.

### *b) The distributed processing approach*

If you want to link several hospital departments for the purpose of sharing information and resources, and speed the system response time, you should consider the distributed processing approach.

Unlike the central processing system (described earlier), distributed processing consists of several computer systems linked through a Local Area Network. In the LAN messages only travel from one host to another, this message transports requests and database information.

Like central processing, distributed processing also has its disadvantages and advantages.

The greatest disadvantage is the initial and system maintenance costs at several levels such as, hardware, software, and database implementation and administration (secrecy and overflow of information). In such a system the hardware and the software required are much more expensive (due to the greater complexity) than in its counterpart. For the purpose of database administration you would probably need a full time employee, since the databases are dispersed (and probably have different structures).

The great advantages are: the system response time and the possibility of linking different kinds of computers (Macintosh, IBM PC, medium UNIX-based computers, Mainframes with proprietary operating systems, etc.) and sharing information between them. Since each computer system has its local

database with sector related information, in the network there only travels information related to other sectors (for instance: the Cardiology Department would need to consult the Neurology Department about a patient's neurologic history), resulting in good response time.

Another advantage is system growth. It is very easy to link another computer system into the network, provided the hardware and software to that particular system exists.

### **The use of relational databases in cardiology. An experience.**

As explained before, a Cardiology Department would benefit greatly from a computer network system, using either central or distributed processing, however, there are two important aspects which will determine the success of either system:

- the database design;
- the software used to implement the needs of the users.

In a medical Department such as Cardiology, where users need on-line database information, a well designed database is crucial. The application and database designer must understand in a clear way the relations between the data tables in order to optimise the search of information through the database and to establish a hierarchy scheme of access (by using passwords), restraining the available information to each user. The database should be designed in order not to have duplicated information, which will result in complicated administration and storage space waste. Our experience in database design tells us that a relational database [22] is the best approach, since it is much easier to establish relations between data, through data fields used to join the database tables.

We have been using the INFORMIX-4GL (and ESQL/C) package for database design and software development with good results. A great advantage of INFORMIX-4GL is its portability (without major changes) through a great number of different types of computers.

#### *The experience*

In the Cardiology Department of the University Hospital of Coimbra is installed a network system based on the central processing philosophy, since December, 1989, with an INFORMIX-4GL application running in the system.

The network consists of a UNIX-based system with several terminals and printers. At this early stage only three terminals are installed.

The application running in the system was designed to match the following goals:

- keep a record of the patient's final report, with the possibility of consulting the patient's cardiologic history;
- report on the usage of beds in the Cardiology Department, either by service or by patient's diagnostic.

The database keeps the following information in its tables:

- tables for:
  - diagnosis
  - patient follow-up
  - services in the Cardiology Department
  - patient's status
  - patient profession
- patient's record with:
  - process number
  - patient name
  - address, telephone
  - other clinical aspects
- patient's final report with:
  - service code
  - bed used by the patient
  - patient process number
  - dates for admission and release
  - diagnostics
  - 20 lines for a brief report

For the moment we are just beginning to feel the benefits of such a system, there have been some problems with the hardware installation resulting in difficulties of usage, but these are now totally solved.

At the present time we are developing a bibliography database for books and journals with cross-reference possibilities and two other cardiologic applications: one for the Intensive Care Unit and another for ECG.

In the near future we intend to develop more applications for other sectors, such as Echocardiography and the Catherization Lab and to link this system to other departments in the hospital and to other hospitals.

All these applications are being developed by a team from SIC, Sofcentro, Informatica do Centro, who are studying the use of INFORMIX-4GL in specialized database applications.

### **The ideal database in a cardiology network**

The ideal database should have some essential characteristics: It should allow better and quicker attendance of patients. All the data available should be kept secure by the use of a password to enter the system, and the information assessed should be limited, depending on professional sectors.

It should give doctors and nurses the possibility of getting more complete information, for clinical screening and research with fewer frustrations, and elaborate reports when the patient leaves the hospital. In summary, it should allow them more time for their patients.

For technical sectors, it avoids wastage of time with incomplete or illegible request forms. It should also be able to create a fully automatic report from the results and transmit it directly to the patient's record. Thus, the pressure in getting test results in a hurry should be largely avoided.

The staff of Outpatients and Admissions should have rapid access to clinical lists, waiting lists and the ward bed state (information which can be printed out

for analysis of hospital activity).

The library contents should also be fully computerized and be easily accessed from any sector in the Department.

Finally, the database should be connected to other departments in the hospital for exchange of information, and if possible, to other hospitals or countries with similar interests.

## Summary

The application of computers in Cardiology has passed beyond the first steps. They have several practical applications which transform them into indispensable tools in any Cardiological Department.

This is a rather complex unit which may benefit from the communication between computerized sectors. This is usually accomplished with local networks (LAN) which allow easy communication and exchange of information between these sectors. These computerized sectors connected by a network have many advantages not only for patients, but also for doctors and other hospital workers.

## References

1. Kember NF (1982) in An Introduction to Computer Applications in Medicine. Edward Arnold Publishers. 5
2. Hampshire N (1988) O Melhor De Dois Mundos. Personal Computer World. Portuguese ed Vol 2: 19:82
3. Hampshire N (1989) A Próxima Geração. Personal Computer World. Portuguese ed. Vol 3:, 25: 78
4. Marchesi C (1989) International drive to define new role for computer technology in medicine. Cardiology Today. Vol 17: 3: P3
5. Brennecke R (1988) A bottleneck model of imaging systems for digital angiocardiology. In Progress in Digital Angiocardiology. Kluwer Academic Publishers. Dordrecht. 337
6. Fehr R (1981) Pulsed-doppler with real-time fourier transform. In Echocardiology. Martinus Nijhoff Publishers. The Hague 443
7. Bruggeman T, Andresen D, Schroder R (1988) Effect of heart rate on the accuracy of detecting ST-segment changes in ambulatory ECG monitoring. Eur Heart J 9 (Suppl 1): 293
8. Kaltenbrunner *et al.* (1988) Performance of intraoperative computer-assisted mapping of ventricular tachycardia in coronary artery disease. Eur Heart J 9 (suppl 1): 293
9. Porento G, Dorffner H, Sochor H (1988) Computer assisted interpretation of Thallium-201 scintigrams using the parallel distributed processing approach. Eur Heart J 9 (Suppl 1): 136
10. Malik M (1988) Camm. computer modelling of heart-pacemaker interaction. Eur Heart J 9 (Suppl 1): 292
11. Killmann R, Dienstl F, Wach P, Eichtinger (1988) Simulation of activation sequence in ventricular myocardium: normal and pathologic cases. Eur Heart J 9 (Suppl 1): 293
12. Malik M Camm (1988) Computer modelling of cardiac repolarisation sequences and recovery processes. Eur Heart J 9 (Suppl 1): 137
13. Kember NF (1982) In an Introduction to Computer Applications in Medicine. Edward Arnold Publishers. 116

14. Piepoli M *et al.* (1989) Noninvasive cardiovascular diagnosis by a microcomputer based expert system. *Eur Heart J* 9 (suppl 1): 137
15. Griffo R, Molini D, Zannini, Mortola G, Vecchio C (1988) Computer assist diagnosis of heart disease (cardio-data-diagnosi). *Eur Heart J* (Suppl 1): 137
16. Kember NF (1982) In an Introduction to Computer Applications in Medicine. Edward Arnold Publishers. 88
17. Signore O, Bertoli D, Marotta G (1988) A data base for echo-doppler results. *Eur Heart J* (Suppl 1): 138
18. Wai G (1989) Take your pick, from client/server to transparency and parallel processing. Distributed processing use a variety of methods to share resources. *Byte* 14:7
19. Walker BJ (1989) Transparent environment. *Byte* 14:7
20. Manson C (1989) Remote control. *Byte* 14:7
21. Rosa C. Redes (1989) Personal Computer World Portuguese ed. 4
22. INFORMIX-SQL User's guide

# CADANS: the nervous system for cardiology

G.T. MEESTER, O.E.H. SARTORIUS, D. FANGGIDAEJ, M.G.J.M. GERRITSEN, J.A.T.G. VAN KESTER and F.M. VAN RAPPARD-VAN DER VEEN

*Interuniversity Institute of Cardiology (ICIN)  
P.O. Box 19258, Utrecht, The Netherlands*

## **Abstract**

The cardiological computer network in The Netherlands is described in this chapter. It consists basically of a local area network in each of the 8 participating university cardiological departments.

These local area networks are connected together in a 'star' format with a centre point in the ICIN location. The network is used for research studies, for example, the REGRESS study, where the effects of a cholesterol lowering drug are investigated. The data is stored in a relational database system. Results of this study will be presented later.

## **Introduction**

The departments of Cardiology in the eight Universities of The Netherlands have been working together for the past fifteen years in the framework of the Interuniversity Cardiology Institute of The Netherlands (ICIN) (Figure 1).

Recently, one of the leading non-university hospitals joined this co-operation (Figure 2).

The primary aim of ICIN is the enhancement of cardiological research, particularly in the fields of multi-centre and multi-disciplinary research. After a few years, postgraduate training also evolved to become a regular task of the institute, together with the organisation of international symposia and the introduction of new technologies in research and patient care.

This collaboration resulted in an effective development of goals, exchange of expertise and in many research projects, both in clinical and basic areas.

Multicentre patient-oriented research studies usually produce fairly large databases with extensive systems for maintenance, management and data-analysis.

In 1987, ICIN realised that in order to ensure the successful future of such studies, a structural approach to data-management was required.

After an introductory feasibility study, carried out by an outside agency, the





*Fig. 1.* Overview of The Netherlands with CADANS participating centres.

planning of a computer-based network for data management was initiated. This resulted in the outline of a plan called CADANS, named after its main task: to set up a *CARDiological DATA Network*.

Basically the plan detailed the construction of a computer network in star-format to connect the participating cardiological centres and the ICIN location.

A wide area network would then link local area networks within each of the departments of Cardiology.

The so-called 'definition study' contained a description of the architecture, construction and management necessary in order to carry out the required assignments.

Discussions, decision-making and planning involved utilising expertise from many areas of cardiology, computer science and management. All cardiology centres contributed to the final plan, either with manpower, advice or otherwise. The final definition study lists all contributors, including those from other outside agencies [1].

The activities were supported by the Ministry of Education and Science, who,

University of Amsterdam  
Free University of Amsterdam  
University of Groningen  
University of Leiden  
University of Maastricht  
University of Nijmegen  
Erasmus University Rotterdam  
University of Utrecht  
St. Antonius Hospital Nieuwegein

*Fig. 2.* List of participating Departments of Cardiology.

after the positive conclusions of the feasibility study, made it possible to carry out the definition studies.

Currently in 1990, the network is already installed and tested. The first multi-centre research project is presently being supported using the recently available infrastructure for database-management.

Aims, methods and details of this research project, falling under the acronym 'REGRESS', are provided in the second part of this chapter.

### **The CADANS network**

When designing a network for data-communication several requirements must be met.

The most obvious requirements are those from the viewpoints of technical performance and integrity of data-transport, data-management and data-analysis. However, because the data is almost always patient-related, strict conditions of privacy and security within the network must be imposed and maintained.

Research-projects within the ICIN framework are carried out by a dedicated project group under the leadership of one or more individuals from one of the participating centres, who carry final responsibility.

The above observations have led to the fundamental architecture of the CADANS-network. This network has the configuration of a star (Figure 3) where each of the participating Cardiology Departments communicates solely

to the centre of the star, where the project management is located. Data-flow is, in principle, one-way only.

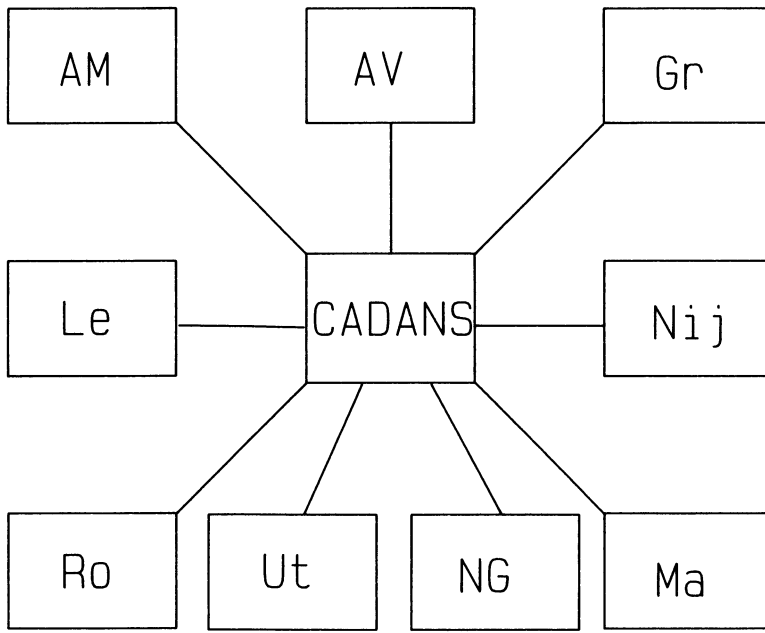


Fig. 3. CADANS wide area network in star configuration.

At each of the points of the star a local area network is provided as a hub for CADANS, consisting of a central server with two or more work-stations (Figure 4). The wide area network uses the digital communication lines as made available by the Telephone authorities (PTT). The central facility is also connected to the Dutch University Computer Network: 'SURFnet', which in turn is connected to other international nets, such as EARN (European Academic Research Net) DFN (Deutsche Forschungs Net) Bitnet in the United States and others. In fact, the Minister of Science and Education appointed CADANS as a centre of expertise within the SURFnet-environment.

When planning the hardware and software facilities, it was decided, whenever possible, to use standard available products, and to keep to the mainstream of connectivity developments as much as possible.

A database is a large assembly of data, systematically stored in a computer with the help of a database management program.

When selecting a database management program for the CADANS network, the following arguments were considered, among others: In a research database it is not always possible to completely specify in the beginning of a project how the data will be entered and interpreted. A flexible data structure is therefore essential. However, in order to avoid compromising the speed of retrieval, this feature should not be over-extended.

A hierarchical database needs a predetermined structure, which is afterwards

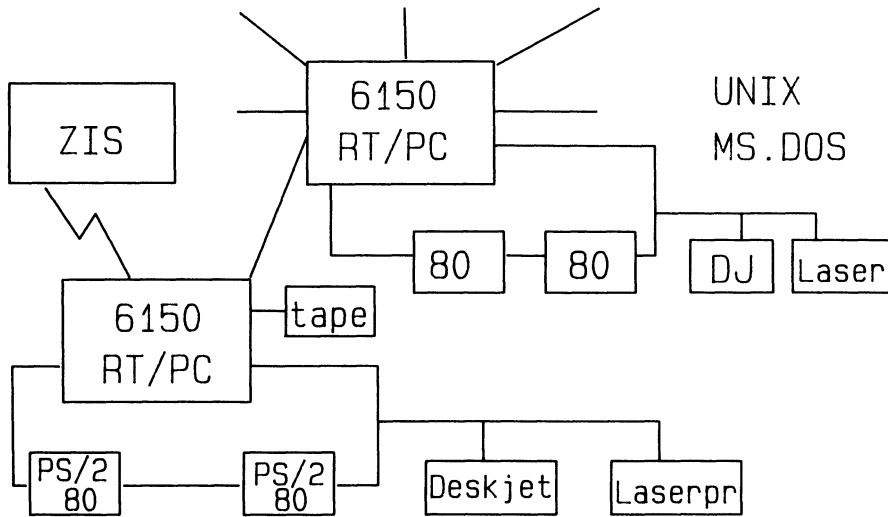


Fig. 4. CADANS local area network with two work stations.

very difficult, if not impossible, to change. For our purposes, therefore, a relational database system, allowing a larger degree of flexibility is a better choice.

A relational database is a collection of related tables. In this collection, tables can be adapted, edited or added during execution of the research project. The RDBMS (Relational Database Management System) should also be able to cope in an elegant manner with missing values (null values).

When a variable such as blood pressure, for instance, is unavailable, it should not automatically be set to zero (0) by the RDBMS. Some systems do this with all empty slots in the tables, later on during data-analysis this can easily lead to misrepresentation.

In addition, the RDBMS should have a built-in connection to analysis software, for reporting the contents and results together with a more statistical evaluation. This facility should preferably be available to the researcher without the obligatory intervention of a computer scientist.

The optimum solution for this requirement is considered by us to be the Structured Query Language (SQL). This language only consists of a small series of commands and provides extensive access to the data stored in the RDBMS.

SQL is designed for the end-user and provides, without professional help, answers to systematic and ad hoc interrogations of the database.

Experience shows that the necessary knowledge of SQL can be acquired in one or two days.

Apart from these demands, our RDBMS should be able to run under two different operating systems, UNIX for the network and MS-DOS for the work stations. These operating systems were chosen for the CADANS network, due to their widespread application in medical research. Furthermore, the selected RDBMS should also be portable, to enable them to function in other environ-

ments, due to the multitude of divergent hardware and software in the departments of Cardiology.

In conclusion, ORACLE<sup>1</sup> was chosen as the RDBMS for the CADANS network. This software package can accommodate large quantities of data, it has many helpful utilities and operates nicely in a network environment (SQL\*net).

### *The network*

The final network is required to embrace circa 300 workstations simultaneously. This endpoint is effectuated in a 'bottom-upwards' strategy, building up the actual network in a step-by-step manner, together with the assembly of the necessary expertise and experience in the field and certain decisions made in the design phases.

### *Work station*

For the elementary work-stations, an 80386 microcomputer<sup>2</sup> was selected, representative of the level of technology of today. The PS/2 microcomputers are mutually connected and in turn, they are all connected to the server key-node in a token-ring and/or Ethernet local area network depending on the local situation.

### *Key-node*

The key-node computer is a RISC-processor<sup>3</sup> as a multi-user, multi-tasking system. It acts as file and printer server (Figure 5).

This computer also supports the connections for the wide area CADANS network. Network facilities, such as back-ups are also supported here.

Furthermore, the key-node will serve as a connection to the Hospital Information Systems where this is required. CADANS as a research network operates, in principle, separately from the health-care facilities.

<sup>1</sup> ORACLE. The Relational Database Management system, Oracle Corporation Belmont, Ca, USA.

<sup>2</sup> IBM Personal System/2. Model 80/111.

<sup>3</sup> IBM 6150 Micro Computer System. Model 125.

<sup>4</sup> PC/TCP. FTP Software, Inc. Wakefield, MA, USA. TCP/IP. IBM Telecommunications Marketing Centre, Raleigh NC, USA.

<sup>5</sup> Interdrive. Implementation of the NSF protocol for the IBM PC. FTP Software Inc. Wakefield, MA. USA. NFS: Network File System. Sun Microsystems Inc. Mountainview, Ca. USA.

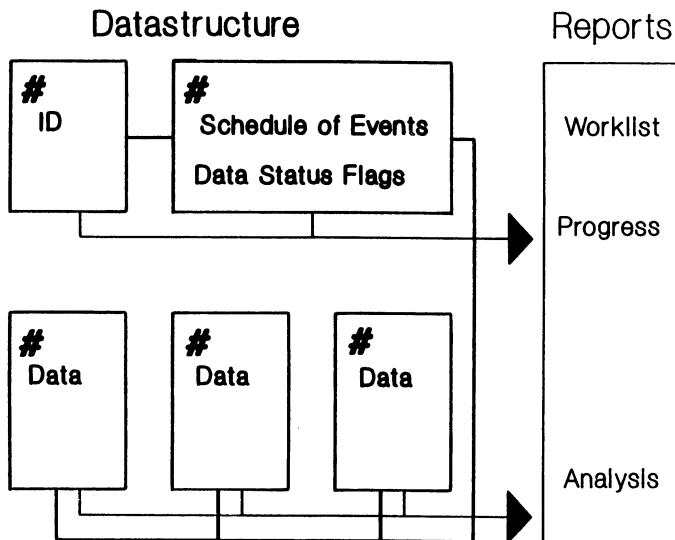


Fig. 5. Configuration of ORACLE for REGRESS.

### Software

The work-stations function under the MS-DOS operating system, probably to be replaced at a later stage by OS/2. The server software is UNIX, in this case of the IBM version AIX (derived from the System V source of AT&T, with many enhancements).

Communication support is provided based on the TCP/IP standard protocols.<sup>4</sup> These are defined both for Ethernet and Token-ring local area networks. The network file system is based on NFS from Sun Microsystems<sup>5</sup> providing transparent file-sharing and remote direct mounting. This protocol also enables remote file access.

In the work-stations a series of application programs is provided for text-processing, graphical presentations, desktop-publishing etc., all easily accessible by way of a custom designed user-interface.

The relational database management system (ORACLE) also provides information management and transaction processing. This RDBMS is designed to support concurrent usage by multiple applications and users. Many security features are available.

Oracle's SQL\*Net handles the communications inside the network and provides the capability of a distributed database. Protocol drivers for TCP/IP are available and functioning.

For data-analysis, SQL\*Plus and SQL\*Report come with this RDBMS, providing formatted reporting and interactive queries of the database. Oracle also provides interfaces to programming languages as Fortran and C.

For statistical evaluation program packages as SPSS<sup>6</sup> and SAS<sup>7</sup> are used.

### *Construction and installation*

After deliberation and consultation with many other network-managers, a single vendor approach was chosen. Delivery of hardware and software, cabling and installation was all placed into the hands of one vendor<sup>9</sup> to simplify the lines of responsibility and communication.

The overall situation is still reasonably complicated because the network connects nine different hospitals with various organisations in differing local situations.

### *Management*

No network will function without a basic organisation and fundamental standardisation. Together with the hardware and software selection, a series of management rules were established. These include:

Rules for:

- privacy and security
- hardware and software configuration maintenance
- problem-solving (help-desk)
- functionality
- responsibilities
- education and training
- back-up

These rules can perhaps best be summarised as follows: 'Everything that is not openly permitted is forbidden...!'

To facilitate use and maintenance of the CADANS network, two special quick-reference manuals were written, one for the LAN system manager and one for the end-user.

### *Acceptance*

The primary basic installation of the CADANS network has been given the character of a pilot study. This pilot study should demonstrate the consequences of the built-in choices, together with their advantages and disadvantages.

Possible weaknesses in the design should be recognised and corrected. Due to the many new elements in the realisation of this research-oriented network, the Ministry of Economic Affairs supports the pilot-study in part.

<sup>6</sup> SPSS/PC+ SPSS Inc. Chicago, Ill, USA.

<sup>7</sup> SAS. SAS Institute Inc. Car, NC, USA.

<sup>9</sup> IBM Nederland NV. Amsterdam, The Netherlands.

In order to test the network facility under realistic conditions, the execution of a true cardiological research project in this new environment is essential.

At the end of 1989, the Scientific Board of the ICN decided to start a study to assess the effects of Pravastatin<sup>8</sup>.

Pravastatin depresses cholesterol synthesis in the liver and the intestine: these organs are together responsible for 90% of all body cholesterol. This drug therefore makes it possible to decrease blood-cholesterol levels and hopefully also to delay or even regress the development of atherosclerosis. This process causes narrowing of blood-vessels resulting in cardiovascular incidents.

The project, going under the acronym of REGRESS, is ideally suited to be carried out by way of a first test for the CADANS network.

The project REGRESS itself and the construction of its database is described in the second part of this chapter.

### **REGRESS: Regression Growth Evaluation Station Study** **Project Leader: JD Barth**

The multicentre study called 'REGRESS' has been initiated in order to investigate the cholesterol-lowering effect on the natural progressive course of human coronary atherosclerosis [6, 7].

This study will make use of the CADANS network for data handling and networking. Many problems arise when a study of this size and complexity is started. The main challenge for the CADANS network in this study will be the data acquisition and management (3000 items per patient, 720 patients in 41 different centres!).

'REGRESS' (*REGression GRowth Evaluation Statin Study*) is a study that will investigate in 'normolipidemic' men, having angina and undergoing coronary cinearteriography, the effects of pravastatin on anatomic and clinical measures of coronary and peripheral atherosclerosis: Pravastatin being a competitive endogenous cholesterol synthesis inhibitor resulting in a lowering of blood cholesterol values.

#### *Study design*

Randomised, double-blind, placebo-controlled 2-year study of 720 men, with baseline and follow-up coronary cineangiography after 2 years. All angiographies will be quantitatively assessed using computer-assisted techniques [8, 9].

Patients with one  $> 50\%$  lesion (visual assessment) and with 'normal' total blood cholesterol values, will be provided with dietary information, and will be stratified into three groups according the initial management plan (i.e.

<sup>8</sup> Selektine (R) developed by E.R. SQUIBB & Sons.



Coronary Artery Bypass Grafting (CABG), Percutaneous Transluminal Coronary Angioplasty (PTCA) or Medical Management).

### **Drug administration**

Forty mg o.d. or placebo administered at bedtime. Patients whose total cholesterol rises above 8.0 mmol/L (hypercholesterolemic level) on re-check not responding to enhanced dietary efforts will have open-label cholestyramine (a cholesterol-lowering bile acid sequestrant) added to their treatment. Patients whose total cholesterol decreases to below 2.0 mmol/L (77 mg/dL) will have dose reduced by one-half due to dangerous low blood cholesterol levels that may interfere with cell wall integrity.

'Matching' patients in the other cohort will receive parallel changes in their treatment regimen.

### **Study evaluation**

Dietary intervention will be given to qualifying patients with anginal symptoms during cardiologic evaluation for angina.

Patients with a qualifying cineangiogram will have determination of coronary myocardial flow reserve and functional evaluation at baseline and end of study. Patients undergoing CABG or PTCA will have repeat functional assessment after the procedure. Patients at three centres will have carotid and femoral ultrasound determinations at six month intervals. Lipids and lipoprotein studies will be done during baseline, and at 2, 6, 12, 18 months and at the end of the study.

Other biochemical parameters will also be assessed. Safety assessments will be done at three month intervals. Provision for study extension (18-24 months) allowing continuation of blind study medication until all patients have completed follow-up cineangiogram is detailed in the protocol. In addition, the relationships of known DNA markers (genetic) with hyperlipidemic response and regression of coronary atherosclerosis will be assessed.

The contribution of CADANS will be crucial to the proper execution of the study in view of the complexity of the study, the necessity of rapid data transfer, the identification of sub-populations, reacting differently to the drug, as compared to those that will show clear progression or clear regression of coronary atherosclerosis, the safety aspects, the laboratories blind data transfer and the epidemiological/statistical management of REGRESS.

### *Database design*

REGRESS and other patient-oriented research studies are often based on the following data-model:

Initially, a series of data is assembled to describe the baseline situation of the

state under investigation. Then the patient undergoes an intervention, such as a drug, surgery or balloon-dilation. This intervention is defined as accurately as possible in a series of variables.

Thereafter, the follow-up study will begin, where the effects of the applied intervention are monitored by means of one or more patient observations, and for the final analysis closely related to each other and the baseline data.

Generally, the final analysis is partly confused by incomplete or missing data, patients lost to follow-up and patients crossing over from one treatment group into another, when two different interventions are compared. A RDBMS is well-suited to store the data of this particular model. Apart from storing the data being studied, the RDBMS can also serve to monitor the logistics of the project, to report progress, to produce work lists, and to perform quality control by input checking and looking for inconsistencies.

Missing data in such studies can be expected to be of three different types: permanently missing data (data unavailable, patient lost to follow-up. etc); temporarily missing data (data presently unavailable, but will come later); irrelevant data (not relevant for this particular patient or situation). For many analyses the difference between these types of incomplete data has to be appreciated.

The data model as implemented for REGRESS under ORACLE is presented in Figure 5.

The database consists of a series of tables mutually linked using a patient identification number. A separate table, related to the identification table (ID) stores the schedule of events. It keeps track of study progress and missing data, thus forming the source for work-lists and progress reports. Before transportation of data to the central computer, the ID data is for reasons of privacy, separated from the observed data.

Data entry will be facilitated by SQL\*Forms, the front end of the database. To support easy data entry the computer screens have almost the same layout as the code sheets with the patient data. During this process checks are performed for consistency and errors. When a form is completed the corresponding data status flag in the events table will be set to a higher value indicating that the form is ready for transport to the central computer. After transport, the data status flag is again set to a higher value, indicating that the data has been transported. In the event of later alterations, the data status is lowered and the form will again be transported.

The implementation of the RDBMS is carried out in three sequential phases. Firstly, the database is configured and tested in a prototype in a stand-alone microcomputer. After successful testing, the RDBMS is installed in all participating centres, data communication to the central facility of CADANS is by means of floppy disks by post. In the final phase, data is communicated to the central database through the digital network.

At present (early 1990) the first phase is almost complete. The prototype has been tested with the data of approximately 100 patients. Data entry modules are available at three locations, now locking as available in Oracle is also implemented.

Standardised reports have been prepared with SQL\* Reportwriter, a menu-driven utility, enabling fast reporting in ASCII format.

Response times in the Oracle RDBMS may increase depending on secondary processing. Every query is checked for consistency, type of requested data, privileges of the user, etc. In addition, load-dependent network-speed and hard-disk access times may influence response time. The RDBMS should therefore also be optimised for speed by defining logical partitions, optimum file management and clusters.

In the second and third phases, two other utilities (IMPORT and EXPORT) take care of data transportation to the central database. In the peripheral stations, new and/or edited records are selected based on their status flags and the data of entry or change. Selected data is assembled in a so-called view, and then via floppy disk or digital line exported to the central database, where they replace existing records.

### *Summary*

Our experiences both with the network and with the relational database are still limited. It has become clear that a computer network is a vulnerable entity where a certain level of discipline in maintenance and updating is essential. The benefits of such a system in terms of increased productivity with regard to file sharing (address lists, literature references, electronic mail) sharing of devices (back-ups, tape-drives, printers, plotters, etc.) and in general for inter-office communications, are impressive.

In order to start using a relational database system, a fairly long introductory period must be anticipated. This depends partly on the experience available, but is generally a minimum of 3-6 months.

The adaptability of this database management system, together with the many available utilities, makes it eminently suitable for the cardiological research environment.

### **References**

1. CADANS – The information network for the Dutch University Cardiology. Report to the Minister of Education & Science by a working group of the ICIN. Project leader: Prof. G.T. Meester MD, PhD.
2. CADANS – Design and implementation of a data network for Cardiology. Meester GT, Sartorius OEH, Fanggideaj D (1987) *Computers in Cardiology*, Leuven Belgium
3. CADANS – A Progress Report. Meester GT, Sartorius OEH, Fanggideaj D *et al.* (1988) *Computers in Cardiology*, Washington DC
4. CADANS – A customised network for Cardiology in The Netherlands. Meester GT, Sartorius OEH, Fanggideaj D, Gerritsen MGJM (1989) 11th Ann. International Conference of IEEE, Seattle, USA
5. Regression of Atherosclerosis. Heart Disease, Update 4. Eugene Braunwald, Blankenhorn DH (eds)
6. Arntszenius AC, Kromhout D, Barth JD *et al.* (1985) Diet, lipoproteins and the progression of

- coronary atherosclerosis: The Leiden Intervention trial. *N Engl J Med* 312: 805-811
7. Barth JD, Jansen H, Kromhout D, Reiber JHC, Birkenhäger JC, Arntzenius AC (1987) Progression and regression of human coronary atherosclerosis. The role of lipoprotein, lipases and thyroid hormones in coronary lesion growth. *Atherosclerosis* 68: 51-58
  8. Reiber JHC, Serruys PW, Kooijman CJ, Slager CJ, Schuurbiers JCH, Den Boer A (1986) Approaches towards standardisation in acquisition and quantitation of arterial dimensions from cineangiograms. In *State of the art in quantitative coronary arteriography*. Reiber JHC, Serruys PW (eds) Martinus Nijhoff Pub. Dordrecht/Boston 145-174
  9. Mancini GBJ (1988) Quantitative Coronary Arteriography: Developments of methods, limitations and clinical applications. *Am J Card Imag* 2: 98-109

# Index

- acceptance 191
- actuarial curves 234
- advantages 192
- aggregate functions 81
- alphanumeric string 8
- analysis 5
- angioplasty 48
- ANSI 66
- archival storage 311
- arrhythmias 344
- ARTEL 155
- ARTEMIS 155
- ASCII 6
- artificial intelligence 87
- Attributes 9
- audit trial 133
- authorization 211
- automatic transfer of data 322
  
- backup 5
- balanced trees (B-trees) 18
- basic concepts 3
- Bayes' theorem 277
- Bayesian analysis 273
- binary search 13
- bits 6
- BMDP (Biomedical Data Package) 38
- bytes 6
  
- C language 65
- CADANS 401
- case retrieval 181
- CCU 143
- CD ROMs 63, 110
- Computer Aided Instruction 61
- Central Processing 395
- checklist 347
- classification 289
- clinical problem solving 113
  
- clinical profiles 42
- closed loop system 350
- codes 8
- coding document 183
- coding systems 291
- Colleague 101
- Compact Cambridge Medline 110
- compliance of users 326
- Comprehensive Core Medical Library (CCML) 103
- computer-aided instruction (CAI) 114, 135
- confidentiality 211
- confounders 46
- congenital heart disease 273, 289
- consistency 229
- coronary angiography 179
  - coronary angiography information systems 197
  - coronary arterial segments 184
  - coronary severity score 188
- coronary surgery 223
- cost of accessing data banks 108
- COSTAR 66
- courseware 133
  
- data analysis 46
  - data capture 300
  - data completeness 300
  - data definition 46
  - 'data driven' approach 162
  - data elements 5
  - data processing 273
  - data retrieval 310
  - data security and privacy 300
  - data types 6
  - data validation 300
  - data-aquisition screens 219
  - data-model 410
- database management systems 4

- administrative data bases 224
- clinical research data bases 225
- database fields 83
- database model 21
- database of images 312
- DATAEASE 16, 34
- Data-Star, Dimdi 105
- DATATRIEVE 95
- dBASE 16, 33, 359
- demographic database 7
- distributed system 36
- external databases 97
- fields 5
- mimimum data base 360
- network database structure 198
- observational databases 41
- online database 98
- quality assurance 43
- quality control data bases 224
- reference databases 98
- Relational Model 24
  - mathematical models 243
- source databases 98
  - observational data collection 45
  - quality of the data 229
  - time-dependent fields 83
  - variable-length fields 10
- decision making 350
- delimiter 10
- derived indices 46
- design 4
- diagnosis
  - Diagnostic Codes 262
  - diagnostic coding system 259
  - diagnostic decisions 273
  - diagnostic 'trees' 259
- Dialorder 107
- 4th Dimension 35
- direct access 16
- Distributed Processing 395
- DOCLINE 107
- document delivery 107
  
- Easynet 99
- echocardiography 69, 387
- editing of data 5
- education 141
- efficiency 198
- Electrocardiography
  - EKG monitoring 344
  - EKG standards 353
  - EKGLab 125
- electronic mail 210
- Embase Plus 103
- entities 88
  
- entity-relationship model 28
- enumerated sets 8
- ergonomy 219
- expert systems 87, 156
- explorer-2 120
  
- file design 9
- File Maker 35
- File Manager 37
- Fileman 23, 37
- files 5
- fixed-field files 16
  - fixed-field files: index sequential 13
  - fixed-field files: indexed file 15
- fixed length record 10
  - fixed-length fields 10
  - fixed-length file 12
  - variable-length fields 10
- flow chart 78
- follow-up 201
- FoxBase 35, 351
- function keys 6
- fundamental knowledge 241
  
- gateways 99
- GPEP 113
- graphic 8
  - graphic resolution 59
  
- Hashed Files 18
- Hashing Algorithm 16
- hazard function 232
- heart transplant patients 201
- HeartLab 125
- help 325
- hidden costs 348
- hierarchical model 22
- high resolution screens 64
  - graphic resolution 60
- Holter monitoring 315
- hyperlink 126
- hypermedia 35, 69, 115
  - Hypercard 35, 69
  - HyperHeart 130
  - hypertalk 70
  - hypertext 35, 116
- hypertension 155
  
- ICDA codes 8
- IMAGE database 182
- IMS system 23
- inference engine 88
- INFORMIX-4GL 397
- INGRES 35
- input of data 5

- integrated transactions 77
- intelligent tutoring system (ITS) 136
- Intelligent Information Facility 100
- intensive care 343
  - coronary care unit 335
  - intensive care patient monitoring systems 347
  - intensive care unit 239
  - invasive monitoring 344
- intersection 27
- invariant data 205
- ISO 68
  
- join 26
- join-tree 80
  
- keys 12
- key words 259
- knowledge processing 158
  - knowledge base 88, 155
  - knowledge resources 124
  
- Laser Card 72
- Leaves 18
- life table 233
- Local Area Network 295, 396
- log of user activity 133
- LOTUS 60, 123
  
- machine readable form 5
- maintenance 5
- managing information 3
- matching 47
- medical connection (DMC) 100
- medical decisions 162
- Medical Information Bus 353
- medical record number 183
- MEDIS 104
- MEDIX 35, 69
- Medizinische Technik (MEDITEC) 105
- MEDLINE 105, 111
- MeSH descriptors 105
- minitel 201
- models
  - Conceptual Data Model 207
  - mathematical models 243
- modem 97
- MQL, Medical Query Language 66
- MS-DOS 56, 382
- multi-tasking 368
- MUMPS 10, 66
- myocardial revascularisation 180
  
- natural join 26
- network 393
  - Network Model 24
    - conceptual Data Model 207
    - neural networks 63
  - NEXT 68
  - NFS 407
  - normalization 24
  - numeric 8
  
  - One-to-many relationship 320
  - One-to-one relationship 320
  - ORACLE 35, 383
  - organ support 345
  - OSI 68
  - out-patient clinic 381
  - outlier analysis 229
  
  - pacemaker 167
  - PACERS 106
  - PARADOX 60
  - parallel processing 53
  - patient data management 346
  - patient duplication 302
  - patient outcomes 43
  - pattern analysis 46
  - pattern recognition 42
  - patterns 46
  - pediatric cardiology 259
  - performance 20, 159
  - personal computer 53
  - Personal Productivity Tools 65
  - personnel 227
  - pointers 8
  - population matching 48
  - predictive value 161
  - primary key 24
  - problem-solving exercises 123
  - program instrumentation 134
  - projection 26
  - Prolog 65
  - “propensity” scores 46
  - PTCA 78, 179
  
  - query language 77, 307
    - entity-relationship model 28
    - MQL, Medical Query Language 67
    - query-by-example (QBA) 84
    - SQL 27
    - SQL-PLUS 383
    - Structured Query Language 27
  
  - randomization 44
  - randomized clinical trials 45
  - receiver operating characteristic (ROC)
    - curves 279
    - ROC analysis 273, 277

- records 5
- recursive partitioning 46
- REGRESS 409
- relations 24, 88
- reporting program 183
- reports 5
- resource utilisation 259
- restriction 9, 26
- revision 44
- risk factor analysis 231
  - risk factor modification 43
  - risk prediction 48
  - risk-stratification 48
- Scientific American Medicine (SAMM) 105
- SCISEARCH 106
- search algorithm 206
- searching the data 5
- selection-tree 81
- semantic network 157
- semantic relation 89
- sensitivity 276
- sequence of data elements 11
- sequential file 12
- serial comparison 198
- shared access 203
- significant relevant 231
- SilverPlatter 111
- simulations 116
- Slice of Life 125
- software portability 58
- soundex 16
- specificity 276
- SPSS (Statistical Package for Social Sciences) 38
- SQL 27
- SQL-PLUS 383
- standardization 56
- statistical analysis 77
  - statistical evaluation 310
  - statistical packages 38
  - statistics-tree 81
  - storage efficiency 20
  - storage supports 57
  - string numeric 8
  - Structured Query Language 27
  - SURFnet 404
  - survival analysis 228
  - SYBASE 36
  - Symphony 60
- TCP/IP 407
- teaching education 113, 141
- technology evaluation 43
- TELETEL 209
- temporal linkage 198
- therapy control 176
- time-varying data 204
- training 60
- transfer of data 313
- transmission speed 58
- TRANSPAC 209
- transputer 53
- Tree-Base 78
- tuple 24
- ultrasound 307
- Unified Medical Language System (UMLS) 121
- UNIX 36, 382
- video-disc 63
- videotex server 201
- workstations 406
- X.25 72
- Zim 37



# Developments in Cardiovascular Medicine

---

1. Ch.T. Lancée (ed.): *Echocardiology*. 1979 ISBN 90-247-2209-8
2. J. Baan, A.C. Arntzenius and E.L. Yellin (eds.): *Cardiac Dynamics*. 1980  
ISBN 90-247-2212-8
3. H.J.Th. Thalen and C.C. Meere (eds.): *Fundamentals of Cardiac Pacing*. 1979  
ISBN 90-247-2245-4
4. H.E. Kulbertus and H.J.J. Wellens (eds.): *Sudden Death*. 1980 ISBN 90-247-2290-X
5. L.S. Dreifus and A.N. Brest (eds.): *Clinical Applications of Cardiovascular Drugs*.  
1980 ISBN 90-247-2295-0
6. M.P. Spencer and J.M. Reid: *Cerebrovascular Evaluation with Doppler Ultrasound*.  
With contributions by E.C. Brockenbrough, R.S. Reneman, G.I. Thomas and D.L.  
Davis. 1981 ISBN 90-247-2384-1
7. D.P. Zipes, J.C. Bailey and V. Elharrar (eds.): *The Slow Inward Current and Cardiac  
Arrhythmias*. 1980 ISBN 90-247-2380-9
8. H. Kesteloot and J.V. Joossens (eds.): *Epidemiology of Arterial Blood Pressure*. 1980  
ISBN 90-247-2386-8
9. F.J.Th. Wackers (ed.): *Thallium-201 and Technetium-99m-Pyrophosphate. Myocar-  
dial Imaging in the Coronary Care Unit*. 1980 ISBN 90-247-2396-5
10. A. Maseri, C. Marchesi, S. Chierchia and M.G. Trivella (eds.): *Coronary Care Units*.  
Proceedings of a European Seminar, held in Pisa, Italy (1978). 1981  
ISBN 90-247-2456-2
11. J. Morganroth, E.N. Moore, L.S. Dreifus and E.L. Michelson (eds.): *The Evaluation of  
New Antiarrhythmic Drugs*. Proceedings of the First Symposium on New Drugs and  
Devices, held in Philadelphia, Pa., U.S.A. (1980). 1981 ISBN 90-247-2474-0
12. P. Alboni: *Intraventricular Conduction Disturbances*. 1981 ISBN 90-247-2483-X
13. H. Rijsterborgh (ed.): *Echocardiology*. 1981 ISBN 90-247-2491-0
14. G.S. Wagner (ed.): *Myocardial Infarction. Measurement and Intervention*. 1982  
ISBN 90-247-2513-5
15. R.S. Meltzer and J. Roelandt (eds.): *Contrast Echocardiography*. 1982  
ISBN 90-247-2531-3
16. A. Amery, R. Fagard, P. Lijnen and J. Staessen (eds.): *Hypertensive Cardiovascular  
Disease. Pathophysiology and Treatment*. 1982 ISBN 90-247-2534-8
17. L.N. Bouman and H.J. Jongsma (eds.): *Cardiac Rate and Rhythm. Physiological,  
Morphological and Developmental Aspects*. 1982 ISBN 90-247-2626-3
18. J. Morganroth and E.N. Moore (eds.): *The Evaluation of Beta Blocker and Calcium  
Antagonist Drugs*. Proceedings of the 2nd Symposium on New Drugs and Devices,  
held in Philadelphia, Pa., U.S.A. (1981). 1982 ISBN 90-247-2642-5
19. M.B. Rosenbaum and M.V. Elizari (eds.): *Frontiers of Cardiac Electrophysiology*.  
1983 ISBN 90-247-2663-8
20. J. Roelandt and P.G. Hugenholtz (eds.): *Long-term Ambulatory Electrocardiography*.  
1982 ISBN 90-247-2664-6
21. A.A.J. Adgey (ed.): *Acute Phase of Ischemic Heart Disease and Myocardial  
Infarction*. 1982 ISBN 90-247-2675-1
22. P. Hanrath, W. Bleifeld and J. Souquet (eds.): *Cardiovascular Diagnosis by  
Ultrasound. Transesophageal, Computerized, Contrast, Doppler Echocardiography*.  
1982 ISBN 90-247-2692-1

## Developments in Cardiovascular Medicine

---

23. J. Roelandt (ed.): *The Practice of M-Mode and Two-dimensional Echocardiography*. 1983 ISBN 90-247-2745-6
24. J. Meyer, P. Schweizer and R. Erbel (eds.): *Advances in Noninvasive Cardiology*. Ultrasound, Computed Tomography, Radioisotopes, Digital Angiography. 1983 ISBN 0-89838-576-8
25. J. Morganroth and E.N. Moore (eds.): *Sudden Cardiac Death and Congestive Heart Failure*. Diagnosis and Treatment. Proceedings of the 3rd Symposium on New Drugs and Devices, held in Philadelphia, Pa., U.S.A. (1982). 1983 ISBN 0-89838-580-6
26. H.M. Perry Jr. (ed.): *Lifelong Management of Hypertension*. 1983 ISBN 0-89838-582-2
27. E.A. Jaffe (ed.): *Biology of Endothelial Cells*. 1984 ISBN 0-89838-587-3
28. B. Surawicz, C.P. Reddy and E.N. Prystowsky (eds.): *Tachycardias*. 1984 ISBN 0-89838-588-1
29. M.P. Spencer (ed.): *Cardiac Doppler Diagnosis*. Proceedings of a Symposium, held in Clearwater, Fla., U.S.A. (1983). 1983 ISBN 0-89838-591-1
30. H. Villarreal and M.P. Sambhi (eds.): *Topics in Pathophysiology of Hypertension*. 1984 ISBN 0-89838-595-4
31. F.H. Messerli (ed.): *Cardiovascular Disease in the Elderly*. 1984  
Revised edition, 1988: see below under Volume 76
32. M.L. Simoons and J.H.C. Reiber (eds.): *Nuclear Imaging in Clinical Cardiology*. 1984 ISBN 0-89838-599-7
33. H.E.D.J. ter Keurs and J.J. Schipperheyn (eds.): *Cardiac Left Ventricular Hypertrophy*. 1983 ISBN 0-89838-612-8
34. N. Sperelakis (ed.): *Physiology and Pathology of the Heart*. 1984  
Revised edition, 1988: see below under Volume 90
35. F.H. Messerli (ed.): *Kidney in Essential Hypertension*. Proceedings of a Course, held in New Orleans, La., U.S.A. (1983). 1984 ISBN 0-89838-616-0
36. M.P. Sambhi (ed.): *Fundamental Fault in Hypertension*. 1984 ISBN 0-89838-638-1
37. C. Marchesi (ed.): *Ambulatory Monitoring*. Cardiovascular System and Allied Applications. Proceedings of a Workshop, held in Pisa, Italy (1983). 1984 ISBN 0-89838-642-X
38. W. Kupper, R.N. MacAlpin and W. Bleifeld (eds.): *Coronary Tone in Ischemic Heart Disease*. 1984 ISBN 0-89838-646-2
39. N. Sperelakis and J.B. Caulfield (eds.): *Calcium Antagonists*. Mechanism of Action on Cardiac Muscle and Vascular Smooth Muscle. Proceedings of the 5th Annual Meeting of the American Section of the I.S.H.R., held in Hilton Head, S.C., U.S.A. (1983). 1984 ISBN 0-89838-655-1
40. Th. Godfraind, A.G. Herman and D. Wellens (eds.): *Calcium Entry Blockers in Cardiovascular and Cerebral Dysfunctions*. 1984 ISBN 0-89838-658-6
41. J. Morganroth and E.N. Moore (eds.): *Interventions in the Acute Phase of Myocardial Infarction*. Proceedings of the 4th Symposium on New Drugs and Devices, held in Philadelphia, Pa., U.S.A. (1983). 1984 ISBN 0-89838-659-4
42. F.L. Abel and W.H. Newman (eds.): *Functional Aspects of the Normal, Hypertrophied and Failing Heart*. Proceedings of the 5th Annual Meeting of the American Section of the I.S.H.R., held in Hilton Head, S.C., U.S.A. (1983). 1984 ISBN 0-89838-665-9

## Developments in Cardiovascular Medicine

---

43. S. Sideman and R. Beyar (eds.): [3-D] *Simulation and Imaging of the Cardiac System*. State of the Heart. Proceedings of the International Henry Goldberg Workshop, held in Haifa, Israel (1984). 1985 ISBN 0-89838-687-X
44. E. van der Wall and K.I. Lie (eds.): *Recent Views on Hypertrophic Cardiomyopathy*. Proceedings of a Symposium, held in Groningen, The Netherlands (1984). 1985 ISBN 0-89838-694-2
45. R.E. Beamish, P.K. Singal and N.S. Dhalla (eds.), *Stress and Heart Disease*. Proceedings of a International Symposium, held in Winnipeg, Canada, 1984 (Vol. 1). 1985 ISBN 0-89838-709-4
46. R.E. Beamish, V. Panagia and N.S. Dhalla (eds.): *Pathogenesis of Stress-induced Heart Disease*. Proceedings of a International Symposium, held in Winnipeg, Canada, 1984 (Vol. 2). 1985 ISBN 0-89838-710-8
47. J. Morganroth and E.N. Moore (eds.): *Cardiac Arrhythmias*. New Therapeutic Drugs and Devices. Proceedings of the 5th Symposium on New Drugs and Devices, held in Philadelphia, Pa., U.S.A. (1984). 1985 ISBN 0-89838-716-7
48. P. Mathes (ed.): *Secondary Prevention in Coronary Artery Disease and Myocardial Infarction*. 1985 ISBN 0-89838-736-1
49. H.L. Stone and W.B. Weglicki (eds.): *Pathobiology of Cardiovascular Injury*. Proceedings of the 6th Annual Meeting of the American Section of the I.S.H.R., held in Oklahoma City, Okla., U.S.A. (1984). 1985 ISBN 0-89838-743-4
50. J. Meyer, R. Erbel and H.J. Rupprecht (eds.): *Improvement of Myocardial Perfusion*. Thrombolysis, Angioplasty, Bypass Surgery. Proceedings of a Symposium, held in Mainz, F.R.G. (1984). 1985 ISBN 0-89838-748-5
51. J.H.C. Reiber, P.W. Serruys and C.J. Slager (eds.): *Quantitative Coronary and Left Ventricular Cineangiography*. Methodology and Clinical Applications. 1986 ISBN 0-89838-760-4
52. R.H. Fagard and I.E. Bekaert (eds.): *Sports Cardiology*. Exercise in Health and Cardiovascular Disease. Proceedings from an International Conference, held in Knokke, Belgium (1985). 1986 ISBN 0-89838-782-5
53. J.H.C. Reiber and P.W. Serruys (eds.): *State of the Art in Quantitative Coronary Arteriography*. 1986 ISBN 0-89838-804-X
54. J. Roelandt (ed.): *Color Doppler Flow Imaging and Other Advances in Doppler Echocardiography*. 1986 ISBN 0-89838-806-6
55. E.E. van der Wall (ed.): *Noninvasive Imaging of Cardiac Metabolism*. Single Photon Scintigraphy, Positron Emission Tomography and Nuclear Magnetic Resonance. 1987 ISBN 0-89838-812-0
56. J. Liebman, R. Plonsey and Y. Rudy (eds.): *Pediatric and Fundamental Electrocardiography*. 1987 ISBN 0-89838-815-5
57. H.H. Hilger, V. Hombach and W.J. Rashkind (eds.), *Invasive Cardiovascular Therapy*. Proceedings of an International Symposium, held in Cologne, F.R.G. (1985). 1987 ISBN 0-89838-818-X
58. P.W. Serruys and G.T. Meester (eds.): *Coronary Angioplasty*. A Controlled Model for Ischemia. 1986 ISBN 0-89838-819-8
59. J.E. Tooke and L.H. Smaje (eds.): *Clinical Investigation of the Microcirculation*. Proceedings of an International Meeting, held in London, U.K. (1985). 1987 ISBN 0-89838-833-3

## Developments in Cardiovascular Medicine

---

60. R.Th. van Dam and A. van Oosterom (eds.): *Electrocardiographic Body Surface Mapping*. Proceedings of the 3rd International Symposium on B.S.M., held in Nijmegen, The Netherlands (1985). 1986 ISBN 0-89838-834-1
61. M.P. Spencer (ed.): *Ultrasonic Diagnosis of Cerebrovascular Disease*. Doppler Techniques and Pulse Echo Imaging. 1987 ISBN 0-89838-836-8
62. M.J. Legato (ed.): *The Stressed Heart*. 1987 ISBN 0-89838-849-X
63. M.E. Safar (ed.): *Arterial and Venous Systems in Essential Hypertension*. With Assistance of G.M. London, A.Ch. Simon and Y.A. Weiss. 1987 ISBN 0-89838-857-0
64. J. Roelandt (ed.): *Digital Techniques in Echocardiography*. 1987 ISBN 0-89838-861-9
65. N.S. Dhalla, P.K. Singal and R.E. Beamish (eds.): *Pathology of Heart Disease*. Proceedings of the 8th Annual Meeting of the American Section of the I.S.H.R., held in Winnipeg, Canada, 1986 (Vol. 1). 1987 ISBN 0-89838-864-3
66. N.S. Dhalla, G.N. Pierce and R.E. Beamish (eds.): *Heart Function and Metabolism*. Proceedings of the 8th Annual Meeting of the American Section of the I.S.H.R., held in Winnipeg, Canada, 1986 (Vol. 2). 1987 ISBN 0-89838-865-1
67. N.S. Dhalla, I.R. Innes and R.E. Beamish (eds.): *Myocardial Ischemia*. Proceedings of a Satellite Symposium of the 30th International Physiological Congress, held in Winnipeg, Canada (1986). 1987 ISBN 0-89838-866-X
68. R.E. Beamish, V. Panagia and N.S. Dhalla (eds.): *Pharmacological Aspects of Heart Disease*. Proceedings of an International Symposium, held in Winnipeg, Canada (1986). 1987 ISBN 0-89838-867-8
69. H.E.D.J. ter Keurs and J.V. Tyberg (eds.): *Mechanics of the Circulation*. Proceedings of a Satellite Symposium of the 30th International Physiological Congress, held in Banff, Alberta, Canada (1986). 1987 ISBN 0-89838-870-8
70. S. Sideman and R. Beyar (eds.): *Activation, Metabolism and Perfusion of the Heart*. Simulation and Experimental Models. Proceedings of the 3rd Henry Goldberg Workshop, held in Piscataway, N.J., U.S.A. (1986). 1987 ISBN 0-89838-871-6
71. E. Aliot and R. Lazzara (eds.): *Ventricular Tachycardias*. From Mechanism to Therapy. 1987 ISBN 0-89838-881-3
72. A. Schneeweiss and G. Schettler: *Cardiovascular Drug Therapy in the Elderly*. 1988 ISBN 0-89838-883-X
73. J.V. Chapman and A. Sgalambro (eds.): *Basic Concepts in Doppler Echocardiography*. Methods of Clinical Applications based on a Multi-modality Doppler Approach. 1987 ISBN 0-89838-888-0
74. S. Chien, J. Dormandy, E. Ernst and A. Matrai (eds.): *Clinical Hemorheology*. Applications in Cardiovascular and Hematological Disease, Diabetes, Surgery and Gynecology. 1987 ISBN 0-89838-807-4
75. J. Morganroth and E.N. Moore (eds.): *Congestive Heart Failure*. Proceedings of the 7th Annual Symposium on New Drugs and Devices, held in Philadelphia, Pa., U.S.A. (1986). 1987 ISBN 0-89838-955-0
76. F.H. Messerli (ed.): *Cardiovascular Disease in the Elderly*. 2nd ed. 1988 ISBN 0-89838-962-3
77. P.H. Heintzen and J.H. Bürsch (eds.): *Progress in Digital Angiocardiography*. 1988 ISBN 0-89838-965-8

## Developments in Cardiovascular Medicine

---

78. M.M. Scheinman (ed.): *Catheter Ablation of Cardiac Arrhythmias*. Basic Bioelectrical Effects and Clinical Indications. 1988 ISBN 0-89838-967-4
79. J.A.E. Spaan, A.V.G. Brusckhe and A.C. Gittenberger-De Groot (eds.): *Coronary Circulation*. From Basic Mechanisms to Clinical Implications. 1987 ISBN 0-89838-978-X
80. C. Visser, G. Kan and R.S. Meltzer (eds.): *Echocardiography in Coronary Artery Disease*. 1988 ISBN 0-89838-979-8
81. A. Bayés de Luna, A. Betriu and G. Permanyer (eds.): *Therapeutics in Cardiology*. 1988 ISBN 0-89838-981-X
82. D.M. Mirvis (ed.): *Body Surface Electrocardiographic Mapping*. 1988 ISBN 0-89838-983-6
83. M.A. Konstam and J.M. Isner (eds.): *The Right Ventricle*. 1988 ISBN 0-89838-987-9
84. C.T. Kappagoda and P.V. Greenwood (eds.): *Long-term Management of Patients after Myocardial Infarction*. 1988 ISBN 0-89838-352-8
85. W.H. Gaasch and H.J. Levine (eds.): *Chronic Aortic Regurgitation*. 1988 ISBN 0-89838-364-1
86. P.K. Singal (ed.): *Oxygen Radicals in the Pathophysiology of Heart Disease*. 1988 ISBN 0-89838-375-7
87. J.H.C. Reiber and P.W. Serruys (eds.): *New Developments in Quantitative Coronary Arteriography*. 1988 ISBN 0-89838-377-3
88. J. Morganroth and E.N. Moore (eds.): *Silent Myocardial Ischemia*. Proceedings of the 8th Annual Symposium on New Drugs and Devices (1987). 1988 ISBN 0-89838-380-3
89. H.E.D.J. ter Keurs and M.I.M. Noble (eds.): *Starling's Law of the Heart Revisited*. 1988 ISBN 0-89838-382-X
90. N. Sperelakis (ed.): *Physiology and Pathophysiology of the Heart*. (Rev. ed.) 1988 ISBN 0-89838-388-9
91. J.W. de Jong (ed.): *Myocardial Energy Metabolism*. 1988 ISBN 0-89838-394-3
92. V. Hombach, H.H. Hilger and H.L. Kennedy (eds.): *Electrocardiography and Cardiac Drug Therapy*. Proceedings of an International Symposium, held in Cologne, F.R.G. (1987). 1988 ISBN 0-89838-395-1
93. H. Iwata, J.B. Lombardini and T. Segawa (eds.): *Taurine and the Heart*. 1988 ISBN 0-89838-396-X
94. M.R. Rosen and Y. Palti (eds.): *Lethal Arrhythmias Resulting from Myocardial Ischemia and Infarction*. Proceedings of the 2nd Rappaport Symposium, held in Haifa, Israel (1988). 1988 ISBN 0-89838-401-X
95. M. Iwase and I. Sotobata: *Clinical Echocardiography*. With a Foreword by M.P. Spencer. 1989 ISBN 0-7923-0004-1
96. I. Cikes (ed.): *Echocardiography in Cardiac Interventions*. 1989 ISBN 0-7923-0088-2
97. E. Rapaport (ed.): *Early Interventions in Acute Myocardial Infarction*. 1989 ISBN 0-7923-0175-7
98. M.E. Safar and F. Fouad-Tarazi (eds.): *The Heart in Hypertension*. A Tribute to Robert C. Tarazi (1925-1986). 1989 ISBN 0-7923-0197-8
99. S. Meerbaum and R. Meltzer (eds.): *Myocardial Contrast Two-dimensional Echocardiography*. 1989 ISBN 0-7923-0205-2

## Developments in Cardiovascular Medicine

---

100. J. Morganroth and E.N. Moore (eds.): *Risk/Benefit Analysis for the Use and Approval of Thrombolytic, Antiarrhythmic, and Hypolipidemic Agents*. Proceedings of the 9th Annual Symposium on New Drugs and Devices (1988). 1989 ISBN 0-7923-0294-X
101. P.W. Serruys, R. Simon and K.J. Beatt (eds.): *PTCA - An Investigational Tool and a Non-operative Treatment of Acute Ischemia*. 1990 ISBN 0-7923-0346-6
102. I.S. Anand, P.I. Wahi and N.S. Dhalla (eds.): *Pathophysiology and Pharmacology of Heart Disease*. 1989 ISBN 0-7923-0367-9
103. G.S. Abela (ed.): *Lasers in Cardiovascular Medicine and Surgery*. Fundamentals and Technique. 1990 ISBN 0-7923-0440-3
104. H.M. Piper (ed.): *Pathophysiology of Severe Ischemic Myocardial Injury*. 1990 ISBN 0-7923-0459-4
105. S.M. Teague (ed.): *Stress Doppler Echocardiography*. 1990 ISBN 0-7923-0499-3
106. P.R. Saxena, D.I. Wallis, W. Wouters and P. Bevan (eds.): *Cardiovascular Pharmacology of 5-Hydroxytryptamine*. Prospective Therapeutic Applications. 1990 ISBN 0-7923-0502-7
107. A.P. Shepherd and P.A. Öberg (eds.): *Laser-Doppler Blood Flowmetry*. 1990 ISBN 0-7923-0508-6
108. J. Soler-Soler, G. Permanyer-Miralda and J. Sagristà-Sauleda (eds.): *Pericardial Disease*. New Insights and Old Dilemmas. Preface by Ralph Shabetai. 1990 ISBN 0-7923-0510-8
109. J.P.M. Hamer: *Practical Echocardiography in the Adult*. With Doppler and Color-Doppler Flow Imaging. 1990 ISBN 0-7923-0670-8
110. A. Bayés de Luna, P. Brugada, J. Cosin Aguilar and F. Navarro Lopez (eds.): *Sudden Cardiac Death*. 1991 ISBN 0-7923-0716-X
111. E. Andries and R. Stroobandt (eds.): *Hemodynamics in Daily Practice*. 1991 ISBN 0-7923-0725-9
112. J. Morganroth and E.N. Moore (eds.): *Use and Approval of Antihypertensive Agents and Surrogate Endpoints for the Approval of Drugs affecting Antiarrhythmic Heart Failure and Hypolipidemia*. Proceedings of the 10th Annual Symposium on New Drugs and Devices (1989). 1990 ISBN 0-7923-0756-9
113. S. Iliceto, P. Rizzon and J.R.T.C. Roelandt (eds.): *Ultrasound in Coronary Artery Disease*. Present Role and Future Perspectives. 1990 ISBN 0-7923-0784-4
114. J.V. Chapman and G.R. Sutherland (eds.): *The Noninvasive Evaluation of Hemodynamics in Congenital Heart Disease*. Doppler Ultrasound Applications in the Adult and Pediatric Patient with Congenital Heart Disease. 1990 ISBN 0-7923-0836-0
115. G.T. Meester and F. Pinciroli (eds.): *Databases for Cardiology*. 1991 ISBN 0-7923-0886-7
116. B. Korecky and N.S. Dhalla (eds.): *Subcellular Basis of Contractile Failure*. 1990 ISBN 0-7923-0890-5
117. J.H.C. Reiber and P.W. Serruys (eds.): *Quantitative Coronary Arteriography*. 1991 ISBN 0-7923-0913-8
118. E. van der Wall and A. de Roos (eds.): *Magnetic Resonance Imaging in Coronary Artery Disease*. 1991 ISBN 0-7923-0940-5

## Developments in Cardiovascular Medicine

---

119. V. Hombach, M. Kochs and A.J. Camm (eds.): *Interventional Techniques in Cardiovascular Medicine*. 1991 ISBN 0-7923-0956-1
120. R. Vos: *Drugs Looking for Diseases*. Innovative Drug Research and the Development of the Beta Blockers and the Calcium Antagonists. 1991 ISBN 0-7923-0968-5
121. S. Sideman, R. Beyar and A. G. Kleber (eds.): *Cardiac Electrophysiology, Circulation, and Transport*. 1991 ISBN 0-7923-1145-0
122. D.M. Bers: *Excitation-Contraction Coupling and Cardiac Contractile Force*. 1991 ISBN 0-7923-1186-8

*Previous volumes are still available*

---

KLUWER ACADEMIC PUBLISHERS – DORDRECHT / BOSTON / LONDON